



MINISTRY OF EDUCATION, YOUTH AND SPORT



# PRINCIPLES OF ASSESSING STUDENTS & TEACHERS

A Facilitator's Training Handbook & Course Materials to Improve Educational Evaluation

Phnom Penh  
November 2022

# TABLE OF CONTENTS

<i>Acknowledgements</i>	iv
<i>Preface</i>	v
<b>1. INTRODUCTION .....</b>	<b>1</b>
1.1 Role of Evaluation in the Teaching-Learning Process	
1.2 Teacher Evaluation	
1.3 Purpose of This Manual	
1.4 Applications of This Manual	
<b>2. HOW TO USE THIS MANUAL .....</b>	<b>5</b>
<b>3. FACILITATOR SESSION PLANS .....</b>	<b>6</b>
<b>PART 1: Basic Principles and Concepts of Student Evaluation</b>	<b>7</b>
3.1 Measurement & Evaluation in Educational Decision-making	
3.2 Types of Measurement & Evaluation: Formative and Summative Evaluation	
3.3 Useful Frameworks for Interpreting Test Scores: Normative and Criterion-referenced Testing	
3.4 Characteristics of Good Tests and Examinations: Concepts of Validity	
<b>PART 2: Practical Strategies for Test Design and Development</b>	<b>25</b>
3.5 Instructional Objectives in the Educational Process	
3.6 Developing Tables of Specification	
3.7 Writing Objective Test Items	
3.8 Writing and Scoring Essay Questions	
<b>PART 3: Question Banking</b>	<b>49</b>
3.9 Conducting Item Analysis and Constructing Question Banks	
<b>PART 4: Principles and Guidelines for Teacher Evaluation</b>	<b>55</b>
3.10 Why We Evaluate Teachers	
3.11 What to Assess during Teacher Evaluation	
3.12 Who Should Evaluate Teachers	
3.13 Organizing Teacher Evaluation and Linkages with Staff Development	
3.14 Bringing the Elements of Teacher Evaluation Together	
<b>4. COURSE MATERIALS .....</b>	<b>74</b>
<b>Handouts for PART 1</b>	<b>75</b>
<b>Handout 1:</b> Why Do We Evaluate?	75
<b>Handout 2:</b> Types of Educational Decision-making	76
<b>Handout 3:</b> How Do We Get Information for Our Decision-making?	77
<b>Handout 4:</b> How Evaluation Purpose Affects Test Design	78
<b>Handout 5a:</b> Types of Evaluation for the Classroom	80
<b>Handout 5b:</b> Summary of the Differences between Summative & Formative Evaluation	86
<b>Handout 6:</b> Useful Frameworks for Interpreting Test Scores	87
<b>Handout 7:</b> Criterion-referenced Scoring	88
<b>Handout 8:</b> Norm-referenced Scoring	89
<b>Handout 9:</b> Comparison between Norm and Criterion Referenced Scoring	92

<b>Handout 10:</b> Checking Your Knowledge on Evaluation Principles	93
<b>Handout 11:</b> Concepts of Assessment Validity	95
<b>Handout 12a:</b> Kinds of Validity	96
<b>Handout 12b:</b> Reinforcement Exercise on Concepts of Validity	100
<b>Handout 13:</b> Some Final Guidelines to Ensure Validity in Your Tests	101
<b>Handouts for PART 2</b>	<b>102</b>
<b>Handout 14:</b> Why Instructional Objectives are Important in Assessment	102
<b>Handout 15:</b> The Role of Instructional Objectives in the Education Process	104
<b>Handout 16:</b> Defining an Instructional Objective	105
<b>Handout 17:</b> Kinds of Instructional Objectives	108
<b>Handout 18a:</b> Guidelines When Writing Instructional Objectives	110
<b>Handout 18b:</b> Using Bloom’s Taxonomy to Develop & Classify Instructional Objectives	111
<b>Handout 19:</b> Developing Tables of Specification	115
<b>Handout 20:</b> About Objective and Subjective Test Questions	119
<b>Handout 21:</b> How to Write Objective Test Questions Effectively	120
<b>Handout 22:</b> Defining Subjective Test Questions and Understanding Their Limitations	134
<b>Handout 23:</b> The Use of Directing Words When Writing Essay Questions	136
<b>Handout 24:</b> Challenges in Scoring Essay Questions	140
<b>Handout 25:</b> Overview of Essay Questions	144
<b>Handouts for PART 3</b>	<b>145</b>
<b>Handout 26:</b> About Item Analysis	145
<b>Handouts for PART 4</b>	<b>154</b>
<b>Handout 27:</b> Knowing Why You Are Evaluating a Teacher	154
<b>Handout 28:</b> Different Kinds of Teacher Evaluation	156
<b>Handout 29a:</b> Know What You Want to Evaluate When You Assess a Teacher	158
<b>Handout 29b:</b> Variations That May Occur in What You Evaluate	162
<b>Handout 30:</b> Who Should Do Teacher Evaluation?	165
<b>Handout 31:</b> Guidelines for Organizing Teacher Conferences	168
<b>Handout 32:</b> Linking Your Teacher Evaluation System to Staff Development	170
<b>Handout 33:</b> Bringing the Elements of Effective Teacher Evaluation Together	172
<b>5. ASSESSMENT TERM GLOSSARY</b> .....	<b>178</b>
<b>ANNEX 1: Sample Teacher Skills Framework</b> .....	<b>183</b>

## **Acknowledgements**

The committee responsible for developing this manual would like to thank all who participated in its development. This includes the various advisers working in Kampuchea Action to Promote Education who shared many internal training documents without any conditions relating to proprietary rights. Special thanks are also due to the instructors and advisers working at the New Generation Pedagogical Research Center at the National Institute of Education who also provided much useful feedback about the document. This greatly helped to improve the many handouts that form a significant part of this document.

## Preface

The present manual was developed by Kampuchea Action to Promote Education and in collaboration with the Ministry of Education, Youth, & Sport to help Cambodian educators working at all levels to improve the quality of both student assessment and teacher evaluation. Most teacher preparation programs in Cambodia do not spend a great deal of time on Principles of Student Assessment. If such principles are covered at all, they are usually treated as a secondary part of some other course or subject. As a result, many teachers are often thrust into the classroom with a poor understanding of how to make the evaluation of their students as valid and reliable as possible. This manual provides a systematic approach to understanding the Principles of Student Assessment that starts with basic concepts such as testing validity and reliability and proceeds step by step to the basic nuts and bolts of proper test development. The manual also provides guidance on how teachers can conduct classroom-based item analysis so that they can determine how well the questions that they have developed actually work (i.e., are they too difficult, are they too easy, etc.).

This manual also includes a special section on Teacher Evaluation that seeks to help schools and educational institutions set up systems of assessment that are effective in promoting teachers' 'professional growth' as well as more traditional forms of teacher assessment focused on 'accountability' or quality assurance. This section does not dictate what schools and institutions should do to establish effective teacher evaluation systems but rather provides flexible guidelines that may be adapted to multiple contexts.

The developers of this manual hope that those educators who complete the course contained within this document will be able to more effectively assess both their students and teachers and be sure that the decisions they make are based on valid data.

# 1. INTRODUCTION

## 1.1 Role of Evaluation in the Teaching-Learning Process

Traditionally, evaluation has tended to be a neglected area in the preparation of teachers not only in Cambodia but in many countries of the world, as well. Prospective teachers studying in training colleges often receive a very short summary of how to do student assessment as part of a broader course rather than as a distinct area of study in its own right. This neglect of educational assessment in the preparation of teachers has led to the perception of evaluation practices as invariable and monolithic, regardless of the context in which they are undertaken. In this respect, evaluation is most commonly thought of as the development and administration of tests for the purpose of assigning grades. While this is certainly one of the most common uses of evaluation, it overlooks its fundamental nature, which is essentially the *generation of information that is then used for a wide range of educational decision-making*.

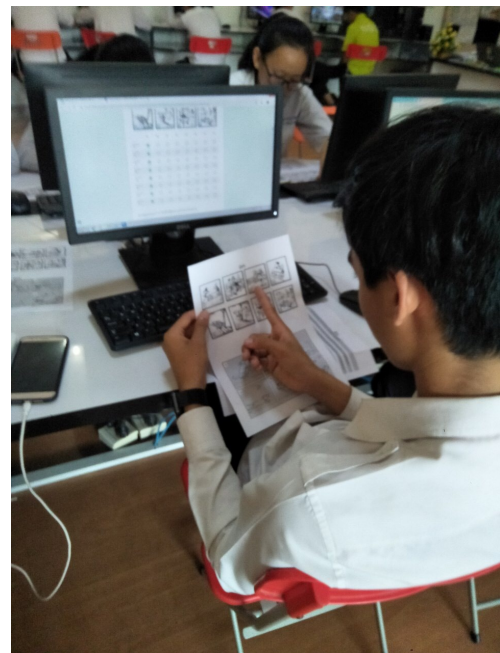
Whenever we make educational decisions, we need information. Such decisions may relate to lesson areas in need of review, promotion of individual students, curriculum revision, awarding of certificates, and numerous other matters. Evaluation is the manner through which we generate the information to make these decisions.

While one of the most common forms that evaluation takes is indeed ‘testing,’ this is certainly not the only form, which it may embody. Rather, evaluation may occur as observations, interviews, project work, or surveys, among others. Particularly in the 21<sup>st</sup> Century, classroom evaluation strategies have been evolving rapidly so that assessments that were never possible 20 years ago are increasingly routine. This includes electronic testing, electronically tracking the amount of time students spend online, etc. Similarly, evaluation standards (e.g., the difficulty level of questions) may be set at a very high level or a very low level. Relatedly, the scoring techniques used in one’s evaluation may also vary ranging from relative rankings of students as is frequently done at the end of each month in Cambodian classrooms or the use of absolute standards of performance (e.g., setting 5 as the criterion for passing on a test of 10 points), as is usually done in most Cambodian classrooms. These variations in evaluation practice should underline the



### ***What is Evaluation?***

*[Evaluation is defined as] . . . the generation of information that is then used for a wide range of educational decision-making.*



***Evaluation in the 21<sup>st</sup> Century:*** A student in a New Generation School takes a test ‘electronically’.

point that one's evaluation design will vary according to the educational decision that one wishes to make (e.g., promotion, selection, etc.). If the evaluation design used is not consistent with the purpose, it is likely that the evaluation conducted will not be very effective.

The list of evaluation practices presented in Box 1 provides a set of examples of evaluation that is in some way faulty. If you have ever found yourself using evaluation in some of the ways described, then it is likely that your evaluation was not as effective as you thought. For those who might be surprised to learn that the examples described in Box 1 are instances of poor evaluation practice, the reasoning behind each case will be discussed in subsequent sections of this manual. For now, the most important point to grasp is that there must be consistency between one's avowed purpose in evaluating (e.g., promotion, selection, instructional planning, etc.) and the nature and design of the evaluation selected.

Following upon the line of thought explained above, the starting point for all evaluation MUST be with determining the 'purpose' of the evaluation to be conducted. For the classroom teacher, there are basically two purposes for which evaluation is generally conducted. One of these purposes refers to *formative* uses while the other refers to those that are *summative* in focus.

The technical terminology used above should be not intimidating. Most teachers are probably already very familiar with the kinds of evaluation strategies entailed by both formative and summative evaluation. For example, homework assignments, weekly quizzes, and question and answer techniques are all examples of *formative evaluation*, which seeks to help the teacher 'form' new instructional planning (e.g., what lessons should be re-taught, which students need the most extra help, etc.). Annual, term, and monthly tests are examples of *summative evaluation*, which the teacher uses to make final marking records that are used in promotional decisions.

### Box 1: Examples of Faulty Evaluation Practices

- **Common Testing Practice:** Grading a test whose purpose is formative and recording student scores in the Classroom Marking Book for promotion decisions.  
*Why It's Wrong:* *Formative evaluation helps teachers know what to re-teach. We never record the grades from formative tests in the Final Marking Book, as this is not its purpose.*
- **Common Testing Practice:** Developing a very difficult test to determine students who should be passed to the next grade.  
*Why It's Wrong:* *Promotional tests should include mostly questions of moderate difficulty with some small number that are either easy or difficult. We should never include many questions of high difficulty in a promotional test. Difficult questions are only appropriate for specialized selection tests such as for a scholarship or special award.*
- **Common Testing Practice:** Using essay questions exclusively on a test to determine students' overall level of achievement.  
*Why It's Wrong:* *Essay questions have very poor sampling coverage in terms of curricular content. Summative tests seeking to assess overall student achievement should include mostly objective questions that can be answered quickly with only a small number of essay questions that focus on higher level thinking skills that objective test items cannot easily assess.*

**If you have ever used any of these common testing practices in your own evaluation, then this manual is especially for you.**

## 1.2 Teacher Evaluation

While much of this manual focuses on the evaluation of students, a special section is also included on the evaluation of teachers (see **Part 4**). Teacher Evaluation, as it is described in this manual, may be undertaken by mentors and peers in the case of **Formative Teacher Evaluation** or by school directors, inspectors, and administrators in the case of **Summative Teacher Evaluation**. In general, most teacher evaluation in Cambodia tends to be ‘summative’ in nature with a strong focus on promoting ‘accountability’ for teacher performance. When we think of teacher inspection, this is the most common kind of teacher evaluation that we usually think of. While this manual also provides guidance on conducting Summative Teacher Evaluation, there is also a strong focus on Formative Teacher Evaluation, which is intended to promote ‘professional growth.’ Such evaluation is often overlooked in the Cambodian context, which may be one reason why traditional forms of Teacher Evaluation (e.g., teacher inspection) have not really been effective in fostering change in teachers’ behavior or classroom practice. It is, therefore, hoped that this manual may also be helpful for educators in establishing school or institution-based Teacher Evaluation Systems that are more balanced and effective than those that currently exist.

## 1.3 The Purpose of This Manual

It commonly happens that teachers and general educators sometimes do not consider the purpose of their evaluation before they start developing their assessment strategies. This results in a mismatch between ‘purpose’ and ‘design.’ In the same way, teachers and educators often do not consider important design principles that relate to validity, question difficulty, and other important evaluation concepts when developing their tests or assessment tools. This manual is designed to address these and similar deficiencies that frequently occur in the way that student/teacher assessment is carried out in Cambodian educational institutions.



**Relevant Content:** The content of this manual has been designed with the following outcomes in mind:

1. Participants can carry out student assessment in a way that ensures harmony between the purpose of evaluation and the design of assessment strategies to fulfill this purpose.
2. Participants can carry out teacher assessment in a way that ensures harmony between the purpose of evaluation and the design of assessment strategies to fulfill this purpose.
3. Participants can use concepts of validity to assess the appropriateness of the assessment tools that they develop.
4. Participants can document their educational objectives in a way that facilitates the evaluation process.
5. Participants can use guidelines for effective test development that utilize an understanding of the use of Tables of Specifications and the advantages and disadvantages of objective and subjective question types.



6. Participants can use concepts of item difficulty and discrimination to determine the appropriateness of a test item on a given test for a given purpose.
7. Participants can develop question banks that archive questions according to analyses of their effectiveness, grade level, and subject.
8. Participants can more effectively organize balanced teacher evaluation systems at their schools or institutions that promote both professional growth and accountability.

#### 1.4 Applications of This Manual in Cambodia

This manual has been designed to facilitate the development and administration of hands-on workshops on educational evaluation. The manual contains a set of user-friendly *Session Plans* to help a trainer to systematically present a training program on Student and/or Teacher Evaluation. This includes fundamental principles in evaluation as well as practical strategies in the design of assessment tools. Each session plan is set out with a suggested time frame, statement of needed preparation, useful materials and resources, and learning objectives. The manual also contains a set of *Participant Course Materials* that should be provided to participants attending the training program so that they can use these as reference documents after the conclusion of the training workshop.



## 2. EXPLANATION ABOUT HOW TO USE THIS MANUAL

This manual uses the following standard symbols to make the manual as user friendly as possible for the trainers using it. These symbols quickly convey the kinds of activities to be used with participants for each step of each session plan.



### **Timing Required for the Lesson**



**Pre-Training Preparation:** Contains information on how to set up your training area for learning activities. It also gives suggestions on how to organize materials needed for the activity.



**Resources & Materials Needed:** This provides an overview of necessary materials. Most of the time these will be very basic things, like pens or paper. Other material in the training sessions will be provided through handouts that are attached in this manual. Sometimes they need to be duplicated by the trainer.



**Learning Outcomes:** Gives a statement of what should have been achieved and assessed at the end of the session.

### Steps and Process Icons

Section 3 of this manual provides a series of training sessions on how to train the primary school teachers who are expected to employ Constructivist Learning techniques in their classrooms. The symbols below are used to help guide the actual training session. These symbols will tell the facilitator quickly what sorts of activities need to be planned for in this part of the training session. This section of the manual includes possible activities, stimulating questions, examples to clarify exercises and optional extra tasks. Although it is advisable to read through the whole lesson clearly from the beginning, especially when used for the first time, trainers can easily see what he or she has to do because of the icons used.



**Action to be Taken:** This symbol indicates that the facilitator must take a concrete action such as passing out a Handout, re-arranging desks, organize groups, etc.



**Questioning Behavior:** This indicates that the facilitator needs to ask a key question to the participants as a prelude to an activity or discussion.



**Discussion:** This symbol indicates that the facilitator must lead a discussion or allow participants to discuss something in their groups.

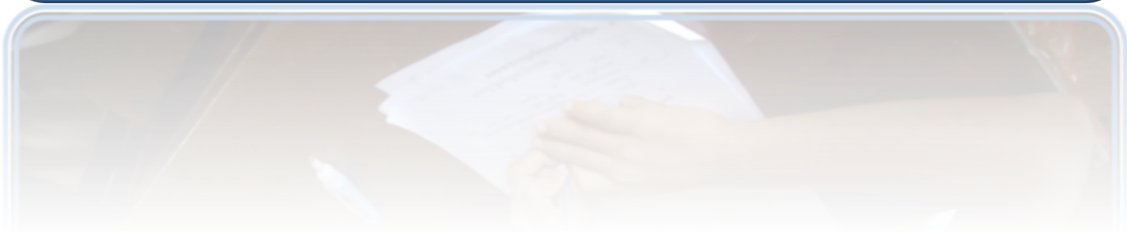


**Writing Tasks:** This symbol indicates that the participants need to write something on poster paper, complete an exercise, or other written task.



**Explanation:** This symbol indicates the facilitator must explain something to participants.

# SECTION 3: Facilitator Session Plans



# PART 1: Basic Principles and Concepts of Student Evaluation

## 3.1 Measurement & Evaluation in Educational Decision-making



**Lesson Time:** 2 Hours and 30 Minutes



### Trainer Preparation:

- Write up the Learning Outcomes of the lesson on a sheet of poster paper to introduce the lesson.
- Make copies of all handouts



### Resources/Materials:

- Poster paper, marker pens
- Poster sheet summarizing the learning outcomes of the lesson.
- **Handout 1:** *Why Do We Evaluate?*
- **Handout 2:** *Types of Educational Decision-making*
- **Handout 3:** *How Do We Get Information for Our Educational Decision-making?*
- **Handout 4:** *How Evaluation Purpose Affects Test Design*



### Learning Outcomes:

1. Participants can *explain* why we evaluate, specifically the links between Educational *Decision-making*, *Information*, and the *Evaluation Process*.
2. Participants can *explain* the different types of educational decision-making (e.g., certification decisions, remedial decisions, etc.) by giving specific examples of each (e.g., Bac II Examination, Scholarship Tests, etc..)
3. Participants can *state* the different methodologies (e.g., written tests, oral tests, surveys, observations, etc.) that one might use for gathering information for educational decision-making.
4. Participants can *determine* how test design (e.g., frequency, question difficulty, scope, etc.) might change depending on the purpose of one's evaluation.

## Training Session Plan

### Outcomes of the Lesson



Place a sheet of poster paper up on the board that summarizes the learning outcomes for the lesson. Explain the outcomes and that these results exemplify what the participants should be able to do at the completion of the lesson.

### Explaining Why We Evaluate

The Facilitator should begin this session by asking a simple question,



*“Why do we evaluate?”*



Write this question on the white board for all participants to see. Try to brainstorm a list with participants on the blackboard about all the possible reasons that we do evaluation. Possible answers might include the following:

- To put marks in our grade books
- To pass and fail students
- To know who the students are that we need to reteach.
- To know how much of our lesson the students understood.
- The Ministry requires us to do evaluation on a regular basis.
- Etc.

If someone says, ‘To get information’ or ‘To make decisions’, stop immediately and bring attention to this answer because this is one of the two responses that we are looking for. If no one gives this response, keep on asking ‘why’ for each of the responses provided above: why do we need to put marks in our grade books, why does the Ministry ask us to do evaluation, etc. until the facilitator can arrive at one of the two desired responses.



Once the facilitator helps participants to arrive at the conclusion that the reason we evaluate is to get INFORMATION and that we need information to make DECISIONS, write the following formula on the board and explain the connection between each term:



**Evaluate → To get Information → For Decision-making**

For example, what happens if we get bad information? Does the nature of our decisions affect the design of our evaluation? How can one’s purpose affect evaluation design?

The Facilitator should explain how confusion about one’s purpose can lead to faulty educational decisions. For example, teachers often administer formative tests (whose purpose is to re-teach students whose understanding is still weak) and then record the marks in their grade books for a ‘summative’ purpose. The marks that they record may no longer represent the final understanding level of the student after they

**EXAMPLE**

have been retaught. If the student fully understands the lesson after reteaching, then the score recorded in the marking book after the formative test does not provide accurate information about the student's 'true' understanding of a given lesson. Thus, a faulty decision may result because teachers used the information generated by their evaluation for a summative purpose when the original purpose was formative in nature. Such problems often arise because teachers are not clear themselves about the 'purpose' of their evaluation.

Next, pass out **Handout 1: 'Why Do We Evaluate'** and review the points made in the previous discussion.

### **Kinds of Educational Decisions & Getting the Information Needed to Make Them**

Following the above discussion, the Facilitator should review the kinds of educational decisions that we may need to make as educators. Pass out **Handout 2: 'Types of Educational Decision-making'** to help provide this explanation. The Facilitator should try to elicit different examples of the various kinds of educational decision-making indicated in the Handout. Write out some of the examples provided by participants in the sample table provided below:



<b>Educational Decision</b>	<b>Examples</b>
<b>Promotion Decisions</b>	<ul style="list-style-type: none"> <li>• Passing students</li> <li>• Repeating students</li> </ul>
<b>Remedial Decisions</b>	<ul style="list-style-type: none"> <li>• Placing students in special groups where they can receive extra help</li> <li>• Levelled Reading Grouping</li> </ul>
<b>Instructional Design Decisions</b>	<ul style="list-style-type: none"> <li>• Revising a lesson plan</li> <li>• Revising instructional materials</li> </ul>
<b>Certification Decisions</b>	<ul style="list-style-type: none"> <li>• Bac II Exam</li> <li>• Awarding Diplomas</li> </ul>
<b>Selection Decisions</b>	<ul style="list-style-type: none"> <li>• NGS Admission Test</li> <li>• Scholarship Awards</li> </ul>

Next, the Facilitator should quickly review some of the evaluation strategies that might be used to collect the needed information to make the educational decisions that were reviewed in the handouts. For this purpose, pass out **Handout 3: 'How Do We Get Information for Our Educational Decision-making?'**

Lead a short discussion to review each of the evaluation methods that might be used for collecting information, many of which are probably already well-known to participants (e.g., written tests, oral tests, surveys, etc.). During the discussion, ask participants which strategies go best with what kinds of educational decisions.



### **How Evaluation Purpose Affects Test Design**

The Facilitator should next start a discussion about how one's evaluation purpose will affect the design of one's test. This discussion gets to the heart of the issue concerning the importance of knowing one's purpose before starting to



design one's evaluation. Start the discussion by first identifying the various parameters to consider in test design and how purpose can affect each of these. To help with this explanation, pass out **Handout 4: 'How Evaluation Purpose Affects Test Design.'** The first page of this Handout explains seven parameters in test design and how these can vary on a test.



The Facilitator should explain each of these parameters to participants as presented in the table below and in the Handout.



Parameter	Options	Parameter	Options
<b>Scoring Procedures</b>	<ul style="list-style-type: none"> <li>• Criterion-Referenced</li> <li>• Normative</li> </ul>	<b>Timing</b>	<ul style="list-style-type: none"> <li>• Often/Infrequently</li> <li>• Continuous/Terminal</li> </ul>
<b>Test Length</b>	<ul style="list-style-type: none"> <li>• Short</li> <li>• Medium</li> <li>• Long</li> </ul>	<b>Kinds of Questions</b>	<ul style="list-style-type: none"> <li>• Objective/Subjective</li> <li>• Product or Process-focused</li> </ul>
<b>Question Difficulty</b>	<ul style="list-style-type: none"> <li>• Easy</li> <li>• Difficult</li> </ul>	<b>Evaluation Strategies</b>	<ul style="list-style-type: none"> <li>• Written Test</li> <li>• Oral Test</li> <li>• Observation</li> <li>• Survey</li> <li>• Etc.</li> </ul>
<b>Content Coverage</b>	<ul style="list-style-type: none"> <li>• Comprehensive</li> <li>• Sampling</li> </ul>		

Following this explanation, the Facilitator should lead a discussion with participants about how each of these parameters might vary in a test's design when the purpose is 'selection.' An example of how each parameter might vary is provided in the Handout. Review the example with participant by going over each parameter, asking questions about why it is best for the test to have the stated characteristics.



Following the review of this example, break participants into small groups and ask them to do a similar exercise on their own when the purpose of the evaluation is 'formative' in nature. Provide each group about 30 minutes for this exercise and then review the answers provided by each group in plenary session.



Compare the answers of each group with preferred answers as provided in the completed table below. Resolve any areas of differences through discussion. Don't forget to remind participants that when the purpose of one's evaluation is formative, the teacher is usually looking to see if students have achieved basic mastery of a majority of the objectives (usually 50% or more of educational objectives). Thus, the questions should be relatively easy. Since we should conduct formative evaluation frequently, the number of objectives evaluated is comprehensive though few in number.

**Evaluation Purpose: Formative**

Parameter	Options	Parameter	Options
<b>Scoring Procedures</b>	<ul style="list-style-type: none"> <li>• Criterion-Referenced</li> </ul>	<b>Timing</b>	<ul style="list-style-type: none"> <li>• Often</li> <li>• Continuous</li> </ul>

<b>Test Length</b>	<ul style="list-style-type: none"> <li>• Short</li> </ul>	<b>Kinds of Questions</b>	<ul style="list-style-type: none"> <li>• Subjective (though objective questions can also sometimes be used)</li> <li>• Process-focused</li> </ul>
<b>Question Difficulty</b>	<ul style="list-style-type: none"> <li>• Easy</li> </ul>	<b>Evaluation Strategies</b>	<ul style="list-style-type: none"> <li>• Written (or Oral) Test</li> </ul>
<b>Content Coverage</b>	<ul style="list-style-type: none"> <li>• Comprehensive</li> </ul>		



## 3.2 Types of Measurement & Evaluation: Formative and Summative Evaluation



**Lesson Time:** 3 Hours and 30 Minutes



### **Trainer Preparation:**

- Write up the Learning Outcomes of the lesson on a sheet of poster paper to introduce the lesson.
- Make copies of all Handouts as listed below.



### **Resources/Materials:**

- Poster paper, marker pens
- Poster sheet summarizing the learning outcomes of the lesson.
- **Handout 5a:** *'Types of Evaluation for the Classroom'*
- **Handout 5b:** *Summary of the Differences between Summative & Formative Evaluation*



### **Learning Outcomes**

1. Participants can *define* different kinds of evaluation (e.g., Formative, Summative, etc.) in terms of their purpose.
2. Participants can *explain* the differences between different kinds of evaluation in terms of when it is used and how it is done.
3. Participants can *use* examples of formative evaluation to help them analyze students' mistakes.
4. Participants can *explain* the similarities and differences between summative and formative evaluation.

## Training Session Plan

### Outcomes of the Lesson



Place a sheet of poster paper up on the board that summarizes the learning outcomes for the lesson. Explain the outcomes and that these results exemplify what the participants should be able to do at the completion of the lesson.

### The Different Kinds of Evaluation and the Relationship with Purpose



Start this session by summarizing the different kinds of evaluation that teachers should know about, most importantly *Summative* and *Formative* Evaluation. Once again point out that the most important way to define a kind of evaluation is by its stated ‘purpose.’ It might also be useful to point out that in English, the name given to different kinds of evaluation often suggests its purpose. For example:

- **Summative:** Based on the word ‘sum,’ as in summing something up.
- **Formative:** Based on the word ‘form,’ such as forming a plan based on information.



It is useful to point this out because the Khmer translation of these terms sometimes does not capture its ‘purpose,’ which is essential for understanding its definition.



After this introduction, the Facilitator should next pass out **Handout 5a:** ‘Types of Evaluation for the Classroom.’ Review as a large group some of the points raised above that are also covered in the handout. Be sure to also review additional kinds of Evaluation such as *Diagnostic* and *Strategic Evaluation*.

**Key Discussion Points:** The following are some important points to raise during the large group review led by the Facilitator:



- *Diagnostic* and *Formative Evaluation* are similar but the main difference is that *Diagnostic Evaluation* is always done ‘before’ teaching while *Formative* evaluation is mainly used during or after teaching.
- The most common example of *Strategic Evaluation* is for placement or admission decisions.
- Both *Summative* and *Formative Evaluation* require the development of what are known as Tables of Specification to systematically plan the tests. However, *Summative Tests* tend to cover more content and a wider range



of skills. Summative tests tend to ‘sample’ the content taught while Formative tests tend to cover all the topics taught because there are usually fewer topics.

- Summative tests require questions of moderate difficulty while Formative tests should focus on questions that are easier.
- In general, Summative Tests use product-oriented questions (e.g., multiple choice questions) while Formative Tests use process-oriented questions so that it is easier for the teacher to analyze students’ mistakes and formulate remedial action.



After the Facilitator’s review of **Handout 5a**, ask participants to work in their small groups to complete the two exercises provided. One exercise deals with ‘Analyzing Students’ Mistakes’ while another requires participants to design a summative test. Give about one hour for this exercise.

### Analyzing Students’ Mistakes

No	Student Work That Indicates a Mistake	Your Analysis of WHY the Student Did This Wrong
1.	$\begin{array}{r} 1,300 \\ - 522 \\ \hline 878 \end{array}$	When subtracting the hundreds column, the student did not realize that it is also necessary to reduce the ‘3’ to ‘2’ so that the final answer should be 778.
2.	$\begin{array}{r} 140 \\ - 21 \\ \hline 120 \end{array}$	The student has misunderstood that 0 -1 in the ones column equals 0.
3.	$\begin{array}{r} 1,300 \\ - 522 \\ \hline 788 \end{array}$	The student has forgotten to subtract 1 from the tens column so that he/she is subtracting 2(0) from 9(0), giving the sum 778.
4.	$\begin{array}{r} 521 \\ + 888 \\ \hline 13,109 \end{array}$	The student is expressing the answer as separate sums from right to left: 1+8=9; 2+8=10; 5+8=13
5.	$\begin{array}{r} 51 \\ + 49 \\ \hline 90 \end{array}$	The student can add individual columns of numbers correctly but does not understand how to carry over units from one column to the next.
6.	$\begin{array}{l} 9 > 11 \\ 10 < 9 \end{array}$	The student has misunderstood that the sign > should always point to the smaller number.
7.	$1 + 2 + 3 = 123$	The student has misunderstood an addition sentence to mean that one simply expresses the answer as a consecutive list of the addends.
8.	recieve niegbor	The student does not understand the spelling rule: ‘i’ before ‘e’ except after ‘c’ and sometimes when followed by ‘g.’
9.	ផ្ទះ ភូមិ	The student has placed adjective before the noun instead of after as is the rule in Khmer.
10.	កី ៃ កើត ភី=ត ក	<p>a) The student has written the letters backwards.</p> <p>b) The student has expressed the vowel incorrectly.</p> <p>c) The student is not able to distinguish the difference between these two letters.</p> <p>d) The student has forgotten to place a ‘hat’ over the letter kaw.</p>

*Exercise: Designing a Summative Test*

Parameter	Suggested Design Feature	Parameter	Suggested Design Feature
<b>Scoring Procedures</b>	<ul style="list-style-type: none"> <li>• Criterion or Norm Referenced</li> </ul>	<b>Timing</b>	<ul style="list-style-type: none"> <li>• Terminal</li> </ul>
<b>Test Length</b>	<ul style="list-style-type: none"> <li>• Medium to long</li> </ul>	<b>Kinds of Questions</b>	<ul style="list-style-type: none"> <li>• Product-oriented</li> <li>• Objective Questions</li> </ul>
<b>Question Difficulty</b>	<ul style="list-style-type: none"> <li>• Moderate</li> <li>• Some Difficult Questions</li> </ul>	<b>Evaluation Strategies</b>	<ul style="list-style-type: none"> <li>• Written Test</li> </ul>
<b>Content Coverage</b>	<ul style="list-style-type: none"> <li>• Multiple Units</li> <li>• Sampling Content</li> </ul>		



When participants have completed the exercises provided in the Handout, discuss together as a large group. The Facilitator may use the answers provided above to help guide the review of participants' work.

**Summarizing the Differences between Summative and Formative Evaluation**



When the Facilitator has completed reviewing the exercises above, pass out **Handout 5b: 'Summary of the Differences between Summative & Formative Evaluation'** to review together the differences between Formative and Summative Evaluation in the matrix provided. This is a good opportunity to answer any questions or queries that participants might have about Summative and Formative Evaluation.



### 3.3 Useful Frameworks for Interpreting Test Scores: Normative and Criterion-referenced Testing



**Lesson Time:** 3 Hours and 30 Minutes



#### **Trainer Preparation:**

- Write up the Learning Outcomes of the lesson on a sheet of poster paper to introduce the lesson.
- Make copies of all Handouts as listed below.



#### **Resources/Materials:**

- Poster paper, marker pens
- Poster sheet summarizing the learning outcomes of the lesson.
- **Handout 6:** *Useful Frameworks for Interpreting Test Scores*
- **Handout 7:** *Criterion-Referenced Scoring*
- **Handout 8:** *Norm-Referenced Scoring*
- **Handout 9:** *Comparison between Norm and Criterion-Referenced Scoring*
- **Handout 10:** *Checking Your Knowledge on Evaluation Principles*



#### **Learning Outcomes**

1. Participants can *define* the meaning of Criterion- and Norm-Referenced Scoring in terms of when it is used and for what purpose.
2. Participants can *use* different scoring frameworks (i.e., criterion and norm-referenced) to interpret the meaning of a test score.
3. Participants can *determine* the situations in which to use criterion and norm-referenced scoring frameworks.
4. Participants can *answer* questions about marking on a 'curve' and how students' marks might be transformed based on a normative scoring framework.
5. Participants can *create* their own curves using the guidelines about Norm-referenced scoring discussed.
6. Participants can *explain* the differences between Criterion and Norm-Referenced Scoring according to different parameters (e.g., Purpose, Evaluation Strategies, Timing, etc.).
7. Participants can successfully *complete* exercises reviewing the evaluation principles discussed up to this point.

## Training Session Plan

### Outcomes of the Lesson



Place a sheet of poster paper up on the board that summarizes the learning outcomes for the lesson. Explain the outcomes and that these results exemplify what the participants should be able to do at the completion of the lesson.

### Introduction to Criterion and Norm-Referenced Measurement

Start this next section by asking the following question:



*Is the meaning of a test score a straightforward matter with little need for interpretation?*



After receiving some responses from participants, the Facilitator should hopefully be able to convince everyone that the answer to the above question is ‘No.’ All test scores are open to interpretation depending on the framework that we are using to interpret them. Two of the most commonly used frameworks used for interpreting test scores are Criterion-referenced Frameworks and Norm-Referenced Frameworks.



Use the example provided below to explain how the same score on a test can be interpreted in different ways. Draw the table on the whiteboard, which presents two sets of identical scores but with totally different interpretations depending on the contextual framework. The set of scores on the left-hand side of the table represent a ‘Promotional Context’ in which the cut-off point for ‘passing’ is 5. In this context, ask participants which students are determined to have passed the test? (*Answer*: Phally, Sovan, Malis, Bun, and Sina). The context shown on the right-hand side of the table represents a ‘Selection Context,’ in which only the top two students can be chosen for a scholarship. In this context, which students have passed the test? (*Answer*: Phally and Sovan). This example should help to demonstrate the idea presented earlier that the same test score is open to different interpretations depending on the context.



#### *Example of Same Sets of Scores Having Different Interpretations*

Student Name	Promotion Decision Context	Student Selection Context
1. Phally	9 ←	9 ←
2. Sovan	8 ←	8 ←
3. Malis	7 ←	7
4. Bun	7 ←	7
5. Sina	5 ←	5
6. Heng	4	4
7. Rith	4	4
<b>Description of Context: →</b>	<ul style="list-style-type: none"> <li>Promotional context.</li> <li>The criterion for passing is ‘5’</li> <li>In this framework, 5 students pass</li> </ul>	<ul style="list-style-type: none"> <li>Selection context.</li> <li>Only the top 2 students can get a scholarship</li> </ul>



The Facilitator should conclude this introduction by distributing **Handout 6: ‘Useful Frameworks for Interpreting Test Scores.’** Quickly review the points

made earlier, emphasizing the the interpretation of test scores will change depending on the use of an *absolute criterion* or a *normative standard*.

### About Criterion-referenced Scoring



After the above introduction, distribute **Handout 7**: ‘Criterion-referenced Scoring.’ Review the handout in plenary session. Lead a guided discussion that covers the following points:



- Criterion-referenced scoring is the most commonly used marking framework and should be familiar to all teachers.
- Criteria are usually set arbitrarily. In this regard, the Facilitator should compare the different criteria used for pass-fail decision in different countries such as the United States and the UK.
- Ask participants how these criteria compare to the one used in Cambodia. Why do you think the criteria are different? Possible answers might include tradition, to combat grade inflation, it is not unreasonable to expect students to know 50% or more of what they learn as a minimum standard, etc.
- Criterion referenced scoring is generally used for purposes involving *Certification, Promotion, or Attainment of Minimum Learning Standards*.
- The use of objectives figures prominently in how criteria are applied because the criteria will indicate how many learning objectives have been achieved by a student.

### About Norm-referenced Scoring



To help the explanation of Norm-referenced Scoring, pass out **Handout 8**: ‘Norm-referenced Scoring.’ Review the handout in plenary session. Lead a guided discussion that covers the following points:



- Norm-referenced scoring shows a student’s standing in comparison with other students; it is not based on an absolute criterion as in Criterion-referenced scoring.
- Norm-referenced scoring can take a number of different forms. For example, if your purpose is ‘selection’ based on a quota of available seats (e.g., for a scholarship), then passing students are simply counted off from the top of the distribution. As in the example shown earlier, if 7 students applied for a scholarship but there are only enough resources for two studnets, then the two top scoring students would be selected.
- Another example of norm-referenced scoring involves marking students on a ‘curve.’ When marking on a curve, teachers re-assign grades to students using the median score as an important point of reference. Teachers find scores relating to particular percentiles (e.g., 90<sup>th</sup> percentile, 10<sup>th</sup> percentile, etc.) and then decide at what percentile they want students to pass. Look at the example of marking on a ‘curve’ provided in **Handout 8** and review it with participants. Be sure to explain the concept of a percentile.



After reviewing the example about creating a ‘curve,’ answer the *Questions for Discussion* at the end of the handout. The following are the answers to help the Facilitator with the resulting guided discussion:

### Answers to the Discussion Questions

**Q1:** 40 students

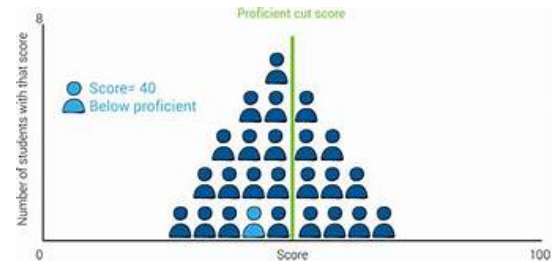
**Q2:** 50<sup>th</sup> Percentile = 4.5 in old distribution; 7 in new distribution; 20 students scored above this or 50% of the students.

**Q3:** 5 would have been the cut-off point for a Criterion-referenced Tests. 50% would have failed if this had been a Criterion-referenced Test.

**Q4:** Under the curve, only 4 students or 10% are failing (i.e., get F scores).

If the teacher uses the 5-point scale, then 8 students or 20% would fail, it depends on how the teacher wants the curve to work.

**Q5:** The teacher's purpose in making this curve was to ensure that more students would pass. In the original distribution, 50% of the students would have failed, which would have been too many.



### Comparison between Norm and Criterion-referenced Scoring

Review the differences and similarities between Norm- and Criterion-referenced Scoring by passing out **Handout 9:** 'Comparison between Norm and Criterion-referenced Scoring.' Review each parameter with participants and discuss how these two scoring frameworks are alike and different.

Pass out **Handout 10:** 'Checking Your Knowledge on Evaluation Principles' and ask participants to complete the questions individually. Provide about 10 to 15 minutes for this exercise. When all participants have completed the exercise, review the answers together using the key provided. Provide any necessary explanations for those participants who may have answered the questions incorrectly.



## 3.4 Characteristics of Good Tests and Examinations: Concepts of Validity



**Lesson Time:** 3 Hours and 30 Minutes



### **Trainer Preparation:**

- Write up the Learning Outcomes of the lesson on a sheet of poster paper to introduce the lesson.
- Make copies of all Handouts as listed below.



### **Resources/Materials:**

- Poster paper, marker pens
- Poster sheet summarizing the learning outcomes of the lesson.
- **Handout 11:** *Concepts of Assessment Validity*
- **Handout 12a:** *Kinds of Validity*
- **Handout 12b:** *Reinforcement Exercise on Concepts of Validity*
- **Handout 13:** *Some Final Guidelines to Ensure Validity in Your Tests*



### **Learning Outcomes**

1. Participants can define validity by giving examples of when a test is valid and when it is not.
2. Participants can explain how validity will affect the decisions that one makes based on the information generated by a test.
3. Participants can distinguish between the different types of validity (e.g., content validity, construct validity, etc.) when examining different case studies that exemplify them.
4. Participants can correctly answer questions about a test's validity when given a description of a particular assessment.
5. Participants can describe specific guidelines that one should follow to ensure that a student assessment is valid.

## Training Session Plan

### Outcomes of the Lesson



Place a sheet of poster paper up on the board that summarizes the learning outcomes for the lesson. Explain the outcomes and that these results exemplify what the participants should be able to do at the completion of the lesson.

### Defining Concepts of Validity



Begin this session by asking participants if they have ever heard of the concept of 'validity' before. Ask people what they think this concept means. Make a list of their answers on the whiteboard. This should help the Facilitator better understand participants' current knowledge and how quickly the facilitator can move through this session.



In trying to interpret participants' answers, the Facilitator should return to an important concept that was raised in **Session Plan 3.1: Measurement & Evaluation in Educational Decision-making**, which describes the link between 'Making Decisions' and the information that assessment generates. That is, we make Decisions based on Information. The Facilitator should remind participants that we can only make good decisions when the data that we receive is accurate. Then ask this question:

*How do we know if the data generated by an assessment is accurate?*



After giving participants some time to think about this question, the Facilitator should come back to the concept of 'validity'. That is, when a test is valid, it should be generating accurate data.



The Facilitator can now get to the crux of the matter about defining validity. Pass out **Handout 11: 'Concepts of Assessment Validity.'** In explaining this Handout, there are three important concepts that participants should understand in order to define validity. Write these concepts on the whiteboard in summary form as follows:



- *The test **Serves the Purpose** for which it was designed*
- *The test is **Relevant** to the information taught*
- *The test is **Reliable***



Next, go through the handout by highlighting the examples provided to further elaborate each of these concepts because they provide frequent mistakes that teachers often make, which undermines the validity of an assessment instrument. But if a teacher can be reasonably sure that a test serves its intended purpose, is relevant, and is reliable, then we can conclude that the test is valid and will enable us to make good educational decisions.



Ask participants if they have ever been in a situation where the mistakes described in the examples given were made? For example, have they ever given a very short test to assess many educational objectives or used a formative test for a summative purpose? Lead a discussion about why such mistakes undermine validity.

### Kinds of Validity

Following the above explanation about defining 'validity', the Facilitator should



next introduce the different kinds of validity that teachers may encounter when studying student evaluation. To better facilitate this explanation, pass out **Handout 12a: 'Kinds of Validity.'**

The Facilitator should point out to participants that there are basically four kinds of validity:

- **Content Validity:** The degree to which a test is 'relevant' to the content taught.
- **Construct Validity:** The degree to which a test is 'relevant' to the thinking skills that have been taught (e.g., remembering, understanding, etc.).
- **Concurrent Validity:** The degree to which a test is 'reliable' or consistent with other similar assessments.
- **Predictive Validity:** The degree to which an assessment can 'predict' an individual's potential of success.



After explaining these definitions, ask participants to review the **Case Study** provided in the Handout about Mr. Sophea's Test on a Mathematics Unit. Read the case study out loud to participants and ask them to consider the questions in small groups (without looking at the answers provided at the bottom of the page!).



- **Question 1:** Did Mr. Sophea's test have *Content Validity*?
- **Question 2:** Did Mr. Sophea's test have *Construct Validity*?
- **Question 3:** If Mr. Sophea administered a new test that covered all 4 lessons and included questions on Memory, Comprehension, and Application, do you think he would get a similar result? Why or why not? Be sure to phrase your answer in terms of Concurrent Validity.

Following the small group discussions, review the answers of each group as a large group. Discuss any discrepancies that may have arisen in the answers of each group and try to resolve any differences in the inverted answers provided at the bottom of the sheet.



Review the other Case Studies provided in the Handout such as the one on Predictive Validity and the case of Albert Einstein and the Case Study on Content Validity (which is optional). Discuss how these case studies affect specific kinds of validity such as Predictive Validity or Content Validity.



### **Reinforcement Activity**

Pass out **Handout 12b: 'Reinforcement Exercise on Concepts of Validity'** in order to check participants' understanding of the concepts presented above. In this Handout, there are 5 short case studies that exemplify issues in Assessment Validity. In their small groups, ask participants to read each case study and identify the kind of validity implied (there may be more than one) and to be prepared to justify their answers in plenary session.



### **CASE STUDIES**

1. Most conventional schools in the world tend to emphasize a limited number of intel-

ligences in their curricular programming such as ‘Verbal Intelligence’ and ‘Mathematical Intelligence.’ If you are someone who has high ‘Social Intelligence’ or high ‘Mechanical Intelligence’, you will probably not do well on most of the tests administered at such schools even though you might be very successful in jobs that require these kinds of skills. What sort of assessment validity is implied in this case study?

**Preferred Answer:** *This case study implies ‘Predictive Validity’ because it suggests that assessment in conventional schools would probably be a poor predictor of someone’s success in life if they have a kind of intelligence other than Verbal or Mathematical.*

2. Most people know that Jack Ma, the Chinese billionaire who created *Alibaba*, was an academic failure during his time at university. What sort of assessment validity is implied by the evaluation that he experienced while at school?

**Preferred Answer:** *This case study implies ‘Predictive Validity’ because it suggests that the assessment outcomes experienced by Jack Ma clearly were not predictive of his actual success in the business world.*

3. Some studies of educational achievement in Cambodia have found that there is often little relationship between the marks that students receive on their internal tests and the marks that they receive on externally administered examinations such as the Bac II Examination or the PISA test. What sort of validity is implied in this case study?

**Preferred Answer:** *This case study most clearly implies ‘Concurrent Validity’ because the assessment results from internal and external tests do not seem to correlate even though they pertain to be evaluating the same thing. This could be caused by such factors as ‘rien kua’ where students often buy their marks from teachers or other factors relating to the content and constructs assessed in which case Content and Construct Validity could also be implied.*

4. Many Cambodian teachers prefer to use open-ended questions and essay questions, which require a great deal of time to answer, when designing their tests, even when they have a great many lessons to evaluate. What sort of validity might this habit affect in terms of student assessment?

**Preferred Answer:** *This case study implies ‘Content Validity’ because it suggests that kinds of question formats used by teachers are not efficient in covering all or most of the content that was taught.*

5. Cambodian teachers receive a great deal of pressure from development partners to change their teaching in a way so that they are teaching more higher order thinking skills such as creativity and evaluation. They are encouraged to use new methods of teaching such as ‘project work’ and ‘problem-based learning’. However, when students are evaluated in their external examinations such as the Bac II Exam, they mainly encounter questions at the level of memory and understanding. What sort of validity is implied in this case study?

**Preferred Answer:** *This case study implies ‘Construct Validity’ because it suggests that external examinations are not evaluating the kinds of thinking skills that teachers may be teaching in their classrooms, especially if they are using project work or PBL methods.*



Give participants about 30 to 40 minutes for this exercise and then review as a large group. Preferred answers are provided above to help the Facilitator guide the discussion.

### **Guidelines to Ensure Validity When Designing Student Assessments**

Bring the session to a close by reviewing specific guidelines that teachers should







observe to ensure that their assessments have validity. The Facilitator should pass out **Handout 13**: ‘Some Final Guidelines to Ensure Validity in Your Tests’ to each participant and review each guideline in plenary session. Be sure to emphasize that the ‘conditions’ of testing as well as the ‘design’ of a test may have an impact on an assessment’s validity. Answer any questions that may arise when reviewing this Handout.



## Part 2: Practical Strategies for Test Design and Development

### 3.5 Instructional Objectives in the Educational Process

   	<p><b>Lesson Time:</b> 3 Hours and 30 Minutes</p> <p><b>Trainer Preparation:</b></p> <ul style="list-style-type: none"><li>• Write up the Learning Outcomes of the lesson on a sheet of poster paper to introduce the lesson.</li><li>• Make copies of all Handouts as listed below.</li></ul> <p><b>Resources/Materials:</b></p> <ul style="list-style-type: none"><li>• Poster paper, marker pens</li><li>• Poster sheet summarizing the learning outcomes of the lesson.</li><li>• <b>Handout 14:</b> <i>Why Instructional Objectives are Important in Assessment</i></li><li>• <b>Handout 15:</b> <i>The Role of Instructional Objectives in the Education Process</i></li><li>• <b>Handout 16:</b> <i>Defining an Instructional Objective</i></li><li>• <b>Handout 17:</b> <i>Kinds of Instructional Objectives</i></li><li>• <b>Handout 18a:</b> <i>Guidelines When Writing Instructional Objectives</i></li><li>• <b>Handout 18b:</b> <i>Using Bloom's Taxonomy to Develop and Classify Instructional Objectives.</i></li></ul> <p><b>Learning Outcomes</b></p> <ol style="list-style-type: none"><li>1. Participants can <i>explain</i> why it is important to use Instructional Objectives and what role they play in the educational process, particularly as this concerns its role in student assessment.</li><li>2. Participants can <i>identify</i> the parts of an Instructional Objective including its 'Content,' its 'Behavioral Construct,' and the 'Conditions' that sometimes form a part of the objective.</li><li>3. Participants can <i>explain</i> the difference between 'General' and 'Specific Objectives.'</li><li>4. Participants can <i>classify</i> the thinking skill level of Instructional Objectives using the framework outlined in Bloom's Taxonomy.</li><li>5. Participants can <i>write</i> 'good' objectives using specific guidelines that are provided in Handouts.</li></ol>
--	--

## Training Session Plan

### Outcomes of the Lesson

Place a sheet of poster paper up on the board that summarizes the learning outcomes for the lesson. Explain the outcomes and that these results exemplify what the participants should be able to do at the completion of the lesson.

### Why Educational Objectives Are Important in Education

Start this session by asking participants the following questions:

*Have you ever taken a long trip somewhere?*

*Have you ever first taken such a trip without first knowing what your destination is?*

*Does knowing the destination help you to prepare for the trip? How does it help you?*

These questions and participants' responses during discussion should help the Facilitator to set the stage for a comparison between the process of writing instructional objectives and the process of setting the destination for a long trip, the needed preparations for the trip, and determining what you have to do to get there.

### Writing Learning Objectives:



### Beginning With the End in Mind

**Key Idea:** The main idea here is that one has to know where one is going before one starts the preparations for a trip. In the same way, one has to know what one's learning outcomes are before one starts teaching or evaluating. Thus, one has to begin one's lesson planning with the end point in mind. Instructional Objectives enable us to do just that.

After completing the discussion and explanation above, pass out **Handout 14:** 'Why Instructional Objectives are Important in Assessment.' Review the first page of the Handout, emphasizing the key idea discussed above. At the end of the Handout, there is a short exercise about preparing for a trip. The Facilitator may ask each group to work on this exercise using the example of a trip across a desert and a mountain range. Each group should complete the table provided by indicating what they need to prepare and how they will use it along the way. Discuss the answers of each group and how such preparations might compare to preparing a lesson plan.

**Example:** Suppose one's objective was for children to know how to use a dictionary. What things might one need to prepare and teach children to achieve this objective? (e.g., providing pocket dictionaries to each student, teaching the order of letters in the alphabet, ordering lists of words according to their alphabetical sequence, etc.).

## The Role Educational Objectives Play in Education

Once the above discussion has been completed, ask participants how we use Instructional Objectives in the education system. Possible answers might include the following:

- To make our lesson plans.
- To know what materials we need for the lesson.
- To help guide the steps in our teaching.
- To help with making tests
- Etc.

Make a list of participants' responses on the whiteboard.

Next, the Facilitator should pass out **Handout 15**: 'The Role of Instructional Objectives in the Education Process.' Review the Handout with participants noting that there are three main roles or purposes of Instructional Objectives in education, namely:

- To Guide Evaluation
- To Plan Instruction Systematically
- To Provide Structure to Curriculum Development

As the Facilitator goes through these purposes, try to match them with the list that was made by participants earlier. Stress also that for purposes of this session, we are focusing on the uses of objectives for instructional and assessment purposes and not curriculum development, since most of those present in the workshop are teachers.

### Defining Instructional Objectives

The next step in this session is to review the actual definition of an Instructional Objective. For this purpose, pass out **Handout 16**: 'Defining an Instructional Objective.' Review the definition provided with participants, bringing attention to the idea that teachers should try to ensure that their Instructional Objective shows a 'visible behavior' at the end of instruction so that we can verify that one or more students has achieved the objective. Sometimes we call such objectives 'behavioral objectives.' When learning outcomes are observable, they are easier to evaluate.

#### **What is an Instructional Objective?**

*" . . . . A clear and unambiguous description of the goals or changes in the student's behavior that the teacher wishes to observe as a result of instruction."*



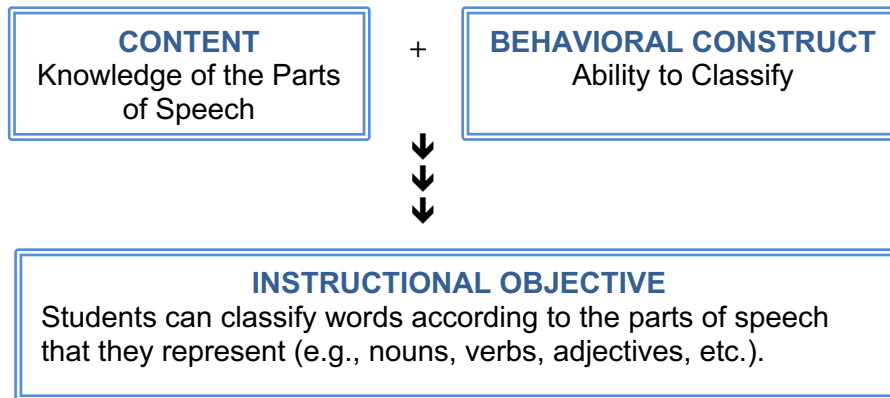
### Learning Objectives

Following the above explanation, the Facilitator should start to review the component parts of an objective. Review these component parts using the examples that are provided in the Handout.



### Parts of an Objective:

- An instructional should contain two or three components. These include the following:
  - The Content
  - The Behavioral Construct
  - The Condition(s) under which the learning should occur (OPTIONAL)



When explaining the use of a conditional statement in an objective, note that this part of an objective is ‘optional.’ Some educators prefer not to include them but others find that they can give an objective greater clarity for what the teacher has to do when teaching (e.g., distribute dictionaries, give maps to students to examine, do an experiment, etc.).

After your explanation, ask participants to try to complete the exercises provided at the end of **Handout 15**, which start with a conditional statement. Give participants about 20 to 30 minutes to do this in their groups and then share with the large group. Check for logic of the resulting Instructional Objectives and also whether they include a good statement of Content and Behavioral Construct.

### Kinds of Instructional Objectives

This next topic to discuss in this session refers to the different kinds of objectives. For this purpose, pass out **Handout 17**: ‘Kinds of Instructional Objectives.’ The Facilitator should explain that there are two kinds of objectives:

**General Objectives:** These are general statements of understanding followed by multiple statements of observable learning outcomes, also known as Specific Objectives.

**Specific Objectives:** These are statements of observable learning outcomes that are used with General Objectives but which may also be used on their own, especially in the context of a teacher’s lesson plan.

Explain to participants that General Objectives may use such general terms as ‘know’ and ‘understand’ because they are later clarified by Specific Objectives. On the other hand, Specific Objectives never use such terms because they must describe ‘observable’ behaviors. Review the examples provided in the Handout to make sure that this is clear.

The Facilitator should explain that educators are not always in agreement about



the use of ‘General’ and ‘Specific’ Objectives, but usually the former are used in formal curriculum documents while the latter are most often used in lesson planning by teachers.

### Guidelines to Use When Writing Instructional Objectives & Bloom’s Taxonomy



The next topic in this session concerns useful guidelines to write effective Instructional Objectives as well as the use of Bloom’s Taxonomy to ensure that a teacher has a good mix of objectives that cover both higher and lower order thinking skills. Start the session by distributing **Handout 18a**: ‘Guidelines to Use When Writing Instructional Objectives.’ There are five important guidelines in this Handout. The Facilitator should review each with participants using the examples provided.



#### Instructional Objective Writing Guidelines

1. Always use verbs that demonstrate observable behaviors when writing ‘specific objectives.’ (e.g., determine, explain, summarize, write, list, etc.)
2. Avoid using words like *strengthen*, *encourage*, *support*, etc. in your objectives as these do not suggest any ‘observable’ cognitive behavior.
3. Never put two behavioral constructs into one objective. (e.g., *students can ‘describe’ and ‘analyze’ the characters of a novel*).
4. Ensure that your objectives include a mix of both lower and higher order thinking skills. Use Bloom’s Taxonomy to help you know which objectives reflect higher order skills and which reflect lower order skills.
5. Classify each objective by the level that it implies in Bloom’s Taxonomy (e.g., Remembering, Understanding, Applying, etc).



The Facilitator should note that Guidelines 4 and 5 require some further explanation, especially if participants have never heard of Bloom’s Taxonomy. To ensure proper understanding of the uses of Bloom’s Taxonomy in writing Instructional Objectives, pass out **Handout 18b**: ‘Using Bloom’s Taxonomy to Develop and Classify Instructional Objectives.’ Next, use the Handout to review the various levels of the taxonomy and in particular the definitions provided for each.



Following the above explanation, ask participants to work in groups to complete the classification of useful verbs used to demonstrate a Behavioral Construct in an objective according to what level of Bloom’s Taxonomy that they exemplify. Be sure to note that one verb may sometimes be used with more than one level in the Taxonomy. Give participants about 30 minutes for this task. When all groups have completed the task, review and discuss as a large group. In order to help the Facilitator in this review process, the following answers are provided below. If participants’ answers happen to be different from the list provided, ask participants to justify their classification based on the answers provided in the table on



Bloom's Taxonomy shown earlier.

**Answers to the Exercise in Handout 18b**

English	Taxonomy Level	English	Taxonomy Level
Adapt	Analyzing Creating	Explain	<i>Understanding</i>
Answer	Remembering Understanding	Formulate	<i>Creating</i>
Analyze	Analyzing	Identify	<i>Understanding</i>
Apply	Applying	Illustrate	<i>Understanding Analysis</i>
Arrange	Understanding Analyzing	Indicate	<i>Remembering Understanding</i>
Calculate	Applying	Interpret	<i>Understanding Analyzing</i>
Categorize	Analyzing	Judge	<i>Evaluating</i>
Clarify	Understanding	Label	<i>Remembering Understanding</i>
Classify	Analyzing	List	<i>Remembering</i>
Combine	Analyzing Creating	Match	<i>Remembering Understanding</i>
Compare	Analyzing	Measure	<i>Applying</i>
Complete	Understanding Applying	Name	<i>Remembering</i>
Compose	<i>Creating</i>	Outline	<i>Understanding</i>
Contrast	<i>Analyzing</i>	Predict	<i>Applying Analyzing</i>
Create	<i>Creating</i>	Recite	<i>Remembering</i>
Critique	<i>Analyzing</i>	Rephrase	<i>Understanding</i>
Defend	<i>Evaluating</i>	Report	<i>Remembering Understanding</i>
Define	<i>Remembering Understanding</i>	Select	<i>Understanding</i>
Demonstrate	<i>Analyzing</i>	Solve	<i>Applying</i>
Describe	<i>Remembering</i>	Specify	<i>Understanding</i>
Determine	<i>Understanding Applying Analyzing</i>	State	<i>Remembering Understanding</i>
Diagram	<i>Understanding Applying</i>	Summarize	<i>Understanding</i>
Differentiate	<i>Analyzing</i>	Synthesize	<i>Creating</i>
Distinguish	<i>Understanding Analyzing</i>	Tell	<i>Remembering</i>
Enumerate	<i>Remembering</i>	Use	<i>Applying</i>
Evaluate	<i>Evaluating</i>		

Following this exercise, try another classification exercise in which participants must read a set of 10 Instructional Objectives and classify them according to the thinking skill that they demonstrate in Bloom's Taxonomy. To save time, this exercise can be done in plenary session. Ask participants to conceal the answers given at the bottom of the page so that they can use their own understanding to do the exercise. Discuss why some participants may have classified some Instructional Objectives differently and try to reach a consensus based on the Taxonomy Table reviewed earlier.





### Writing Instructional Objectives Using the Guidance Provided

The final task in this session requires participants to actually write their own objectives in their small groups. The final exercise for this purpose is provided at the end of **Handout 18b**. Pass out poster paper so that each group can write the objectives that they composed. The Facilitator should review the directions with participants so that they actively apply the Instructional Objective writing guidelines that were discussed earlier. The Facilitator should explain to participants that they may choose whatever ‘behavioral construct’ they want for the objective that they write, as long as it is logical.

#### Topical Areas for Instructional Objective Writing Exercise

- **English:** Vocabulary Words
- **Mathematics:** The idea of a set
- **Science:** The elements of the Periodic Table
- **Khmer:** The differences and similarities of characters in the story, Thum-Thiew
- **Geography:** Geographical maps
- **History:** The causes of World War II



Give participants about 45 minutes to complete this exercise, then review together as a large group, comparing the similarities and differences in the objectives that each group wrote.

## 3.6 Developing Tables of Specification



**Lesson Time:** 2 Hours



### **Trainer Preparation:**

- Write up the Learning Outcomes of the lesson on a sheet of poster paper to introduce the lesson.
- Make copies of all Handouts as listed below.



### **Resources/Materials:**

- Poster sheet summarizing the learning outcomes of the lesson.
- **Handout 19:** *Developing Tables of Specification*



### **Learning Outcomes**

1. Participants can *explain* what a Table of Specifications is and why we should use them in developing tests.
2. Participants can *answer* questions about the design of a test by looking at the Table of Specifications used to plan the test.
3. Participants can *answer* questions about how a Table of Specifications can strengthen a test's validity by examining case studies about poor test design.
4. Participants can *use* Tables of Specifications to make tests for their own students.

## Training Session Plan

### Outcomes of the Lesson



Place a sheet of poster paper up on the board that summarizes the learning outcomes for the lesson. Explain the outcomes and that these results exemplify what the participants should be able to do at the completion of the lesson.



### What is a Table of Specifications?

The Facilitator should start this session with some questions about how teachers design their tests.



*How do you currently make tests for your students?*

*How confident are you that the way you make tests ensures high levels of validity?*

*Have you ever heard of a Table of Specifications before?*



Such questions will help the Facilitator to know whether participants have ever heard of Tables of Specification before or whether they have actually used them. If participants have heard of Tables of Specification before and actually used them, then it will be easier to review this session quickly. If they have not heard of them before, then it may require the full amount of time put aside for this lesson to deliver the content.



Assuming that most participants have not heard of Table of Specifications, the Facilitator should start the session by passing out **Handout 19:** ‘Developing Tables of Specification.’ Review the definition of a Table of Specifications provided and why it can be useful for test development.



*A Table of Specifications is defined as a test blueprint that enables the development of a test, which is ‘balanced’ and ‘relevant’ in terms of the content and thinking skills taught.*

### The Format of a Table of Specifications



Next, the Facilitator should use the Handout to show an example of the format of a Table of Specifications, bringing attention to its provisions to list Content Areas (usually lessons or sub-units within a lesson) on the left hand side and thinking skills identified in Bloom’s Taxonomy along the top. Review the process of completing a Table of Specifications as indicated in the Handout:

#### How Does One Fill in a Table of Specifications?

- 1) Determine the content areas that you want to test. These can be lesson names or topics from one lesson.
- 2) Reflect on how you taught each content area. Did you ask students to simply remember things, understand things, apply principles, etc.?
- 3) Fill in the numbers in your table:
  - a. The numbers can be either the **number of questions** for each subject area/construct or the **number of points** on the test for each area.
  - b. If the numbers represent questions, all questions should be worth the same number of points.
- 4) Make totals for your content and constructs. The number of points or questions

that you have indicated for each topic and construct should reflect the amount of time and emphasis that you placed on each topic/construct when teaching. If it does not, then there is a problem with the test's Content or Construct Validity.



Following this explanation, the Facilitator should ask participants to read the three Case Studies provided in the Handouts as a large group. Then, ask participants to try to answer the questions at the end of each Case Study. Lead a guided discussion on the Case Studies using the sample answers provided below:

### Case Study Answers to Help Guide Discussion

**Case Study 1:** The test lacks *Content Validity* because each of the three lessons tested should have had the same number of questions or points. Instead, 80% of the questions or points were devoted to the last lesson; therefore, the test lacks overall validity. Using a Table of Specifications could have reminded the teacher to review her lesson plans, which would have shown her that there should have been an equal number of questions or points devoted to each lesson. These specifications could have, then, been easily built into her Table of Specifications.

**Case Study 2:** The test lacks *Construct Validity* because during the teaching process, students were asked to mainly remember facts and concepts and were not expected to demonstrate 'knowing' at the level of 'understanding' or 'application,' as required in the test. If the teacher had used a Table of Specifications, this would have reminded him to review the Instructional Objectives in his lesson plans at which time he would have realized that all of his objectives emphasized 'remembering' and not 'understanding' or 'application'. Teachers have to test what they teach in order for tests to be valid. If the teacher wanted the test to include 'knowing' at the level of 'understanding' and 'application,' then he should have addressed this need during the teaching process. Developing a Table of Specifications would have demonstrated clearly to the teacher that his test questions should be placed in the remembering column of the table only.

**Case Study 3:** Once again, the test lacks *Content Validity* because too many of the questions overemphasize one lesson at the expense of the others, which were all given an equal amount of time in teaching. As in Case Study 1, a Table of Specifications would have helped the teacher to see clearly that too many questions had been assigned to the last lesson of the unit.

### Reading a Table of Specifications



The topic in this session focuses on how to read a Table of Specifications. For this purpose, the Facilitator should bring participants' attention to the example of a completed Table of Specifications for Khmer Language. Look at the table together and answer the questions together in a plenary session. The answers to the questions are provided below to help the Facilitator lead a guided discussion on reading a Table of Specifications.

### Questions for Discussion



- 1) How many content areas are there in this test?  
**Answer:** *There are 4 content areas tested as indicated in the first column of the table.*
- 2) How many skills are measured in this test?  
**Answer:** *There are 4 thinking skills that are tested including Remembering, Understanding, Applying, and Analyzing.*
- 3) Which content area has the most emphasis? The least emphasis?



**Answer:** The Content Area with the most emphasis is 'Grammar' because it has 20 questions assigned to it. On the other hand, Vocabulary, Poetry, and Composition have less emphasis than Grammar but are assigned an equal number of questions (see the Total Column on the far righthand side of the table). To be valid, this assignment of questions should also reflect the same amount of emphasis given to each topical area during teaching.

- 4) Which construct area has the most emphasis? The least emphasis?

**Answer:** The Skill Area with the most emphasis is 'Applying' because it has 20 questions assigned to it (see the Total Row at the very bottom of the table). On the other hand, 'Remembering' has the least number of questions assigned to it (only 5 questions). To be valid, this assignment of questions should also reflect the same thinking skills that were actually taught during class. The teacher should make sure that the thinking skill level indicated in his Instructional Objectives are properly reflected in the Table of Specifications.

- 5) If the numbers in the table represent the number of questions, how much is each question worth on a test with 100 points.

**Answer:** Assuming that each numeral represents the number of questions, the value of each question would be 2 points since 100 points in total divided by 50 questions is 2.



Following the completion of the above discussion, the Facilitator should ask participants if they have any further questions on Tables of Specifications and provide explanations accordingly.



## 3.7 Writing Objective Test Items



**Lesson Time:** 6 Hours



### Trainer Preparation:

- Write up the Learning Outcomes of the lesson on a sheet of poster paper to introduce the lesson.
- Make copies of all Handouts as listed below.



### Resources/Materials:

- Poster paper, marker pens
- Poster sheet summarizing the learning outcomes of the lesson.
- Textbooks for all grades and subjects
- **Handout 20:** *About Objective and Subjective Test Questions*
- **Handout 21:** *How to Write Objective Test Questions Effectively*



### Learning Outcomes

1. Participants can correctly *define* an ‘objective question.’
2. Participants can *identify* the key characteristics of ‘objective’ questions.
3. Participants can *explain* the difference between ‘objective’ and ‘subjective’ questions.
4. Participants can *explain* when it is most appropriate to use objective questions and when it is appropriate to use subjective questions.
5. Participants can *indicate* whether a specific question format (e.g., extended essay, multiple choice, etc.) is subjective or objective in nature.
6. Participants can *identify* the advantages and disadvantages of specific kinds of objective questions (e.g., True-False, Multiple Choice, Matching, etc.).
7. Participants can *indicate* the various thinking skills in Bloom’s Taxonomy that are associated with each objective question.
8. Participants can effectively *write* objective questions including Short Answer, Matching, True-False, Multiple Choice, and Classification Questions using the guidelines that have been provided and discussed.

## Training Session Plan

### Outcomes of the Lesson



Place a sheet of poster paper up on the board that summarizes the learning outcomes for the lesson. Explain the outcomes and that these results exemplify what the participants should be able to do at the completion of the lesson.



### Why We Need a New Terminology to Describe Question Types That Moves Beyond 'Open' and 'Closed' Questions



The Facilitator should start this session by reviewing some common terminologies about describing question-types in Cambodia. The most commonly used typology in this regard is what are called 'open' and 'closed' questions. For purposes of this session, we will avoid using this terminology because it misses a key element in how assessment specialists distinguish between question types – this refers to the way that questions are scored. The Facilitator should bring participants' attention to how teachers score questions by presenting the following two question examples:

- *The Angkorian Period in Cambodia began under what King? \_\_\_\_\_*
- *Which Angkorian King was the greatest king of them all? Be sure to justify your answer with examples and logical arguments.*

---

---

---

---



Write each of these questions on the board for the entire group to consider. Then, the Facilitator should ask them to respond to the following questions:



*How many answers are possible for the first question?*

*How many answers are possible for the second question?*

*Would you consider both of these questions to be open-ended questions? Why or Why not?*

*How would a teacher assign points to each student response when correcting these questions?*



The Facilitator should lead a discussion about the three questions above. When leading the discussion, the Facilitator should be sure to make the following points:

### Guided Discussion Points

- It is clear that the first question about the founder of the Angkorian Period has only ONE possible answer: Jayavaraman II. Any other response to this question would be wrong.
- The second question could have MULTIPLE answers that are all correct depending on how well the student justifies his selection of a great king.
- According to the definition of an open-ended question, both questions would appear to be open ended even though the way that we score each question seems to be quite different. The first question would be scored 'dichotomously,' which means that it is either completely correct (full points) or completely wrong (zero

points). There is no grey area. On the other hand, the second question could involve the award of points that are full, partial, or zero depending on how well the student answers the question. This is quite different from the first question.

- The fact that both questions are considered to be ‘open-ended’ even though the way that we score them is quite different suggests that using this terminology (i.e., open-ended, closed) for describing questions is not very useful.

### What is the Difference between Objective and Subjective Questions?

At the end of the above discussion, the Facilitator is now ready to pass out **Handout 20: ‘About Objective & Subjective Test Questions.’** The Facilitator should use this Handout to introduce a much better typology for describing question types that includes the following two kinds of questions:

- Objective Questions
- Subjective Questions

This typology of describing question types is most commonly used in other countries and emphasizes the differences in how an educator scores a given question. If a question has only one possible answer and we score it dichotomously (i.e., the student receives a full award of points or zero), then the question is considered to be ‘objective.’ There is only one correct answer and this will not change regardless of who corrects the question. If a question has multiple possible answers and the award of points varies among those correcting the question, then the question is considered to be ‘subjective.’ Thus, the way that we define a question type depends heavily on how the question is scored and whether there is any variation in how points are awarded by those scoring the question.

### Kinds of Objective and Subjective Questions

The Facilitator should be sure to bring participants’ attention to how we classify commonly used question formats in terms of being either ‘objective’ or ‘subjective.’ This is noted in the Handout. Review why such questions as Multiple Choice, True-False, Matching, etc. are considered to be objective questions (**Answer:** because there is always only ONE correct answer to such questions and that they are scored dichotomously).

Essay Questions on the other hand are considered to be ‘subjective’ because answers may vary from student to student and point awards may be either full, partial, or zero. The Facilitator should explain that we will consider how to write good Essay Questions in another section but for now we will first consider how to write effective objective questions.

### Characteristics of Objective Questions

The Facilitator should start the next topic of this session by passing out **Handout 21: ‘How to Write Objective Questions Effectively.’** The Facilitator should note that this is a long and very important Handout that provides guidelines for writing effective objective questions in each of the various formats that they may take (e.g., Short Answer, True-False, Multiple Choice, etc.). But before getting into these guidelines, the Facilitator should review the key point that Objective Questions can only have ONE answer and if this is not the case, then it means

that the teacher has either written the question in a faulty way or that the question is not objective in terms of its classification. Review the definition of an Objective Question provided in the Handout:



Next, the Facilitator should note the **advantages** and **disadvantages** of Objective Questions. For example, their ability to be answered quickly means that they can be used to cover a great deal of content in a short period of time. This makes them ideal for the development of Summative Tests in which there is usually a great deal of content to be covered.

### ***What is an Objective Question?***

*An Objective Question is a kind of question for which there is only ONE correct answer. When scored, the question must be evaluated as either completely right or completely wrong.*

Review and discuss the other advantages and disadvantages of Objective Questions provided in the Handout:



#### **Format of Objective Questions**

- Such questions have multiple formats including Fill-in the blank, Matching, True-False, and Multiple Choice.

#### **Scoring Objective Questions**

- Scoring is done 'dichotomously', meaning that objective questions are marked completely right or completely wrong.
- There is only ONE Correct Answer

#### **Advantages**

- Effective for covering a lot of content
- Students can answer quickly leading to time efficiencies
- Easy to correct with high inter-rater reliability
- Can be used to measure constructs at the level of Memory, Understanding, Application, and Analysis

#### **Disadvantages**

- Cannot be used to measure creativity or synthesis level skills
- Cannot be used to measure process skills
- Difficult to write

### **Important Guidelines to Effectively Write Objective Questions of Different Formats**



The remaining content in this session focuses exclusively on improving participants' understanding of how to write objective questions of different formats (e.g., Multiple Choice, True-False, etc.) well. The Facilitator should note that participants are probably already aware of each of these different question formats but that teachers often write them in a faulty manner. Indeed, it is important

to note in this regard that objective questions are much more difficult to write than Essay Questions. Thus, the guidelines to be discussed going forward will help to provide a structured discussion on effective question writing that covers the following points:

- **The Basic Structure of a Question Format:** Whether it is a Short Answer Question or Multiple Choice, the Handout gives a clear description of the basic structure of each kind of question format.
- **Subvarieties of the Question Format:** The Handout provides examples of the different varieties and sub-varieties that a particular question type may have.
- **When to Use it:** The Handout discusses in what contexts and for what sorts of content and thinking skills a particular question format is most appropriate (e.g., to assess memory level skills, higher order skills, etc.).
- **Strengths & Weaknesses:** The Handout also discusses the Advantages and Disadvantages of each question format (e.g., its ability to prevent guessing, how difficult it is to write, its flexibility, etc.).
- **Guidelines on Writing Specific Question Formats:** The most important discussion in the Handout will provide actual rules for writing objective questions of different formats. These writing guidelines are supplemented with actual examples, each of which exemplify a given guideline.



The Facilitator should present each of the question formats described in **Handout 21** using a question-and-answer type approach to better gauge how familiar with each question participants are. The Facilitator should place a particular focus on common mistakes in writing different objective questions using the examples provided. Ask participants if they can see some of the faulty characteristics in the examples provided before actually telling them what the problems are.

### Writing Exercises

When the Facilitator has completed the presentation of all question formats, pass out textbooks, poster paper, and marker pens to each small group. It is advisable to make sure that participants are organized into small groups according to their subject area of expertise. Using the textbooks, ask participants to choose any lesson from the textbooks and write one or more objectives for the lesson followed by questions that use three or more of the different question formats that were discussed earlier. Provide participants at least one hour to complete this activity. Participants should write their objectives and corresponding questions on poster paper so that it can be presented to the large group.



When participants have completed this exercise, ask them to tape their poster paper sheets on the walls around the classroom to present them to the large group. The Facilitator should help to critique the questions that participants have written in terms of the guidelines earlier provided. When critiquing participants' work, be sure to consider some of the following issues:



- Are the questions developed by participants able to measure the thinking skills indicated in the objectives? For example, if participants have developed an instructional objective at the application level, does the question



also assess students' ability to apply principles and/or concepts as well?

- Do the question formats used by participants show variety?
- Do the questions follow the writing guidelines explained earlier? For example:
  - Are the stem and responses of multiple-choice questions properly structured?
  - Do questions follow the structural format that was explained in the Handout?
  - Do questions include any 'clues' that might promote guessing behavior?
  - Do the responses in a question overlap with one another?
  - Are there always enough responses to minimize guessing behavior (e.g., Multiple Choice Questions should always have at least 4 responses; Matching Questions should have more responses than things to be matched, etc.).
  - Etc.

When the Facilitator has completed critiquing each group's work, provide some time for wrap-up and additional questions from participants about Objective Question types.

## 3.8 Writing Essay Questions



**Lesson Time:** 3 Hours



### Trainer Preparation:

- Write up the Learning Outcomes of the lesson on a sheet of poster paper to introduce the lesson.
- Make copies of all Handouts as listed below.



### Resources/Materials:

- Poster paper, marker pens
- Poster sheet summarizing the learning outcomes of the lesson.
- **Handout 22:** *Defining Subjective Test Questions and Understanding their Limitations*
- **Handout 23:** *The Use of Directing Words When Writing Essay Questions*
- **Handout 24:** *Challenges in Scoring Essay Questions*
- **Handout 25:** *Overview of Essay Questions*



### Learning Outcomes

1. Participants can correctly *define* what a Subjective Question is.
2. Participants can *explain* the difference between Extended Response Essay Questions and Restricted Response Essay Questions.
3. Participants can *explain* the limitations of using essay questions to assess students' understanding in terms of multiple factors including their ability to sample content, the amount of time needed to answer essay questions, sampling reliability, and the use of directing words.
4. Participants are able to *use* Directing Words effectively when writing essay questions.
5. Participants can *explain* some of the key challenges that occur when scoring essay questions in terms of issues of reliability.
6. Participants are able to *explain* what the Halo Effect is and how this plays a role in scoring essay questions.
7. Participants are able to *explain* different methods for scoring essay questions including the 'Analytical Scoring Method' and the 'Global (or Holistic) Scoring Method.'

## Training Session Plan

### Outcomes of the Lesson



Place a sheet of poster paper up on the board that summarizes the learning outcomes for the lesson. Explain the outcomes and that these results exemplify what the participants should be able to do at the completion of the lesson.



### Characteristics of Subjective Questions



The Facilitator should start this session by reflecting once again on the differences between objective questions and subjective questions, also called Essay Questions. The Facilitator may summarize these differences by drawing the following table on the whiteboard:



Factor	Objective Questions	Subjective Questions
Scoring	Dichotomous	Scores may vary depending on the scorer
Response Generation	Students choose pre-determined responses	Students must generate responses that comprise sentences or paragraphs



The Facilitator may either try to elicit responses from participants when completing this table or simply explain the table at one time. The important thing to note is that Objective and Subjective Questions are mainly distinguished by the way that they are scored and how students generate responses, as noted above. Be sure that these two key differences are well understood.



After providing the above explanation, the Facilitator is now ready to distribute **Handout 22:** 'Defining Subjective Questions and Understanding their Limitations.' Review the definition of a Subjective Question in the Handout and link this definition with the table that was drawn on the board above.



### *What is a Subjective (Essay) Question?*

*Subjective Questions generally take the form of what is known in English as an Essay Question. An Essay is a 'free response' question that requires students to produce a written response in sentence or paragraph form, rather than to select the correct response from a number of alternatives or to generate a short word or phrase. Subjective Questions are not scored dichotomously and point awards for a student's response may vary from scorer to scorer.*

### Different Kinds of Subjective Questions

After reviewing the basic definition of a Subjective Question, the Facilitator is now ready to help participants distinguish between the two basic types of Subjective Questions noted on the Handout:

- ***Extended Response Essay Questions***
- ***Restricted Response Essay Questions***



The Facilitator should explain each of these two types of Subjective Questions using the explanation provided on **Handout 22**. In order to summarize the differences between these two types of questions, draw the following table on the whiteboard and ask participants to help complete it based on the explanation provided:

*(Complete the table below in plenary)*

	<b>Extended Response Question</b>	<b>Restricted Response Question</b>
Advantages	<ul style="list-style-type: none"> <li>Allows students to show their creativity</li> <li>Useful for Instructional Objectives at the level of Evaluation &amp; Creating</li> </ul>	<ul style="list-style-type: none"> <li>Useful for Instructional Objectives at the level of Remembering, Understanding, Applying, and Analysis.</li> <li>Easier to score than Extended Response Essays</li> <li>Has higher Reliability</li> </ul>
Disadvantages	<ul style="list-style-type: none"> <li>Difficult to score</li> <li>Low Reliability</li> </ul>	<ul style="list-style-type: none"> <li>Not useful for measuring thinking skills at the level of Evaluation &amp; Creating</li> </ul>

The Facilitator should fill in the table with the large group using the above example as a guide for the ensuing discussion.

### **The Limitations of Subjective Questions and Why They Require Caution When Used**

The Facilitator should start this next topic with a question:

*Many teachers like using Essay Questions for their end of semester tests. Do you think this is a good idea? Why or Why not?*

The Facilitator should lead a large group discussion to answer this question. Use the discussion guide below to help lead the discussion:

#### **Guided Discussion Points**

- It is the opinion of most assessment specialists that exclusive use of Essay Questions on summative tests is not good practice because they have *limited content sampling ability* and *low reliability*.
- One important reason for the problem of *limited sampling coverage of curricular content* is that Essay Questions take a long time to answer. Thus, only a few questions can be asked in the limited time available.
- Essay Questions may be used on summative tests but should not represent more than 20% of the total points on the test as a general rule of thumb.
- Limit the use of Essay Questions to higher order thinking skills such as Evaluating and Creating because objective questions cannot assess these skills.

After the large group discussion using the guided discussion points above, the Facilitator should next review the limitations of Subjective Questions, mainly

Essays, that are described in **Handout 22**. These points should echo many of the same points that were made in the discussion above.

### The Importance of Directing Words in Writing Essay Questions

The Facilitator should next introduce the topic of ‘Directing Words’ by passing out **Handout 23**: ‘The Use of Directing Words When Writing Essay Questions.’ Directing Words refer to the key verbs in the Essay Question that tell students what they should include in their answers. Review the guidelines for choosing Directing Words and ensuring that students understand the meaning of these words, as this is explained in **Handout 23**. In particular, the Facilitator should note that Directing Words directly correspond with the various levels of thinking skill as described in Bloom’s Taxonomy. Bring special attention to commonly worded questions at various levels of Bloom’s Taxonomy, as shown in the table below:

*The Chart below illustrates some appropriate Directing Words for different levels of thinking in Bloom’s Taxonomy. . .*

Remembering	Understanding	Applying	Analyzing	Evaluating	Creating
-What . . . -Where . . . -When . . . -Who . . . -Define . . . -Outline . . . -State . . .	-Why . . . -How . . . -State in your own words -Show . . . - Demonstrate -Summarize	-Apply . . . -What would happen if . . . -What best illustrates . . . -Explain how X would react to . . . -Illustrate . . . -Prove . . . -Demonstrate . . .	-What relationship exists between . . . -Identify the main idea . . . -Analyze . . . -Distinguish -Examine . . .	-Draw conclusions . . . -Defend the idea that . . . -Evaluate . . . -Compare . . .	-Propose an alternative to . . . -Devise . . . -How else would you . . .

After reviewing the topic of Directing Words, ask participants to try to develop Essay Questions at each level of Bloom’s Taxonomy as these relate to the famous story, *Tum Tiew*. This exercise is explained at the end of **Handout 23**. Pass out poster paper and marker pens to each small group for this purpose. Give participants about 30 minutes for this exercise. When they are done, ask each group to present their questions to the whole group. The Facilitator should help to critique each presentation according to the guidelines presented.

### Challenges in Scoring Essay Questions

The final topic in this session concerns scoring Essay Questions. The Facilitator should start this section by noting to participants that while Essay Questions are easier to write than objective questions, they are much more difficult to score. This is the opposite of objective questions, which are difficult to write but easy to score. To better explain the challenges of scoring Essay Questions, the Facilitator should distribute **Handout 24**: ‘Challenges in Scoring Essay Questions.’ As noted in a previous Handout, Essay Questions have significant limitations in terms

of their reliability. The award of points for the same student response may vary between teachers and may even vary for the same teacher depending on when he or she corrected the question. This problem is sometimes referred to as *inter-rater reliability*.



In order to demonstrate the problem of reliability, the Facilitator should do the exercise provided in **Handout 24**. In this exercise, the Faciliator asks participants to read a sample of a student response to an Essay Question to assign a score between 1 to 10, based on their own assessment. Give participants about 5 minutes to read and score the student’s response. Then, do a survey to find out the spread of scores using the table provided in the Handout. Indicate how many participants awarded 1 point, 2 points, etc using the table provided below:



When the Facilitator has completed the survey of participant score awards, ask participants to answer the questions provided in the Handout. Discuss these answers as a large group. Some of the key points to highlight in this discussion include the following:

Points Awarded	How Many Participants Gave This Score?
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

- Teachers vary from one to the other in awarding point scores to the same response.
- Scores tend to cluster towards the middle, because it is ‘safe’ to avoid an extremely high or low score.
- If clearer criteria had been provided for scoring this question, this might have lessened the amount of variation in scoring; however, even when providing clear criteria for scoring, it is likely that some variation will still occur. That is, we can never achieve the same level of reliability that exists for objective questions when compared to marking an Essay Question.

### Two General Approaches for Scoring Essay Questions



The Facilitator should next note that two methods of scoring Essay Questions exist to help teachers address problems relating to reliability. These methods are known as the following:

- *Analytical Method*
- *Global Method*



In order to effectively explain these scoring methods, the Facilitator should come back to **Handout 24**: ‘Challenges in Scoring Essay Questions.’ Explain that this Handout provides useful guidelines for correcting Essay Questions using either the Analytical or Global Methods. But the Facilitator should first emphasize once again that Essay Questions present significant difficulties for issues relating to Reliability as we already saw in the survey exercise that was just done earlier.

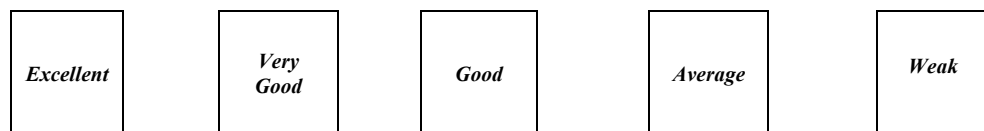
The Facilitator should review the techniques used in each scoring method using the Discussion Guide below:

### Analytical Method

- The teacher should write a model answer to the question. Alternatively, the teacher may also choose a student answer that is a model as well.
- The teacher should create a system that identifies specific criteria for giving points. The criteria identified by the teacher may have different weighting. Take the following example below for a question worth 10 points:
  - Logic and Organization of the Answer: 5 points
  - Effective Writing Style and Presentation: 3 points
  - Spelling and Grammar: 1 point
  - Neatness: 1 point
- Notice in that in the set of criteria described above, the teacher puts more weight on the ‘substance’ than the ‘form’ of the answer. Breaking down the scoring criteria in this way helps to strengthen reliability even though it creates quite a lot of extra work for the teacher.

### Global Method (Sometimes also called the Holistic Method)

- The teacher should create 4 or 5 spaces on their desk to organize papers from best to worst:



- Read a student’s paper and place it into one of the categories organized on the teacher’s desk as in the diagram above. Suppose the first student answer read by the teacher goes into the ‘Excellent’ space; but let’s further suppose that a second answer read by the teacher is even better than the first one. This may require the first paper to go into the ‘Very Good’ category while the second paper goes into the ‘Excellent’ space. The teacher may continuously adjust the classification of papers as he reads through students’ answers.
- This method can be very time consuming and may require re-reading students’ answers several times before a final judgement can be made.

After presenting this explanation, the Facilitator should take questions from participants and discuss any comments that they might have. During this discussion, the Facilitator should be sure to explain the problem of what is known as the ‘Halo Effect.’ This is a common problem in many areas of judgement, including the correction of Essay Questions. The most common example of the Halo Effect is when teachers give excessively high marks to students who have nice handwriting and present their answers in a very neat form, even though their answers may not have much substance. Another example of the Halo Effect is when students engage in what is called ‘bluffing,’ which occurs when students use flowery language in their answers to disguise the fact that they actually do not know the answer. The Halo Effect is not a problem in objective questions, which is one of their primary strengths. Nevertheless, teachers must always be on their guard against the Halo Effect when scoring Essay Questions.



### Overview of Essay Questions

Finally, the Facilitator is ready to summarize this session on Essay Questions. For this purpose, pass out **Handout 25: 'Overview of Essay Questions.'** This Handout provides a good summary of the key characteristics of Essay Questions including their strong and weak points. For example, Essay Questions are easy to write but difficult to score. They are excellent for measuring higher order thinking skills but because of the great difficulty involved in scoring them, they should be used sparingly in examinations. The Facilitator should close the session by taking any remaining questions from participants.

Factor	Characteristics
How easy to design?	<ul style="list-style-type: none"><li>• Essay Questions are relatively easy to write</li></ul>
Level of Thinking Measured?	<ul style="list-style-type: none"><li>• Restricted Response Essay Questions are appropriate for assessing Comprehension, Application, and Analysis.</li><li>• Extended Response Essay Questions are appropriate for Evaluation and Creation.</li></ul>
How Efficient to Cover Content?	<ul style="list-style-type: none"><li>• Because Essay Questions take a long time to answer, the amount of content that they can be used to cover is highly limited.</li></ul>
Impact of Guessing?	<ul style="list-style-type: none"><li>• Guessing is not an issue in Essay Question Responses. However, Essay Questions are susceptible to the Halo Effect when scoring.</li></ul>
Dangers of Irrelevant Clues that influence responding?	<ul style="list-style-type: none"><li>• Essay Questions rarely give irrelevant clues to students to help them to respond.</li></ul>
Scoring?	<ul style="list-style-type: none"><li>• Essays are very difficult and time-consuming to score and suffer from very low reliability.</li></ul>

## Part 3: Question Banking

### 3.9 Conducting Item Analysis and Constructing Question Banks



**Lesson Time:** 4 Hours



#### **Trainer Preparation:**

- Write up the Learning Outcomes of the lesson on a sheet of poster paper to introduce the lesson.
- Make copies of all Handouts and other forms as listed below.
- The trainer should bring one or more sets of examination papers from an actual test administration that occurred in someone's school. Specific questions from these tests will be analyzed by participants using Item Analysis procedures.



#### **Resources/Materials:**

- Poster paper, marker pens
- Poster sheet summarizing the learning outcomes of the lesson.
- Samples of actual examination papers answered by students
- Item Data Tabulation Sheet
- Item Analysis Score Card
- **Handout 26:** *About Item Analysis*



#### **Learning Outcomes**

1. Participants can *explain* the Definition, Purpose, and Advantages of using Item Analysis.
2. Participants can *define* Item Difficulty and Discrimination
3. Participants can calculate Indices of Difficulty and Discrimination using standardized formulae.
4. Participants can *organize* Item Analysis procedures including (i) the ranking and selection of student test papers; (ii) encoding data on standardized tabulation forms; (iii) determining Indices of Difficulty and Discrimination based on tabulated data; and (iv) transferring data to an Item Analysis Score Card.
5. Participants can *interpret* data generated by an Item Analysis in a way that enables them to improve the functioning of specific questions.
6. Participants can *explain* what a Question Bank is.

## Training Session Plan

### Outcomes of the Lesson



Place a sheet of poster paper up on the board that summarizes the learning outcomes for the lesson. Explain the outcomes and that these results exemplify what the participants should be able to do at the completion of the lesson.



### The Definition, Purpose, and Advantages of Item Analysis



The Facilitator should start this session with a series of questions about the tests that teachers give to their students such as the following:

- *When you give tests, do you see big fluctuations in the performance of your students from test to test?*
  - **Discussion Notes:** Test scores among students often fluctuate and this may be a result not of changes in learning outcomes but rather in the validity of the questions that have been developed.
  - **Discussion Notes:** Problems in the validity of questions may suggest that they are too easy or too difficult and do not discriminate well between the best performing students and those who are not performing well.
- *How confident are you that the questions that you put in your tests are functioning well?*
  - **Discussion Notes:** If teachers said they are confident in the functioning of their tests, what sort of evidence did they base this conclusion on? Such conclusions should be based on firm evidence from question analysis but the sad truth is that most teachers never analyze their questions after they have been administered; therefore, they really have no idea how well their tests are functioning.
- *Do you ever re-use questions on tests from year to year or do you just throw away the questions that you wrote and write new questions?*
  - **Discussion Notes:** Many teachers do not re-use questions because they are afraid that students may have seen old tests, which gives them unfair advantage. Sadly, throwing questions away from year to year means that teachers are losing a valuable educational asset because analyzed questions can be improved in terms of their validity, thereby raising the overall quality of testing instruments.



The Facilitator should lead a guided discussion on these questions using the discussion notes provided under each question above. The main point that the Facilitator wants to make after this discussion is that one's tests will only yield valid evaluations of student learning if the questions are functioning properly (e.g., not too hard, not too easy, aligned to the purpose for which they were designed, etc.). Unfortunately, teachers usually throw away used questions and start with new questions on the next test instead of analyzing how well they functioned and making improvements accordingly. Throwing away unanalyzed questions, therefore, results in the loss of a potentially valuable resource.



The Facilitator should next note that today's session seeks to help teachers analyze their test questions, make improvements in questions based on solid evidence, and improve the quality of classroom tests in the process. The best way to achieve this outcome is through a process called **Item Analysis**.



Pass out **Handout 26: 'About Item Analysis'** to all participants. Explain the first page of the Handout (Sheet 1) focusing on the Definition, Purpose, and Advantages of Item Analysis. To some degree, these issues may already have been understood as a result of the earlier discussion. The Facilitator should review these points again for reinforcement, as well as any outstanding points in the Handout not already covered.



### Information Provided by Item Analysis

The Facilitator should next move to the kind of information that Item Analysis can generate to help teachers know whether a question is functioning properly. For this purpose, ask participants to look at Sheet 2 of **Handout 26**.

Start the discussion by asking participants whether a question in which all students answered correctly (or incorrectly) is a good question? Clearly, when everyone is responding correctly or incorrectly, the question is operating at the extremes and probably not generating accurate information about students' understanding. The Facilitator should explain that, as we saw in a previous section of this manual, summative tests should contain questions of 'moderate' difficulty. For formative tests, however, we may want questions in the 'easy' range.

*But how do we know if a question is of easy, moderate, or high difficulty?*

The Facilitator should explain that Item Analysis will tell the teacher how easy or difficult a question is.

Moving to the next issue raised in the Handout, review what teachers think about a question in which the poorly performing students answer correctly while those students who are stronger performers answer incorrectly. In general, we expect that questions should be able to discriminate well between the high and low performers. The question described in the Handout on Sheet 2 is, therefore, a poorly performing question because it has low discrimination power. Once again, Item Analysis will tell a teacher how well a question can discriminate between high and low performing students.

In summary, the Facilitator should stress that Item Analysis will tell the teacher the **Difficulty Level** and the **Discriminating Power** of a question.

### Steps in Conducting Item Analysis

The Facilitator should next ask participants to turn to Sheet 3 of **Handout 26**. This sheet explains how to organize the Item Analysis process and tabulate data. For this purpose, the Facilitator should organize the participants into two or three large groups of perhaps 5 to 10 persons each.

Next, pass out three separate sets of test papers from three different classes. The test papers should be actually marked tests that have been completed by students and comprise mainly 'objective test questions.' Each test paper should have a clearly visible total mark at the top of the front page of the test that summarizes each student's total score. The number of test papers should be anywhere between 30 to 50 papers for each group.

Explain to each group that the first step in conducting an Item Analysis involves ranking the test papers from the highest to the lowest score. Once this has been done, the groups should create what is called an Upper Group and a Lower Group. Usually, 27.5% of the top scoring students are assigned to the Upper Group while the 27.5% of the lowest scoring students are assigned to the Lower Group.<sup>1</sup> For example, if a test group has 40 students in it, one would take the top 11 scoring students to form the Upper Group (i.e.,  $40 \times 27.5\% = 11$ ) and the 11

<sup>1</sup> Kelly, T. L. (1939). The Selection of Upper and Lower Groups for the Validation of Test Items. *Journal of Educational Psychology*. Vol. 30, p.p. 17-24.



lowest scoring students to form the Lower Group. The remaining student test papers in the middle are put aside and not used in the Item Analysis.



The Facilitator should next explain that participants are now ready to tabulate the test data. In order to tabulate test data, one has to use the Item Data Tabulation Form shown in **Handout 26**. To get some practice in using this form, the Facilitator should distribute additional tabulation forms to each group. For each question on a test, one has to use two tabulation forms. One form is used to tabulate the response patterns of the Upper Group and another form is used to tabulate the

**Item Data Tabulation Form**

Question No: \_\_\_\_\_      Group: \_\_\_\_\_      Question Type: \_\_\_\_\_  
 Date: \_\_\_\_\_      Size: \_\_\_\_\_      Grade Level: \_\_\_\_\_

No	a	b	c	d	e
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					

No	a	b	c	d	e
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					

No	a	b	c	d	e
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					

Tabulation Summary Area	
Response	Number
a	
b	
c	
d	
e	
Omit	

response patterns of the Lower Group. After completing the information required at the top of the form, one should next look at the response to Question 1 by the first student in the Upper Group. If the student chose Response 'a' for the question, one would check (✓) the box under Column 'a'. Then, go to the next student in the

group and see what response he/she chose for Question 1. If that student chose Response 'd,' then check the box under Column d for Row 2 and continue in this way until the responses of all students in the Upper Group to Question 1 have been recorded. Follow the same procedure for students in the Lower Group on a separate Tabulation Form and summarize the response patterns of both groups in the small box at the far-right hand side of the form. Finally, staple the two tabulation forms for Question 1 together so that they do not become separated.



Following this explanation, ask all participant groups to continue analyzing test questions for each question on the tests that they received, using two tabulation forms per question. For the information required by calculating Indices of Difficult and Discrimination, wait until the completion of the following section.

### Calculating Indices of Difficulty and Discrimination



The Facilitator should next start a discussion that explains how to calculate Indices of Difficulty and Discrimination. Start first with finding the **Difficulty Level** of a question. For this purpose, ask participants to look at Sheet 4A in **Handout 26**, where they can see the formula for calculating a question's difficulty level:

$$\text{Index of Difficulty} = \frac{N_u + N_l}{N_{u+l}}$$



The Facilitator should explain that this formula essentially means that one counts up the number of students answering a given question correctly in both the Upper

and Lower Group and then dividing by the total number of students in both groups. In the example provided, there are 8 students in the Upper Group and 8 in the Lower Groups for a total of 16 students. Five students in the Upper Group answered a question correctly whereas only 2 could do so in the Lower Group, so the total number of students answering correctly was '7.' Therefore:

$$7 \div 16 = 0.44 \times 100 = 44\%$$

The Index of Difficulty for this question is 44%. Using the interpretive scale provided in Sheet 4A, this means that the question is of 'moderate' difficulty.



Following this example, ask participants to calculate the Index of Difficulty for some of the questions that they have already analyzed and be prepared to interpret what the Difficulty Index that they calculate actually means, using the interpretive scale provided in Sheet 4A of the Handout.

The Facilitator should next move to showing participants how to calculate the Index of Discrimination. This explanation can be found in Sheet 4B of **Handout 26**. Remind participants that Discrimination refers to the ability of a question to demonstrate correct responding among high performing students and incorrect responding among more poorly performing students.

The Facilitator should continue to explain that the formula to find the Index of Discrimination is similar to the formula for Difficulty with some important differences. First, we subtract the number correct responders in the Lower Group from the Upper Group (instead of adding them together). Then we divide by the total number of students in a single group (instead of taking the combined number of students in both groups). Therefore, the formula reads as follows:

$$\text{Index of Discrimination} = \frac{N_u - N_l}{N_g}$$

Using the example from before, one would calculate the Index of Discrimination as follows:

$$(5 - 2) \div 8 = 3 \div 8 = 0.38$$

Using the interpretive scale provided in Sheet 4B, this means that the question has a moderate level of Discrimination, since 0.38 falls within the range of 0.20 to 0.39. Questions with moderate or high levels of discrimination are considered to be functioning properly but questions with low or even negative discrimination need to be revised or even thrown out completely.



Following this example, ask participants to calculate the Index of Discrimination for some of the questions that they have already analyzed and be prepared to interpret what the Discrimination Index that they calculate actually means, using the interpretive scale provided in Sheet 4B of the Handout.

### **Interpreting Your Findings from Item Analysis**

Once the Facilitator has completed the above explanations about Difficulty and Discrimination, the participants are now ready to go to the last step of Item Analysis – this refers to actually *interpreting* what the data from the analysis actually means.



In order to prepare participants for making interpretations about data generated by Item Analysis, bring their attention to the Item Analysis Score Card shown in

the Handout. Point out that one score card is used for one question. There are two tables shown in a score card because this helps to show the user the evolution of

a question after multiple administrations with respect to its functionality. Explain what information goes into each space of the card, as noted at the bottom of the figure to the left.

After providing the above explanation, review the examples given in Sheet 5 of **Handout 26**. Ask the questions provided and discuss participants' responses based on the examples given.

Following the review of examples described above, the Facilitator should ask participants to complete their own Item Analysis Score Cards based on the data they generated in their tabulation sheets. Ask participants to write in their data into an Item Score Card. Pass out additional Score Card sheets as needed.

**Item Analysis Score Card**

Section:		Difficulty Index:		Section:		Difficulty Index:	
Question No:		Discrimination Index:		Question No:		Discrimination Index:	
Item Stem:							
Distrac- tors	Upper Group	Lower Group	Total	Distrac- tors	Upper Group	Lower Group	Total
A				A			
B				B			
C				C			
D				D			
E				E			
Context (Picture/Diagram)				Context (Picture/Diagram)			

**Label Explanation**

**Section:** This refers to the section of the test from which the question comes (e.g., Section A, Section I, etc.)

**Question No:** The number of the question in the test or its sequence number in a Question Bank.

**Difficulty Index:** Based on the agreed formula

**Discrimination Index:** Based on the agreed formula

**Item Stem:** Reproduces the question here.

**Upper Group:** The number students in the Upper Group who chose each response (e.g., a, b, etc).

**Lower Group:** The number students in the Lower Group who chose each response.

**Total:** The number of students in a single group

**Content:** This space provides the opportunity for teachers to reproduce a question in its entirety including any accompanying diagrams, pictures, etc.

Then, ask participants to present their findings to the large group and indicate whether a question needs to be revised in any way. Be sure to ask participants why a question needs to be revised (e.g., too easy, too difficult, does not discriminate well between high and low performing students, etc.). Discuss each group's analysis in plenary.

### Question Banking

When the discussion has been completed, the Facilitator should point out that when we bring together multiple question analyses in one place, we have created what is called a Question Bank. The collected Item Analysis Score Cards form the content of the Question Bank. Questions in a Bank should be organized by topical areas so that they can be easily accessed. Be sure to urge participants to create their own Question Banks when they return to their schools. Such banks can be formed at the class level or for an entire school.

# PART 4: Principles & Guidelines for Effective Teacher Evaluation

## 3.10 Knowing Why You Are Evaluating a Teacher



**Lesson Time:** 2 Hours and 30 Minutes



### Trainer Preparation:

- Write up the Learning Outcomes of the lesson on a sheet of poster paper to introduce the lesson.
- Make copies of all handouts



### Resources/Materials:

- Poster paper, marker pens
- Poster sheet summarizing the learning outcomes of the lesson.
- **Handout 27:** *Knowing Why You Are Evaluating a Teacher*
- **Handout 28:** *Different Kinds of Teacher Evaluation*



### Learning Outcomes:

1. Participants can *explain* why we evaluate teachers in terms of two, broad, general purposes: Accountability and Teacher Growth.
2. Participants can *formulate* opinions about the effectiveness of traditional types of teacher evaluation that focus on accountability.
3. Participants can *explain* how Formative and Summative Teacher Evaluation are different.
4. Participants can *describe* the characteristics of a Dual Teacher Evaluation System and the degree to which such systems are effective.

## Training Session Plan

### Outcomes of the Lesson



Place a sheet of poster paper up on the board that summarizes the learning outcomes for the lesson. Explain the outcomes and that these results exemplify what the participants should be able to do at the completion of the lesson.

### Explaining the Reasons That We Evaluate Teachers

The Facilitator should begin this session by asking a simple question,



*“Why do we evaluate teachers?”*



Write this question on the white board for all participants to see. Try to brainstorm a list with participants on the blackboard about all the possible reasons that we do teacher evaluation. Possible answers might include the following:

- To assure that teachers are held accountable for their performance.
- To remove incompetent teachers.
- To recommend increases in teacher payment
- To make reports to supervisors about teacher performance.
- To extend teacher contracts.
- The Ministry requires that teachers are evaluated regularly.
- To help teachers to improve their practice (*if anyone suggests this purpose, write it on the board as well but keep it for later in the discussion*).



After making a list of possible responses, ask participants if they think that doing evaluation for most of the reasons cited (except for improving teachers’ practice) is likely to have any positive impact on teachers’ actual teaching practice?



After some short discussion, the Facilitator should conclude that the answer is probably ‘No.’ Explain further that most of the purposes for teacher evaluation in Cambodia focuses on monitoring teachers’ performance, but that it seems to have little effect on actual practice. This is something that teacher evaluation in Cambodia needs to address. This is not to say that the other purposes of evaluation cited are unimportant, only that they do not address what should be a key goal of teacher evaluation, i.e., improving teachers’ classroom practice.



Next, pass out **Handout 27: ‘Knowing Why You Are Evaluating a Teacher,’** and review some of the conclusions discussed above. In particular, make a note that educators often evaluate teachers mechanically without knowing clearly ‘why’ they are doing an evaluation or what they should do with the results. It is even more unlikely that the way teachers are evaluated will have any impact on their actual performance.



The Facilitator should next organize an exercise in which participants can think more deeply about some of the ideas raised above. Ask participants to look at the exercise on page 2 of Handout 27 (What Do YOU Think about Teacher Evaluation in Cambodia?) and respond to the following opinion:

*Most of the teacher evaluation that occurs in Cambodia seems to focus on ‘traditional’ evaluation purposes rather than helping teachers to improve their practice.*

Pass out poster paper to each group in the room and ask them to write their views about the opinion expressed above. Do they agree or disagree and why? Provide about 30 minutes for the small group discussions and then ask each group to place their report sheets on the board and compare opinions. Hopefully, most participants will see that the way teacher evaluation is carried out in Cambodia today probably has little effect on improving teachers’ practice.



### Different Kinds of Teacher Evaluation

The previous discussion about the purposes of Teacher Evaluation (i.e., Traditional Purposes focused on Accountability and Improving Teachers’ Practice) lead into the present subtopic on ‘Kinds of Teacher Evaluation.’ To begin the discussion, the Facilitator should pass out **Handout 28**: ‘Different Kinds of Teacher Evaluation’ to each participant. As we saw in the early part of this manual, the terms Formative and Summative Evaluation also apply to Teacher Evaluation. To help explain this, write the following matrix on the blackboard for all participants to see:

Evaluation Purpose	Kind of Evaluation
Accountability (Traditional Teacher Evaluation)	??
Improving Teachers’ Classroom Teaching	??

Next, ask what kind of evaluation they think goes with evaluation purposes related to ‘Accountability’? What kind goes with ‘Improving Teachers’ Classroom Practice’? Clearly, the answer is that the former goes with Summative Evaluation while the latter goes with Formative Evaluation. Complete the matrix like this:

Evaluation Purpose	Kind of Evaluation
Accountability (Traditional Teacher Evaluation)	Summative
Improving Teachers’ Classroom Teaching	Formative

Then, the Facilitator should start reviewing the matrix that defines the characteristics of Formative and Summative Teacher Evaluation in the table provided in Handout 28. The Facilitator should note the sharp differences between each kind of evaluation according to the parameters provided (e.g., Operational Focus, Timing, Common Instruments, etc.). These differences in focus should

once again underline that most teacher evaluation in Cambodia is Summative, not Formative.



Following this explanation, the Facilitator should introduce the concept of a ‘Dual Teacher Evaluation System.’ Explain that this is the kind of teacher evaluation system in a school or institution where there is one system that mixes elements of both Summative and Formative Evaluation. Handout 28 makes the following conclusion about Dual Teacher Evaluation Systems:

*Far from leading to growth and improvement, dual evaluation systems often demotivate teachers and create resistance and resentment of the assessment process.*



Participants may or may not agree with this conclusion (especially if they are administrators who are often expected to serve a dual function in their school where they are both a ‘policeman’ as well as someone who is supposed to mentor teachers). To deepen the discussion on the above view about Dual Teacher Evaluation Systems, pass out some poster paper to each group and ask them to answer the questions provided in Handout 28 on the paper provided. The name of this exercise is: ‘Does Cambodia Have a Dual Teacher Evaluation System?’ Give participants about 30 to 35 minutes to formulate their answers. Group members may have divergent viewpoints about this issue so make sure that each group represents all opinions.



When participants have completed their discussions, ask each group to report back on their responses to the questions provided. The Facilitator needs to be on his/her toes to challenge groups who believe that Dual Teacher Evaluation Systems work well. The literature demonstrates that they do not work well because teachers will not be receptive to changing their teaching practice when they feel threatened by the process, as is often the case in Dual Systems. This is the most important point that the Facilitator needs to make in this section to ensure that participants try to keep the processes of Formative and Summative Teacher Evaluation ‘separate’ upon their return to their schools and institutions.



## 3.11 What to Assess During Teacher Evaluation



**Lesson Time:** 2 Hours and 30 Minutes



### Trainer Preparation:

- Write up the Learning Outcomes of the lesson on a sheet of poster paper to introduce the lesson.
- Make copies of all handouts



### Resources/Materials:

- Poster paper, marker pens
- Poster sheet summarizing the learning outcomes of the lesson.
- **Handout 29a:** *Knowing What You Want to Evaluate When You Assess a Teacher*
- **Handout 29b:** *Variations That May Occur in What You Evaluate*



### Learning Outcomes:

1. Participants can *identify* appropriate subskills that further define key teaching skills such as Instructional Planning, Instructional Delivery, Classroom Management, etc.
2. Participants can *modify* lists of key teaching skills to come up with their own teacher assessment framework.
3. Participants can *explain* how the focus of an evaluation might vary in different contexts.
4. Participants can *identify* the challenges that might occur in setting the focus of an evaluation for teachers in different situations (e.g., new teacher, experienced teacher, etc.).
5. Participants can *determine* possible solutions that might effectively address the challenges that evaluators might encounter when evaluating teachers in different situations (e.g., new teachers, experienced teachers, etc.).



## Training Session Plan

### Outcomes of the Lesson



Place a sheet of poster paper up on the board that summarizes the learning outcomes for the lesson. Explain the outcomes and that these results exemplify what the participants should be able to do at the completion of the lesson.

### Knowing What to Assess During a Teacher Evaluation



The Facilitator should begin this session by stating what is obvious – Teacher Evaluators must know well the skills that teachers should possess, and they should themselves be proficient in using these skills (i.e., they should be experienced teachers). But even with this background, Teacher Evaluators need to be able to organize their knowledge of teacher skills in a coherent framework. There are already many teacher skill frameworks available in the research literature as well as those compiled by the Ministry of Education, Youth, and Sport. Such frameworks can help evaluators to focus in on specific skills that they want teachers to improve, depending on where the greatest needs are.



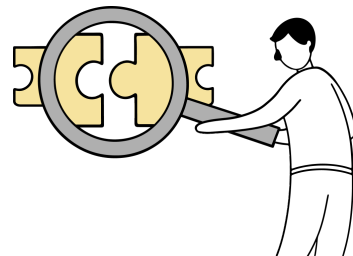
An example of a teacher skill framework is provided in **Handout 29a: 'Knowing What You Want to Evaluate When You Assess a Teacher.'** Pass out this handout to all participants and review each of the skills listed. Note that there are 8 Skill Areas (A to H) and a total of 25 discrete teaching skills that are identified.



Ask participants if there are any Skill Areas that could be added to this framework? Does this framework seem relevant to Cambodia? Are there any changes that they might want to see in the organization of these skills to make them more relevant to Cambodia? These kinds of questions will help participants to think about the organization of the framework more deeply as well as its relevance to their schools and institutions.<sup>2</sup>



Following the above discussion, participants are now ready to start thinking in more detail about specific subskills that would better define each of the 25 skill areas listed in the framework in Handout 29a. In their groups, ask participants to try to identify appropriate subskills that define in more detail selected teaching skills. An example is provided in Handout 29a for Instructional Delivery. After reviewing this example, ask participants to identify subskills for a different skill area in the framework using the Exercise format provided in Handout 29a. Pass out poster paper to each group so that they can define subskills for any two skill areas that they select. Give participants about 35 to 45 minutes to complete this exercise.



When participants have completed this exercise, ask each group to present

<sup>2</sup> For the reference of the Facilitator, a full framework is provided in **Annex 1** of this document to help with subsequent discussions on the identification of subskills.



their teaching skill analyses to the large group. The Facilitator should review what groups have done to identify any omissions. Use the full framework document provided in **Annex 1** of this document to help identify any possible omissions in the analyses that participants present.

**Variations That May Occur in Setting the Focus and Content of a Teacher Evaluation**



**Variations in Evaluation Content/Focus According to Purpose:** The Facilitator should next provide an explanation about how the content and focus of a Teacher Evaluation may vary depending on the purpose, setting, and who is being evaluated. Once again, the Kind of Evaluation (Formative or Summative) will have a big impact on variations in the focus of the Teacher Evaluation.

Draw the following matrix on the board to help demonstrate how variations in evaluation focus may occur:



Evaluation Type/Purpose	Flexibility in Setting Content & Focus
Formative Evaluation (Growth Purpose)	<ul style="list-style-type: none"> <li>Highly flexible</li> <li>Focus and content are usually set in agreement with the teacher</li> </ul>
Summative Evaluation (Accountability Purpose)	<ul style="list-style-type: none"> <li>Not so flexible</li> <li>Focus and content set by central office</li> </ul>



Explain that when a Teacher Evaluation is done for purposes of promoting ‘professional growth and improvement’ (i.e., Formative Evaluation), the content and focus of the assessment is usually set consensually with the teacher. This means that the teacher has a great deal of input into what skill areas they would like to improve or perhaps they would like to learn how to use a new teaching methodology. In the case of an ‘accountability’ purpose (i.e., Summative Evaluation), there is little flexibility in setting the content and focus of the evaluation because this is usually set by a central office.



Following the above explanation, pass out **Handout 29b:** ‘Variations That May Occur in What You Evaluate.’ Review again how a Teacher Evaluation might vary in focus and content depending on the purpose and context of the evaluation.



**Variations in Evaluation Content/Focus According to Teacher Background:** Next, the Facilitator should review that part of Handout 29b that discusses variations in setting the content/focus of an evaluation according to the teacher’s background. There are two types of teacher backgrounds to consider in this case. One refers to ‘new’ teachers while another refers to ‘experienced’ teachers. In the case of experienced teachers, they are likely to have much more input into the content of the evaluation that occurs, especially in the case of growth-focused assessments. In the case of new teachers, a more structured assessment may occur that focuses on more basic needs and teaching skills. Review the table below with participants to be clear on how variations in set-

ting the focus/content of an evaluation may vary by both purpose and teacher background. Allow for questions and discussion as needed.

**How to Determine the Focus of Your Evaluation by Teacher Background**

Teacher Status	Growth Purpose	Accountability Purpose
Experienced Teacher	<ul style="list-style-type: none"> <li>Review the Teacher's career path plan to determine areas of interest.</li> <li>Organize a conference with the teacher to determine areas of interest for growth.</li> <li>Review a school program that has set priorities for changes that it expects from teachers.</li> </ul>	<ul style="list-style-type: none"> <li>Observe the teacher using minimum competency lists of teacher skills.</li> <li>Review records of attendance, participation in school planning, interactions with parents and students, and other professional behaviors based on standardized sets of criteria.</li> </ul>
New Teacher	<ul style="list-style-type: none"> <li>Observe the teacher to determine areas of need.</li> <li>Conference teachers to determine areas of interest for growth.</li> </ul>	<ul style="list-style-type: none"> <li>Observe the teacher using minimum competency lists of teacher skills.</li> </ul>



Following the above explanation, pass out some poster paper to participants and ask them to complete the exercise at the end of Handout 29b. This exercise asks participants to think about some of the challenges that they might face in setting the focus/content of a teacher evaluation with new and experienced teachers and how they might resolve some of these challenges.

Teacher Status	Growth Purpose	Accountability Purpose
Experienced Teacher	<p>What are likely to be the challenges to setting the focus of a teacher evaluation?</p> <ul style="list-style-type: none"> <li><u>Teacher is not motivated.</u></li> <li>_____</li> <li>_____</li> </ul> <p>What could be the solutions?</p> <ul style="list-style-type: none"> <li><u>Peer Pressure.</u></li> <li>_____</li> <li>_____</li> </ul>	<p>What are likely to be the challenges to setting the focus of a teacher evaluation?</p> <ul style="list-style-type: none"> <li>_____</li> <li>_____</li> <li>_____</li> </ul> <p>What could be the solutions?</p> <ul style="list-style-type: none"> <li>_____</li> <li>_____</li> </ul>
New Teacher	<p>What are likely to be the challenges to setting the focus of a teacher evaluation?</p> <ul style="list-style-type: none"> <li>_____</li> <li>_____</li> <li>_____</li> <li>_____</li> </ul> <p>What could be the solutions?</p> <ul style="list-style-type: none"> <li>_____</li> <li>_____</li> </ul>	<p>What are likely to be the challenges to setting the focus of a teacher evaluation?</p> <ul style="list-style-type: none"> <li>_____</li> <li>_____</li> <li>_____</li> <li>_____</li> </ul> <p>What could be the solutions?</p> <ul style="list-style-type: none"> <li>_____</li> <li>_____</li> </ul>



Ask participants to complete the exercise on the poster paper provided and then present the challenges and solutions that they identified. Take about 30 to 40 minutes for this exercise. Review and compare the responses together as a large group.

## 3.12 Who Should Evaluate Teachers



**Lesson Time:** 2 Hours and 30 Minutes



### **Trainer Preparation:**

- Write up the Learning Outcomes of the lesson on a sheet of poster paper to introduce the lesson.
- Make copies of all handouts



### **Resources/Materials:**

- Poster paper, marker pens
- Poster sheet summarizing the learning outcomes of the lesson.
- **Handout 30:** *Who Should Do Teacher Evaluation*



### **Learning Outcomes:**

1. Participants can *explain* why the same personnel should not be conducting both Formative and Summative Teacher Evaluation.
2. Participants can *indicate* those who are the most suitable personnel in a school or institution to carry out Formative Teacher Evaluation and Summative Teacher Evaluation.
3. Participants can *identify* the challenges in achieving a suitable division of labor in a school or institution with regards to Teacher Evaluation.
4. Participants can *identify* appropriate measures to achieve a suitable division of labor in a school or institution with regards to Teacher Evaluation.

## Training Session Plan

### Outcomes of the Lesson



Place a sheet of poster paper up on the board that summarizes the learning outcomes for the lesson. Explain the outcomes and that these results exemplify what the participants should be able to do at the completion of the lesson.

### Characteristics of Teacher Evaluators Focused on Professional Growth



The Facilitator should start this session by asking the following question:

*What is/are the most important characteristics of Teacher Evaluator whose purpose is to promote professional growth?*

List some of the answers provided by participants on the board. Some likely responses might include:



- Trust
- Sensitive
- Good Communication Skills
- Good at building relationships
- Etc.



shutterstock.com · 250176199



All such responses are correct; however, the most important response is probably the ability to



build **‘trust.’** Start a discussion in which participants consider whether a school director or vice director can have a relationship with a teacher in which he/she will share their weaknesses and problems in teaching? For those who say ‘yes, it is possible’, be sure to ask whether participants are being naïve or not. Explain that it is quite normal for subordinates to usually hold back their thoughts on their weaknesses and problems because of a fear of the boss’ power over them. Thus, it is usually difficult for supervisors such as school directors to carry out Teacher Evaluation focused on professional growth.



Pass out **Handout 30:** ‘Who Should Do Teacher Evaluation.’ Review some of the ideas discussed above in the handout on pages 1-2. Emphasize that the best persons to do Teacher Evaluation whose purpose is ‘professional growth’ are Mentors or Technical Grade Leaders while those School Directors or Vice Directors are more suited for Teacher Evaluation whose purpose is ‘accountability.’ Draw the following matrix on the board to emphasize these points:



Evaluation Purpose	Personnel in the Best Position to Fulfill This Purpose
Professional Growth	<ul style="list-style-type: none"> <li>• Mentors</li> <li>• Technical Grade Leaders</li> </ul>
Accountability for Performance	<ul style="list-style-type: none"> <li>• School Directors</li> <li>• School Vice Directors</li> <li>• Department Heads</li> <li>• DOE/POE Staff</li> </ul>



The discussion above should come back again to the problem of Dual Evaluation Systems discussed earlier and what happens when the same person does both Formative and Summative Teacher Evaluation. Usually, such systems do not work very effectively for the reasons explained in the handout.

### The Feasibility of Separating Teacher Evaluation Functions



When the discussion above has been completed, ask participants to think about their own schools and institutions and the challenges that might be encountered in trying to create a suitable division of labor in terms of who is doing Teacher Evaluation. After identifying the challenges, what might be some of the solutions? In organizing this task, use the exercise provided in Handout 30. Pass out poster paper to each small group and ask participants to use the template provided to frame their discussions. Give participants about 40 minutes for this exercise. When participants have completed the exercise, ask each group to share their thoughts with the large group.



During group presentations, the Facilitator should compare group responses and give feedback. The following ideas and thoughts are provided below to help the Facilitator lead a guided discussion. These responses are simply suggested but not required for sharing with participants.

Suggested Ideas to Help with a Guided Discussion	
Challenges in Achieving a Suitable Division of Labor	Possible Solutions
<ul style="list-style-type: none"> <li>• Shortage of personnel</li> <li>• School personnel do not have enough time to do teacher evaluation</li> <li>• Teachers do not trust evaluators</li> <li>• Lack of suitable tools for evaluators to use to evaluate teachers</li> <li>• Teacher evaluation is not a high priority</li> <li>• Teachers do not cooperate with efforts to</li> </ul>	<ul style="list-style-type: none"> <li>• Create a position of mentor in a school or institution who can focus on professional growth issues.</li> <li>• Request additional staff from MoEYS or NIE.</li> <li>• Reduce the responsibilities of staff dedicated to teacher evaluation so that they have more time to focus on teacher evaluation.</li> <li>• Promise teachers that evaluation focused on professional growth will be kept confidential.</li> <li>• Evaluators should share confidential personal stories of their own problems and failures to build trust.</li> <li>• Prioritize teacher evaluation activities in the school improvement plan.</li> </ul>

### 3.13 Organizing Teacher Evaluation and Links with Staff Development



**Lesson Time:** 2 Hours and 30 Minutes



#### **Trainer Preparation:**

- Write up the Learning Outcomes of the lesson on a sheet of poster paper to introduce the lesson.
- Make copies of all handouts



#### **Resources/Materials:**

- Poster paper, marker pens
- Poster sheet summarizing the learning outcomes of the lesson.
- **Handout 31:** *Guidelines for Organizing Teacher Conferences*
- **Handout 32:** *Linking Your Teacher Evaluation System to Staff Development*



#### **Learning Outcomes:**

1. Participants can *explain* what should be included in a pre-conference/post-conference for Formative Teacher Evaluation.
2. Participants can *explain* what should be included in a pre-conference/post-conference for Summative Teacher Evaluation.
3. Participants can *do a role play* that exemplifies the differences between a pre-conference and a post-conference in different contexts (e.g., Formative Context, Summative Context).
4. Participants can *explain* how Teacher Evaluation and a Staff Development System can support one another.
5. Participants can *use* a concrete example in the handout to explain how Formative Teacher Evaluation and a school-based Staff Development Program are linked.

## Training Session Plan

### Outcomes of the Lesson



Place a sheet of poster paper up on the board that summarizes the learning outcomes for the lesson. Explain the outcomes and that these results exemplify what the participants should be able to do at the completion of the lesson.

### Teacher Conferences as a Tool to Better Plan Teacher Evaluation



The Facilitator might consider starting this session by asking participants how many of them have ever participated in an individualized teacher conference. Indicate the percentage of those who have ever participated in a conference on the board. Ask further how many participants have ever participated in an individualized conference ‘before’ the evaluation took place? Write the percentage of participants who have ever experienced a ‘pre-conference’ on the board and



compare the two numbers. In general, teacher conferences are rarely used in Cambodia to plan teacher evaluation. Often, teacher evaluations are planned without any input from the teacher though follow up conferences after an evaluation has already occurred are probably more common.



Following this short survey, the Facilitator should explain that individualized teacher conferences are an important tool through which to both ‘plan’ a teacher evaluation as well as provide ‘confidential feedback.’ Conferences can be used for both Formative and Summative Teacher Evaluation, but tend to be easier when doing ‘growth-focused’ assessments. This is because teachers are less threatened by conferences with a Formative purpose where they generally have more input into setting the focus of the evaluation. In contrast, ‘accountability’ focused assessments usually have a pre-determined focus and the conference is just to inform the teacher of what the focus will be. Teachers have little input into such conferences.



Next, the Facilitator should pass out **Handout 31: ‘Guidelines for Organizing Teacher Conferences’** to each participant. Review this Handout with participants emphasizing some of the following discussion points:



- Teachers have much more input in setting the parameters and goals of a conference in a Formative Teacher Evaluation than in a Summative Evaluation.
- Pre- and Post-conferences in Summative Evaluation tend to be more focused on information-sharing about goals, decisions made, etc. rather than giving



the teacher an opportunity for input into the evaluation.

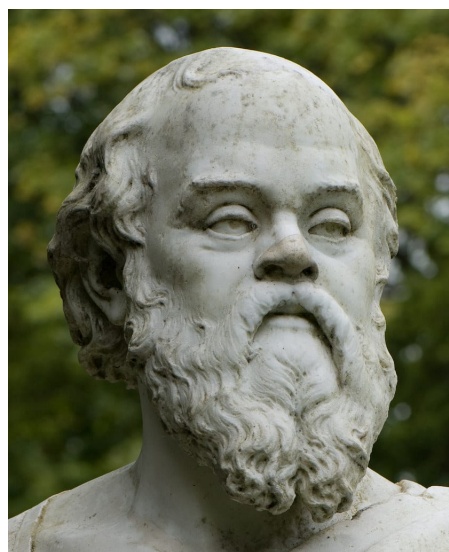
- Reaching a ‘consensus’ on the goals of the Teacher Evaluation is a unique feature of pre-conferences that are focusing on professional growth.
- Using the Socratic Method during conferencing in which evaluators ask questions rather than dictate what the teacher did ‘right’ or ‘wrong’ is a very effective technique for maintaining a non-threatening tone of a conference, especially in the case of a Formative Evaluation.

The Facilitator should provide ample opportunity for participants to ask questions on the guidelines for organizing teacher conferences, based on the Handout provided.

**Optional Exercise:** If time permits, it is strongly suggested that the Facilitator consider organizing a role play in which participants act out a pre- or post-conference in either a Formative or Summative setting. Use the directions provided for the Role Play Exercise in Handout 31 to help participants organize the role play. Make sure that participants are clear whether the conference that they act out is a Pre-Conference or a Post-Conference and whether the purpose is ‘Formative’ or ‘Summative.’ The content of the conversation during the role play may involve setting goals, identifying processes for data collection, sharing information, using Socratic techniques to provide feedback, etc. Allow as many groups to present their role plays as time permits.

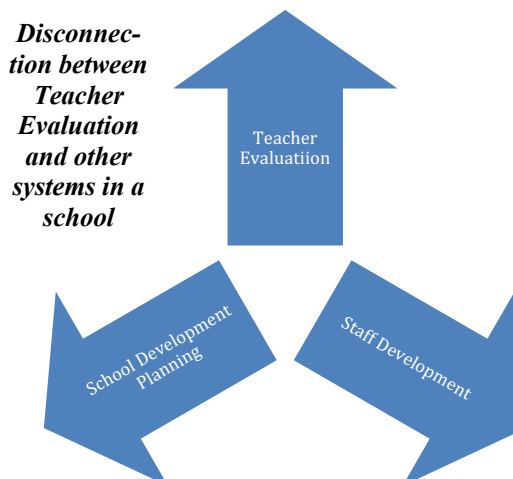
### What is the Socratic Method?

The Socratic Method is a teaching technique developed by the Greek philosopher Socrates. In this method, a teacher helps students discover knowledge on their own by asking probing questions in a way that helps students explore the underlying beliefs that shape their views and opinions.



### Linking Teacher Evaluation & Staff Development

In this next session, the Facilitator wants to help participants understand the disconnection that often exists between Teacher Evaluation, School Development Planning, and Staff Development. Often, these process occur independently of each other with little connection. This is unfortunate because there are



so many ways in which these processes can help support and reinforce each other.



To emphasize this point, draw the diagram (shown above) on the board to emphasize the problem of disconnection between these three elements of programming in a school or institution.



Next, the Facilitator should distribute **Handout 32: 'Linking Your Teacher Evaluation System to Staff Development'** to all participants. Review the Handout together being sure to emphasize the following discussion points:

#### Suggested Discussion Points

- Formative Teacher Evaluation helps to identify real teacher professional needs, which can inform the content of a Staff Development Program.
- It may also be possible to link both Teacher Evaluation and Staff Development with School Improvement Planning so that there is as much overlap between these three elements as possible.
- Staff Development planning is most compatible with Formative Teacher Evaluation, not Summative Teacher Evaluation.
- When Teacher Evaluation, Staff Development, and School Planning are all moving in the same direction and reinforcing each other, it is more likely that this will impact positively on the effective delivery of the school's curricular program.



Following this discussion, review the concrete example provided in Handout 32 and go over the discussion questions provided to be sure that everyone understands the example that was explained:

#### General Discussion Questions

Read the example provided above and try to answer the following questions.

1. Who in this school assessed the teachers?  
*Answer: Technical Grade Leaders*
2. How many need categories did assessors identify?  
*Answer: Five categories were identified – how to organize group work, classroom management, using ICT, using the library more effectively, and using PBL.*
3. Through what process did the SMC identify general priorities to improve the school?  
*Answer: Through the School Improvement Planning process.*
4. What were the commonalities between the SIP and teacher assessment?  
*Answer: (1) Improved teaching practices involving student-centered learning, especially Cooperative Learning and PBL; (2) Using ICT in education*
5. What staff development activities were identified to link the SIP and teacher evaluation?  
*Answer: The following Staff Development activities were identified:*
  - *A foundational workshop for teachers on PBL and Cooperative Learning (5 days)*
  - *2 Reflection Workshops (1 day each)*
  - *Opportunities for team teaching with TGLs using new techniques*
  - *One exposure visit to see a New Generation School where such methods are more commonly used.*
  - *Training every Saturday for teachers on basic ICT techniques. Providing 50% matching funds from school funds to help teachers buy a used laptop.*



## 3.14 Bringing the Elements of Teacher Evaluation Together



**Lesson Time:** 3 Hours



### **Trainer Preparation:**

- Write up the Learning Outcomes of the lesson on a sheet of poster paper to introduce the lesson.
- Make copies of all handouts



### **Resources/Materials:**

- Poster paper, marker pens
- Poster sheet summarizing the learning outcomes of the lesson.
- **Handout 33:** *Bringing the Elements of Effective Teacher Evaluation Together*



### **Learning Outcomes:**

1. Participants can *identify* the key attributes that should characterize a Formative Teacher Evaluation (e.g., Assessment tools with open-ended responses, links with professional development systems, etc.).
2. Participants can fully *explain* the key attributes that should characterize a Formative Teacher Evaluation (e.g., Assessment tools with open-ended responses, links with professional development systems, etc.).
3. Participants can *identify* the key attributes that should characterize a Summative Teacher Evaluation (e.g., Assessment tools with both closed and open-ended responses, fixed content focus, implementation by administrators, etc.).
4. Participants can fully *explain* the key attributes that should characterize a Summative Teacher Evaluation (e.g., Assessment tools with both closed and open-ended responses, fixed content focus, implementation by administrators, etc.).
5. Participants can *distinguish* the key differences between a Formative and a Summative Teacher Evaluation System.
6. Participants can *identify* the challenges and solutions that they might encounter when setting up a Teacher Evaluation System at their school or institution using the guidelines provided.

## Training Session Plan

### Outcomes of the Lesson



Place a sheet of poster paper up on the board that summarizes the learning outcomes for the lesson. Explain the outcomes and that these results exemplify what the participants should be able to do at the completion of the lesson.

### Defining Expectations When Participants Return to their Schools & Institutions

The Facilitator should begin this session by noting that it is the final session in the training workshop as this relates to Teacher Evaluation. The expectation now is that participants will be able to use what they have learned to set up teacher assessment systems at their schools or institutions that promote both ‘professional growth’ and ‘accountability.’



Hopefully, these purposes will NOT be served by a Dual Teacher Evaluation System. Of course, we must realize that we must balance our expectations with some of the constraints that exist in the real situation at our schools and institutions. This



session will hopefully help participants to sum up what the expectations for setting up a local Teacher Evaluation System should be and some of the challenges that we might face in trying to follow the guidelines to actually set up such systems.

### Key Elements of a Formative Teacher Evaluation System



Start a discussion of the key elements of a Formative Teacher Evaluation System by distributing **Handout 33: ‘Bringing the Elements of Effective Teacher Evaluation Together’** to each participant. Identify and review each of the elements shown in the diagram in the handout using the explanations provided in the chart below. These are the discrete elements of an ideal Formative Teacher Evaluation System:



#### Elements of a Formative Teacher Evaluation System

1. Assessment Tools with open-ended responses
2. Using unscored tools
3. Clear communication of purpose (i.e., professional growth)
4. Consensual identification of content
5. Pre-and Post-conferencing
6. Use of Mentors or Peers as evaluators (to achieve separation of functions)
7. Links with Continuous Professional Development

The Facilitator should try to present these elements in a way that fosters dis-

discussion with participants. Several ideas have already been discussed in previous sessions. Try using some of the following questions to elicit responses and ideas from participants about the organization of a Formative Teacher Evaluation System:

#### **Questions to Elicit Discussion about Formative Teacher Evaluation**

- Why do you think that assessment tools should be open-ended and unscored in Formative Evaluation?
- Who sets the content focus of a Formative Evaluation?
- How can such evaluations be made as non-threatening as possible?
- What role do pre- and post-conferences play in a Formative Evaluation and what would happen if such conferences were not used?
- What form might linkages between Formative Teacher Evaluation and Staff Development take?
- Why do you think it is important not to use administrators to do Formative Teacher Evaluation?

These are only some of the questions that Facilitators might use to ensure that there is discussion with participants about the guidelines provided for setting up a Formative Teacher Evaluation System. Use these and similar questions that the Facilitator comes up with on his/her own to avoid doing this review of the handout as a lecture.

#### **Key Elements of a Summative Teacher Evaluation System**

When discussions on Formative Teacher Evaluation guidelines have been completed, the Facilitator should next move to a discussion of Summative Teacher Evaluation using the same handout. Identify and review each of the elements of a Summative Teacher Evaluation System shown in the diagram in the handout using the explanations provided in the chart below. These are the discrete elements of an ideal Summative Teacher Evaluation System:

#### **Elements of a Summative Teacher Evaluation System**

1. Assessment Tools with open and closed-ended responses
2. Use of Scored Tools
3. Clear communication of purpose (i.e., accountability)
4. Identification of teachers to evaluate through sampling
5. Fixed Content Focus
6. Pre-and Post-conferencing
7. Use of Administrators as evaluators (to achieve separation of functions)

Once again, the Facilitator should try to present these elements in a way that fosters discussion with participants. Try using some of the following questions to elicit responses and ideas from participants about the organization of a Summative Teacher Evaluation System:

#### **Questions to Elicit Discussion about Summative Teacher Evaluation**

- Why do you think that assessment tools should need to be scored for a Summative Evaluation?
- How is the content of a Summative Teacher Evaluation set?

- What role do pre- and post-conferences play in a Summative Evaluation and what would happen if such conferences were not used?
- Why do you think it is better for administrators to be responsible for Summative Teacher Evaluation?
- Why do you think sampling is used to select teachers for evaluation?
- Do you see any similarities at all between Summative and Formative Evaluation?

### **Identifying the Challenges and Solutions of Setting Up Effective Teacher Evaluation Systems**



Following the review of Formative and Summative Teacher Evaluation Systems, the Facilitator should next ask participants to complete the exercise provided at the end of Handout 33. The purpose of this exercise is to help everyone to think clearly about the real situation at their schools or institutions (e.g., staffing shortages, lack of time, little interest among teachers to improve their teaching, etc.) in order to identify what these challenges might be and what the possible solutions are. Pass out a sheet of poster paper to each small group and using the template provided in the handout, ask participants to answer the given questions, writing out their answers as concisely as possible. Give participants about 40 minutes for this exercise.



When participants have completed their discussions in their small groups, ask each group to present their thoughts and ideas to the whole group. The Facilitator should ask probing questions about what participants have presented, particularly in cases where they might reject some of the guidelines provided. Many of the challenges are likely to be same from group to group (e.g., staffing shortages) so the Facilitator should check to see how the solutions differ between groups in order to give groups alternative solutions to solving the same problems.



# SECTION 4: Course Materials



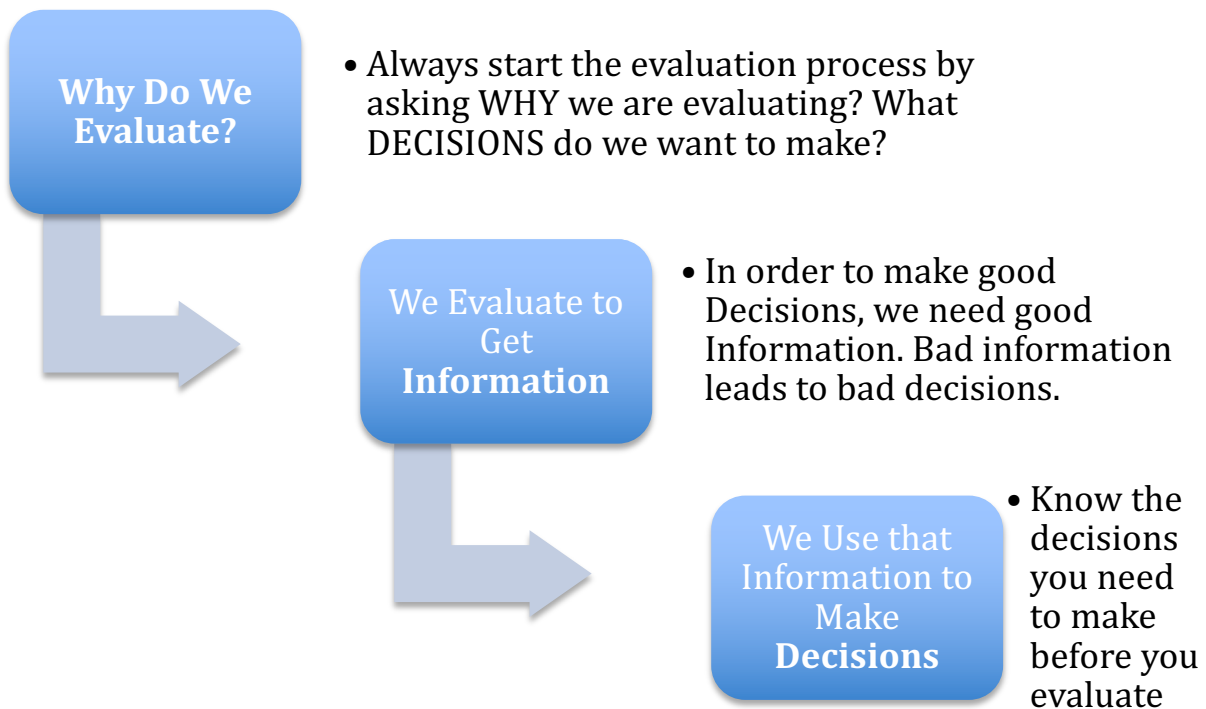
# HANDOUTS FOR PART 1

## HANDOUT 1: Why Do We Evaluate?

### Understanding the Purpose of Evaluation

Many teachers often do assessments without knowing clearly ‘why’ they are doing the evaluation. Often teachers do evaluation mechanically because it is something that the Ministry asks them to do. It is important for every teacher to reflect on what decisions need to be made **before** they start organizing their evaluation because the kinds of decisions that need to be made may affect the design of the evaluation.

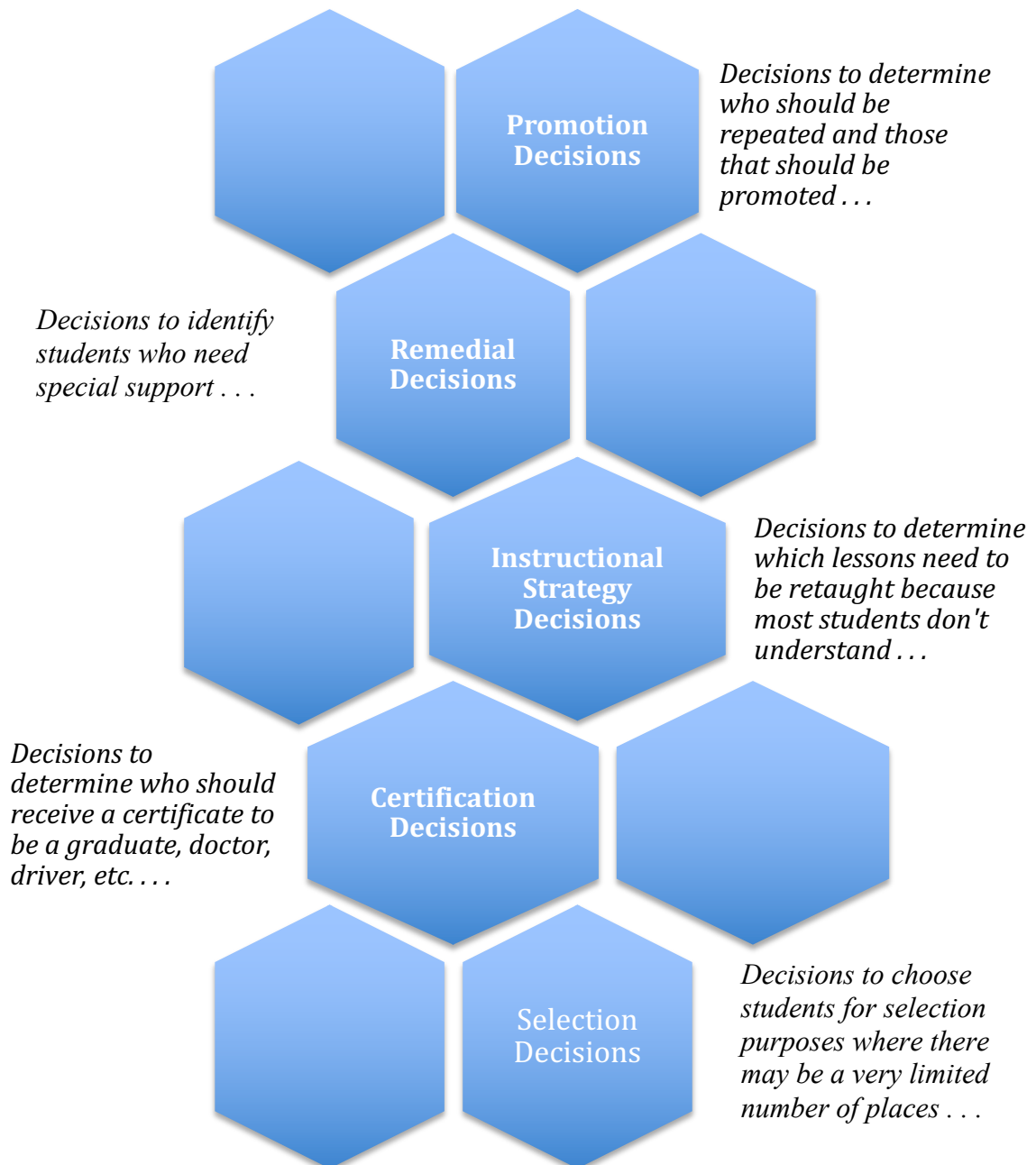
In order to make **good decisions**, we need **good information**. If the design of the assessment that we organize does not match the purpose of the evaluation (i.e., the decisions that need to be made), we may get bad information. Bad Information will lead to Bad Decisions.





## HANDOUT 2: Types of Educational Decision-making

We do evaluation to get **information** to make decisions. These decisions are often very different and may include the following **purposes** . . . .



## **HANDOUT 3: How Do We Get Information for Our Decision-making?**

There are many ways that an evaluator can use to get information in order to make decisions. Some of these ways may be suitable for the classroom and others may be suitable for use outside of the classroom, depending on our purpose.

Some of the instruments and methods used to collect information may include the following:

### **Evaluation Strategies to Gather Information**

#### **Tests**

- Written Tests
- Oral Tests
- Electronic Tests

#### **Observations**

- Checklists
- Open-ended Responses

#### **Interviews**

- Person-to-Person
- Telephone

#### **Surveys**

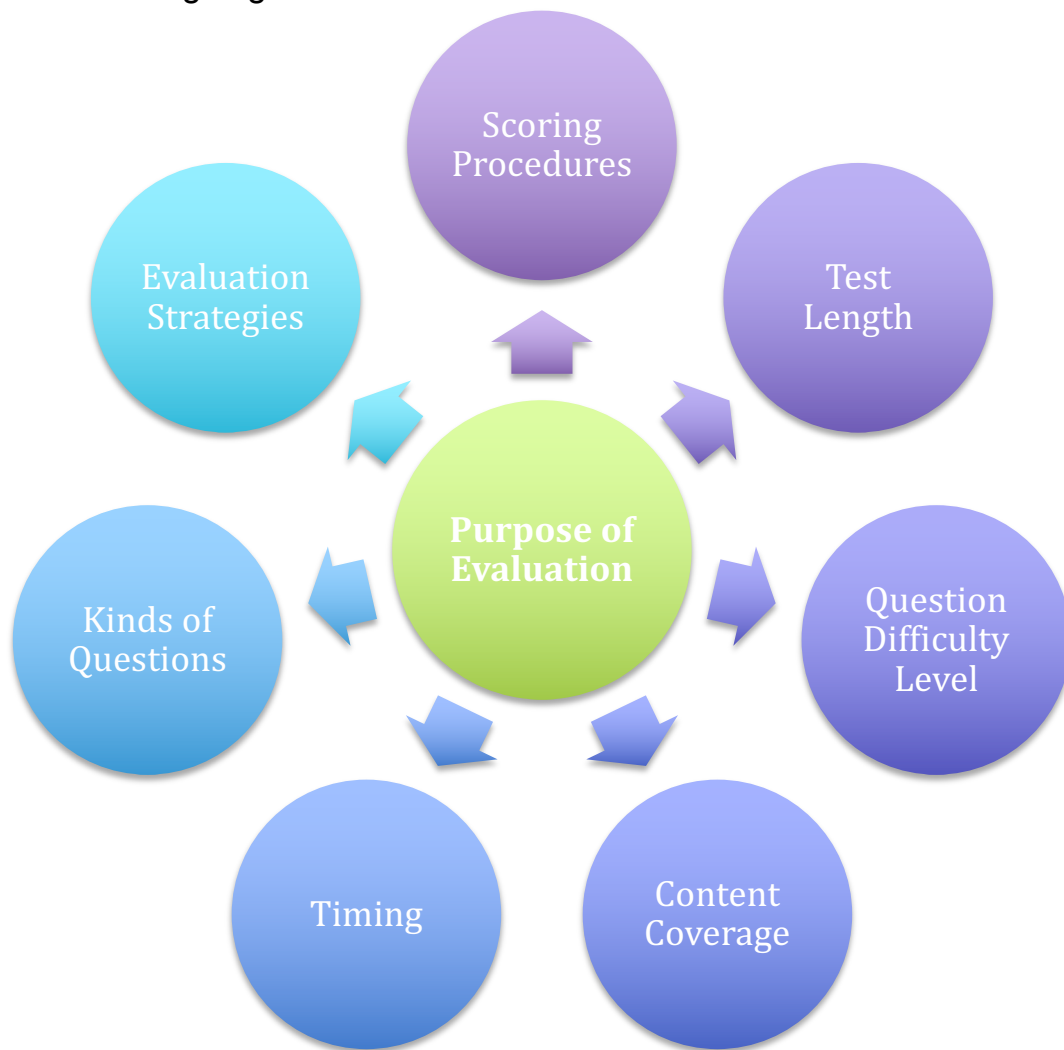
- Questionnaires
- Focus Group Discussions

#### **Demonstrations**

- Demonstrating a Lesson
- Doing an Experiment
- Etc.

## HANDOUT 4: How Evaluation Purpose Affects Test Design

**Explanation:** Knowing the Purpose of one's evaluation is important because different purposes may affect the design of the assessment tool that one develops. There are 6 parameters to consider as summarized in the diagram provided below. Under each parameter, there may be different options to consider when designing one's evaluation.



Parameter	Options	Parameter	Options
<b>Scoring Procedures</b>	<ul style="list-style-type: none"> <li>• Criterion-Referenced</li> <li>• Normative</li> </ul>	<b>Timing</b>	<ul style="list-style-type: none"> <li>• Often/Infrequently</li> <li>• Continuous/Terminal</li> </ul>
<b>Test Length</b>	<ul style="list-style-type: none"> <li>• Short</li> <li>• Medium</li> <li>• Long</li> </ul>	<b>Kinds of Questions</b>	<ul style="list-style-type: none"> <li>• Objective/Subjective</li> <li>• Product or Process-focused</li> </ul>
<b>Question Difficulty</b>	<ul style="list-style-type: none"> <li>• Easy</li> <li>• Difficult</li> </ul>	<b>Evaluation Strategies</b>	<ul style="list-style-type: none"> <li>• Written Test</li> <li>• Oral Test</li> <li>• Observation</li> <li>• Survey</li> </ul>
<b>Content Coverage</b>	<ul style="list-style-type: none"> <li>• Comprehensive</li> <li>• Sampling</li> </ul>		

## An Exercise and an Example

**Directions:** Consider the evaluation purposes provided below and determine how the purpose might affect the design of a test based on the parameters reviewed above.

**Evaluation Purpose:** Selection

Parameter	Suggested Design Feature	Parameter	Suggested Design Feature
Scoring Procedures	<ul style="list-style-type: none"> <li>• Normative</li> </ul>	Timing	<ul style="list-style-type: none"> <li>• Once</li> </ul>
Test Length	<ul style="list-style-type: none"> <li>• Medium to Long</li> </ul>	Kinds of Questions	<ul style="list-style-type: none"> <li>• Mostly Objective</li> <li>• Product-focused</li> </ul>
Question Difficulty	<ul style="list-style-type: none"> <li>• Medium to High Difficulty Level</li> </ul>	Evaluation Strategies	<ul style="list-style-type: none"> <li>• Written Test</li> </ul>
Content Coverage	<ul style="list-style-type: none"> <li>• Sampling of Content</li> </ul>		

**Discussion:** When one’s purpose is ‘selection’ (e.g., for a scholarship or limited number of places in a school or university), it means that the number of candidates selected will be but a small handful of those applying. This suggests that one would only choose the top scoring candidates up to a point that equals the number of available places. That is, the scoring procedure would be ‘normative.’ The test should be medium to long to sample as much content as possible and the questions should tend towards the high difficulty level to get the most qualified candidates. In order to sample as much content as possible, one should use objective, product-focused questions, which can be answered quickly by candidates. This would be a one-time test since the purpose is a one-time selection.

**Evaluation Purpose:** Remedial Decisions

Parameter	Suggested Design Feature	Parameter	Suggested Design Feature
Scoring Procedures	•	Timing	•
Test Length	•	Kinds of Questions	•
Question Difficulty	•	Evaluation Strategies	•
Content Coverage	•		

**Discussion:** \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

## HANDOUT 5a: Types of Evaluation for the Classroom

**Explanation:** Different kinds of evaluation are usually characterized according to the purpose, which they serve. Two of the most important kinds of evaluation used by classroom teachers include:



### Summative Evaluation & Formative Evaluation

**The Name of the Evaluation Shows Its Purpose:** Notice how in English the name of a particular kind of evaluation suggests its purpose. For example, the name *Summative* comes from the word ‘summarize’ or the purpose of summarizing a student’s academic achievement. Similarly, the name *Formative* comes from the English word ‘form’ or the purpose of forming one’s educational plans. In Khmer language, the term generally used to name Formative evaluation does not suggest its purpose. For this reason, we will generally use the English term to refer to Formative Evaluation.

**Other Types of Evaluation:** There are many other kinds of evaluation to be aware though Formative and Summative Evaluation are the most important to know for classroom teachers. Other kinds of evaluation to be aware of include:

**Diagnostic Evaluation:** This is a form of pre-assessment that allows a teacher to determine students’ individual strengths, weaknesses, knowledge, and skills prior to instruction. It is primarily used to diagnose student difficulties and to guide lesson and curriculum planning. In this sense, it is similar to Formative Evaluation but only occurs *before* instruction begins and not during or after instruction.

**Strategic Evaluation:** This is a kind of evaluation that is done for very specific purposes such as determining a student’s **entry status** (e.g., placement in an advanced class or a remedial support class), **selecting** a student for a scholarship of which only a few are available, or **classifying** a student’s rank on a particular task or set of tasks. Strategic Evaluation is usually only conducted *once* to meet the specific purpose for which it was intended.



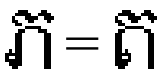

## Formative Evaluation:

<b>What is it?</b>	<b>A kind of evaluation that occurs before or during instruction for the purpose of guiding instructional planning and/or improve the effectiveness of one's teaching.</b>																									
<b>When is it used?</b>	<p>Formative Evaluation can be used for one or all of the following purposes:</p> <ol style="list-style-type: none"> <li>1 Forming educational plans for a new group of students whose abilities in a specific subject are not yet known.</li> <li>2 Assessing the effectiveness of currently used instructional strategies in the classroom.</li> <li>3 Reviewing a specific topic to determine areas in need of re-teaching.</li> <li>4 Identifying individual students who need remedial help in a given lesson.</li> </ol>																									
<b>How is it done?</b>	<p>The following pointers should help teachers to conduct Formative Evaluation effectively:</p> <ol style="list-style-type: none"> <li>1 <b>First</b>, identify the specific learning unit that one wishes to evaluate. Usually the learning unit evaluated should have a limited domain with a small number of related objectives.</li> <li>2 <b>Second</b>, be sure the learning area evaluated has a documented set of objectives.</li> <li>3 <b>Third</b>, develop a test plan that covers all the objectives of the learning unit. The number in parentheses indicates the number of questions asked.</li> </ol> <p><b>Subject:</b> Set Theory</p> <table border="1" data-bbox="411 1279 1401 1563"> <thead> <tr> <th>Topic</th> <th>Remembering</th> <th>Understanding</th> <th>Application</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>Idea of a Set</td> <td>Objective A (2) Objective B (2) Objective C (2)</td> <td></td> <td></td> <td>6</td> </tr> <tr> <td>General Set Notation</td> <td>Objective D (2)</td> <td>Objective E (2) Objective F (2)</td> <td></td> <td>6</td> </tr> <tr> <td>Kinds of Set</td> <td>Objective H (2)</td> <td></td> <td>Objective G (2)</td> <td>4</td> </tr> <tr> <td><b>Total</b></td> <td><b>10</b></td> <td><b>4</b></td> <td><b>2</b></td> <td><b>16</b></td> </tr> </tbody> </table> <p>*Note: 2 Questions per objective.</p> <ol style="list-style-type: none"> <li>4 <b>Fourth</b>, set questions according to the test plan developed. The questions should be of a relatively low level of difficulty and should enable teachers to observe the student's <i>process of thinking</i>. Usually, an equal number of test questions are set for each objective. In the above plan, two questions are set for each objective.</li> <li>5 <b>Fifth</b>, After correcting the results, the teacher should try to see which questions/objectives students got wrong the most to determine areas in need of re-teaching.</li> </ol>	Topic	Remembering	Understanding	Application	Total	Idea of a Set	Objective A (2) Objective B (2) Objective C (2)			6	General Set Notation	Objective D (2)	Objective E (2) Objective F (2)		6	Kinds of Set	Objective H (2)		Objective G (2)	4	<b>Total</b>	<b>10</b>	<b>4</b>	<b>2</b>	<b>16</b>
Topic	Remembering	Understanding	Application	Total																						
Idea of a Set	Objective A (2) Objective B (2) Objective C (2)			6																						
General Set Notation	Objective D (2)	Objective E (2) Objective F (2)		6																						
Kinds of Set	Objective H (2)		Objective G (2)	4																						
<b>Total</b>	<b>10</b>	<b>4</b>	<b>2</b>	<b>16</b>																						

## Exercise: What's the Problem? Analyzing Students' Mistakes

*Directions: Formative Evaluation involves reviewing students' processes of thinking in order to know how best to reteach certain concepts. Review the student work provided below and indicate the mistake in their reasoning (not the mistake itself) that led the student to solve the problem incorrectly.*

<i>No</i>	<i>Student Work That Indicates a Mistake</i>	<i>Your Analysis of WHY the Student Did This Wrong</i>
1.	$\begin{array}{r} 1,300 \\ - 522 \\ \hline 878 \end{array}$	
2.	$\begin{array}{r} 140 \\ - 21 \\ \hline 120 \end{array}$	
3.	$\begin{array}{r} 1,300 \\ - 522 \\ \hline 788 \end{array}$	
4.	$\begin{array}{r} 521 \\ + 888 \\ \hline 13,109 \end{array}$	
5.	$\begin{array}{r} 51 \\ + 49 \\ \hline 90 \end{array}$	
6.	$\begin{array}{l} 9 > 11 \\ 10 < 9 \end{array}$	
7.	$1 + 2 + 3 = 123$	
8.	<p style="text-align: center;"><i>recieve</i> <i>niegbor</i></p>	
9.	<p style="text-align: center;">ជំងឺ តុច្ចា</p>	

<p><b>10.</b></p>	<p>        </p>	<p>a)</p> <p>b)</p> <p>c)</p> <p>d)</p>
-------------------	---	---



## Summative Evaluation:

<b>What is it?</b>	<b>A kind of evaluation that occurs at the end of instruction for the purpose of summarizing and documenting student achievement.</b>																																													
<b>When is it used?</b>	Summative Evaluation is the kind of assessment with which teachers are most familiar. It is the typical graded test administered at the END of a lesson to determine students' achievement. The Final Examination is the most classic example of a summative test.																																													
<b>How is it done?</b>	<p>The following pointers should help teachers to conduct Summative Evaluation effectively:</p> <ol style="list-style-type: none"> <li><b>First</b>, determine the range of topics that the summative test will cover. Usually summative tests cover a broader domain of content than do formative tests.</li> <li><b>Second</b>, be sure the learning area evaluated has a documented set of objectives.</li> <li><b>Third</b>, develop a test plan that 'samples' the range of objectives that have been identified. Due to its broader domain of content, the summative test cannot cover all objectives as in a formative test but must contain a representative sample instead. In summative tests, some objectives may be weighted more heavily than others.</li> </ol> <p><b>Subject: Science (Unit 1 on Plants and Unit 2 on Health)</b></p> <table border="1" data-bbox="435 1243 1401 1594"> <thead> <tr> <th>Topic</th> <th>Remembering</th> <th>Understanding</th> <th>Application</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>1. Natural Environments</td> <td>3 Questions</td> <td></td> <td>1 Questions</td> <td>4</td> </tr> <tr> <td>2. Usefulness of Plants</td> <td></td> <td>5 Questions</td> <td></td> <td>5</td> </tr> <tr> <td>3. Life Cycle of Plants</td> <td>2 Questions</td> <td>2 Questions</td> <td>1 Question</td> <td>5</td> </tr> <tr> <td>4. Plant Classifications</td> <td>3 Questions</td> <td>2 Questions</td> <td></td> <td>5</td> </tr> <tr> <td>5. The Skeleton &amp; Muscles</td> <td>1 Question</td> <td>3 Questions</td> <td>2 Questions</td> <td>6</td> </tr> <tr> <td>6. The Eyes</td> <td>1 Question</td> <td></td> <td>1 Questions</td> <td>2</td> </tr> <tr> <td>7. Dengue &amp; Malaria</td> <td>2 Questions</td> <td></td> <td>4 Questions</td> <td>6</td> </tr> <tr> <td><b>Total</b></td> <td><b>12 Questions</b></td> <td><b>12 Questions</b></td> <td><b>9 Questions</b></td> <td><b>33</b></td> </tr> </tbody> </table> <p>*Note: 2 Questions per objective.</p> <ol style="list-style-type: none"> <li><b>Fourth</b>, develop questions for the test according to the test plan. The questions should be of a relatively moderate level of difficulty. Since remedial instruction is not the purpose of summative evaluation, teachers should try to use objectively scored and product-focused questions (e.g., Multiple Choice Questions) on their summative tests as much as possible.</li> <li><b>Fifth</b>, grade test papers according to the weighting scheme specified in the test plan.</li> </ol>	Topic	Remembering	Understanding	Application	Total	1. Natural Environments	3 Questions		1 Questions	4	2. Usefulness of Plants		5 Questions		5	3. Life Cycle of Plants	2 Questions	2 Questions	1 Question	5	4. Plant Classifications	3 Questions	2 Questions		5	5. The Skeleton & Muscles	1 Question	3 Questions	2 Questions	6	6. The Eyes	1 Question		1 Questions	2	7. Dengue & Malaria	2 Questions		4 Questions	6	<b>Total</b>	<b>12 Questions</b>	<b>12 Questions</b>	<b>9 Questions</b>	<b>33</b>
Topic	Remembering	Understanding	Application	Total																																										
1. Natural Environments	3 Questions		1 Questions	4																																										
2. Usefulness of Plants		5 Questions		5																																										
3. Life Cycle of Plants	2 Questions	2 Questions	1 Question	5																																										
4. Plant Classifications	3 Questions	2 Questions		5																																										
5. The Skeleton & Muscles	1 Question	3 Questions	2 Questions	6																																										
6. The Eyes	1 Question		1 Questions	2																																										
7. Dengue & Malaria	2 Questions		4 Questions	6																																										
<b>Total</b>	<b>12 Questions</b>	<b>12 Questions</b>	<b>9 Questions</b>	<b>33</b>																																										

## Designing a Summative Test

**Directions:** Given what you know about the purpose of Summative Evaluation, what sorts of characteristics do you think that a summative test would have? Clarify the various characteristics of a summative test by completing the table provided below. Make any necessary clarifications in the space provided below for Discussion.

**Evaluation Purpose:** Summative Decision-making

Parameter	Suggested Design Feature	Parameter	Suggested Design Feature
Scoring Procedures	•	Timing	•
Test Length	•	Kinds of Questions	•
Question Difficulty	•	Evaluation Strategies	•
Content Coverage	•		

**Discussion:** \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

## **HANDOUT 5b: Summary of the Differences between Summative & Formative Evaluation**

<b>Formative Evaluation</b>	<b>Summative Evaluation</b>
1. Coverage of short lesson segments	1. Coverage of longer lesson segments and units
2. Has a 'comprehensive' coverage of the objectives that have been taught since the content domain is narrow	2. 'Samples' the objectives that have been taught since the content to be assessed is extensive
3. Administered continuously	3. Administered terminally
4. Questions have a narrower range of difficulty levels tending towards easier questions.	4. Questions have a wider range of difficulty levels tending towards questions of moderate or high difficulty (but mostly moderate)
5. Uses a Criterion-referenced scoring framework	5. Uses either Criterion or Norm-Referenced scoring frameworks
6. Uses 'process-oriented' questions	6. Uses 'product-oriented' questions that can be answered quickly to maximize coverage.
7. Most often used for remedial and re-teaching purposes and not for final grading	7. Important in final grading for promotion purposes

## **HANDOUT 6: Useful Frameworks for Interpreting Test Scores**

### **QUESTION:**

Is the meaning of a test score a straightforward matter with little need for interpretation?



### **ANSWER:**

**No!** . . .All educational measurement is open to interpretation. The interpretation of test scores should occur in a framework that is in harmony with the purpose of the assessment.



Two of the most important and frequently used interpretive frameworks employed for educational decision-making are known as:  
**Criterion-referenced Scoring & Norm-referenced Scoring**

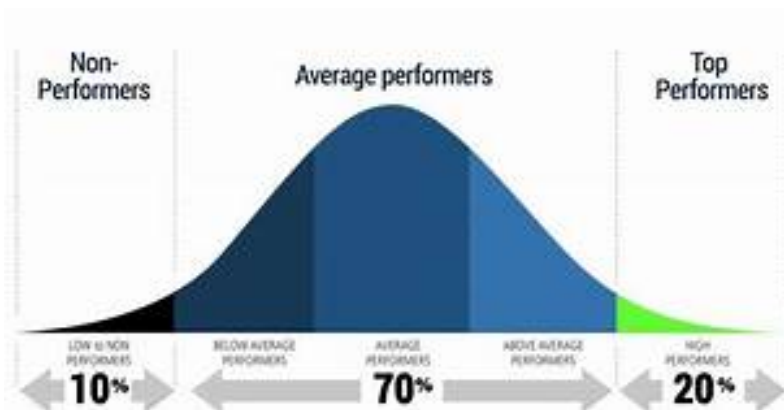
## HANDOUT 7: Criterion-Referenced Scoring

<b>What is it?</b>	<b>Criterion-referenced Scoring is defined as assessment that is based on each student’s ‘mastery’ of course objectives. Within this framework, a fixed ‘criterion’ set by the teacher, the school, or the Ministry defines content mastery (e.g., 50%, 65%, etc.)</b>			
<b>Is it common?</b>	Yes! . . .Whenever you have given a test and used a pass mark of 5 points out of 10, you have used a criterion-referenced scoring framework for interpreting an educational measurement.			
<b>How do you set the criterion?</b>	There are no fixed rules for setting a criterion. The teacher, the school director, or the Ministry simply make a judgment regarding the criterion that seems to be most suitable.			
<b>Did you know this?</b>	Criteria for passing a test vary widely from country to country. In Cambodia, the Ministry of Education, Youth, & Sport defines content mastery as 5 points out of 10 or more. Here are some other examples of assessment criteria used in other countries:			
	United States: Netherlands: United Kingdom:	65% out of 100 55% out of 100 50% out of 100	New Zealand: India: Cambodia	50% out of 100 40% out of 100 5 out of 10
<b>When should I use Criterion-Referenced Scoring?</b>	<p><b>To answer this question, several factors must be considered:</b></p> <p><b>Factor 1: Purpose of the Assessment:</b> When the purpose of one’s assessment is to make educational decisions that concern <i>Promotion, Certification, or Attainment of Minimum Level Competencies</i>, then a Criterion-Referenced scoring framework is recommended.</p> <p><b>Factor 2: The Nature of the Objectives:</b> School systems will often use Criterion-Referenced scoring when they have clearly identified educational objectives whose attainment they wish to assess.</p> <p>When objectives are documented clearly, the criterion for mastery can be pegged to the attainment of a minimum number of objectives (e.g., 50% of all documented objectives).</p>			

## HANDOUT 8: Norm-Referenced Scoring

<b>What is it?</b>	<p>Norm-referenced Scoring is defined as assessment that reflects a student’s standing or position in comparison with other students. In such frameworks, even if a student studies hard, he may only receive an average grade if it so happens that everyone else also studies hard too.</p>
<b>Is it common?</b>	<p>Norm-referenced scoring is most often used in higher education contexts and sometimes at secondary school level. Schools sometimes use this framework if their goal is to select students for some specific purpose. In some cases, if all students do poorly on an examination, a teacher may decide to grade students on what is known as a ‘curve.’</p>
<b>Did you know this?</b>	<p>When Cambodian teachers classify students by their rank at the end of each month, they are using Norm-Referenced Scoring.</p>
<b>When should I use Norm-Referenced Scoring?</b>	<p><b><i>To answer this question, several factors must be considered:</i></b></p> <p><b>Factor 1: Purpose of the Assessment:</b> Norm-Referenced scoring is most appropriate when the purpose of one’s assessment is for educational decisions relating to <i>Selection, Admission, or Classification</i>.</p> <p><b>Factor 2: Nature of the Students Who Are Being Assessed:</b> When the students being assessed are very diverse in nature encompassing school populations with widely differing facilities and resources, Norm-Referenced Scoring may be highly desirable to ensure that there are provisions in place to push lower scoring students over the passing line (using a curve).</p>
<b>How is it done?</b>	<p><b>First,</b> remember the goal of Norm-referenced Scoring. In this kind of scoring, the teacher’s goal is to produce a distribution of student performance that conforms with what is known as the Normal Distribution. In this distribution, the majority of students cluster around the center while the highest and lowest performing students can be found in the tails of the distribution.</p>

## The Normal Distribution



**Second**, find the middle of your distribution. In the Normal Distribution, the middle of the curve represents the 'average' level of performance. The middle of the distribution becomes the anchor on which your scoring is based.

**Third**, make up a new scale of marks based on the distance from the middle score. In this way, the teacher can make more students pass than ordinarily might have been the case (in cases where too many students performed poorly) or alternatively make more students fail if too many students scored at the higher range of the scale. This rearrangement of scores can be done through the use of percentiles. A percentile is a statistic that represents the percentage of students who stand above or below a certain mark.

Follow the steps below to create a normative scoring framework for a particular test:

1. Take the number of students in the entire class and estimate some key percentiles (e.g., 90<sup>th</sup> percentile; 80<sup>th</sup> percentile, 75<sup>th</sup> percentile, etc.). In the example provided, there are 40 students in the distribution. This would mean that the 4<sup>th</sup> student from the top would be at the 90<sup>th</sup> percentile while the 4<sup>th</sup> student from the bottom would be in the 10<sup>th</sup> percentile.
2. Mark the scores representing key percentile points such as the 90<sup>th</sup>, 80<sup>th</sup>, etc.
3. Designate new scores to shift the scale in the direction that you want. In the example provided below, too many students are failing. That is, more than half

scored less than 5 points, which is the cut-off for a passing mark. In the example provided, the teacher has shifted the scale so that 4.5 (the 50<sup>th</sup> percentile) becomes 7. All other scores now also shift upwards accordingly.

4. Based on the new scale, 20% of students are counted as high performers (A), 70% are middle performers (B, C, or D) and 10% are non-performers (F).

Old Score	Number	Percentile	New Score	Letter Grade	
10	I		10	A	20%
9.5	I		10	A	
9	II	----- 90 <sup>th</sup>	10	A	
8.5	I		10	A	
8	I		10	A	
7.5	I		10	A	
7	I	----- 80 <sup>th</sup>	9.5	A	70%
6.5	II		9	B	
6	III	----- 75 <sup>th</sup>	8.5	B	
5.5	IIII		8	B	
5	III		7.5	C	
4.5	IIII	----- 50 <sup>th</sup>	7	C	
4	II		6.5	C	
3.5	II		6	C	
3	II		5.5	D	
2.5	II	----- 25 <sup>th</sup>	5	D	
2	II		4.5	D	
1.5	II		4	D	
1	IIII	----- 10 <sup>th</sup>	3.5	F	10%
	<b>40</b>				<b>100%</b>

In this example, the teacher wants to ensure that 90% of students 'pass' and 10% fail. This is the resulting curve.

In the curve, 4.5 was the original median score (50<sup>th</sup> percentile) which has now been changed to 7

## Questions for Discussion

- 1) How many students took this test?
- 2) What score represented the 50<sup>th</sup> percentile or median score? How many students scored above this score? What percentage does this comprise of the total distribution of students?
- 3) If this had been a criterion-referenced test in your school, what would have been the cut-off point for failing? How many students would have failed with this criterion for failing?
- 4) With the norm-referenced framework that the teacher made, how many students are now failing?
- 5) What do you think the teacher's purpose was in making this curve?



## HANDOUT 9: Comparison between Norm and Criterion-Referenced Scoring

Parameter	Norm-Referenced	Criterion-Referenced
Purpose	<b>Used for:</b> <ul style="list-style-type: none"> <li>• Surveying abilities and skills,</li> <li>• Determining individual differences</li> <li>• Discriminating between students</li> <li>• Admission decisions</li> </ul>	<b>Used for:</b> <ul style="list-style-type: none"> <li>• Assessing master of content</li> <li>• Determining achievement of minimum competency standards</li> <li>• Making Certification decisions</li> </ul>
Evaluation Type	Summative Only	Summative or Formative
Interpretation of Performance	<ul style="list-style-type: none"> <li>• Relative interpretations</li> <li>• Comparison with performance of others</li> </ul>	<ul style="list-style-type: none"> <li>• Absolute standard</li> <li>• Comparison of performance to an absolute standard</li> </ul>
Content Coverage	<ul style="list-style-type: none"> <li>• Objectives sampled are usually from a broad domain</li> </ul>	<ul style="list-style-type: none"> <li>• Objectives sampled can be from a broad or narrow domain</li> </ul>
Timing	<ul style="list-style-type: none"> <li>• Infrequent/Terminal</li> </ul>	<ul style="list-style-type: none"> <li>• Frequent or Terminal</li> </ul>
Test Length	<ul style="list-style-type: none"> <li>• Tend to be long</li> <li>• Relatively large number of items</li> </ul>	<ul style="list-style-type: none"> <li>• Can be either long or short in length</li> </ul>
Question Difficulty	<ul style="list-style-type: none"> <li>• Wider range of difficulty levels representing varied question difficulty levels</li> </ul>	<ul style="list-style-type: none"> <li>• Narrower range of difficulty levels concentrating on moderate to easy difficulty</li> </ul>
Evaluation Strategies	<ul style="list-style-type: none"> <li>• Written</li> </ul>	<ul style="list-style-type: none"> <li>• Written</li> </ul>

## **HANDOUT 10: Checking Your Knowledge on Evaluation Principles**

**Directions:** Based on your knowledge of Formative and Summative Evaluation, answer the questions below by choosing the ONE response that is the most correct.

1. The evaluation methodology selected by a teacher should be determined by the:
  - a. content area.
  - b. purpose of the evaluation.
  - c. course objectives.
  - d. student competencies.
2. A summative test should NOT be used for which of the following purposes?
  - a. Reviewing a lesson
  - b. Promoting a student
  - c. Evaluating overall achievement
  - d. Professional certification.
3. A monthly test in a Cambodian school is an example of a:
  - a. formative test.
  - b. norm-referenced test.
  - c. summative test.
  - d. certifying examination.
4. Process-oriented questions are most appropriate on a:
  - a. formative test.
  - b. summative test.
  - c. selection test.
  - d. certification test.
5. Certification decisions should ideally be based on evaluation procedures that:
  - a. measure relative achievement.
  - b. primarily involve observations.
  - c. involve paper and pencil tests exclusively.
  - d. specify a minimum competency level.
6. The most frequently expressed purpose for using Norm-Referenced testing is for:
  - a. Discriminating among learners.
  - b. Determining certification decisions.
  - c. Remedial testing
  - d. Determining mastery of content.
7. Which of the following best describes the characteristics of a summative test?
  - a. Narrow domain of content that samples all the objectives taught.
  - b. Broad domain of content that samples all the objectives taught.
  - c. Broad domain of content that assesses all objectives comprehensively.
  - d. Narrow domain of content that samples all the objectives taught.

8. Which of the following is not an example of Criterion-Referenced scoring?
- a. Bac II Examination
  - b. Monthly tests
  - c. Monthly ranking of children
  - d. Semester examinations

Answers:  
1. b  
2. a  
3. c  
4. a  
5. d  
6. a  
7. b  
8. c

## HANDOUT 11: Concepts of Assessment Validity

When ever a teacher designs a test, he or she should always keep in mind the following guideline:

*Educational Decision-making should be based on measurements derived from tests that are VALID.*

### Validity: What is it?

The following key concepts will help you to understand what validity is:

Validity is defined as the extent to which a test 'serves the purpose' for which it was designed

- **Example:** Recording the marks of a formative test in the Grade Book for promotion invalidates the purpose for which the test was designed.

A test is valid when it is 'relevant' to the information that it seeks to evaluate.

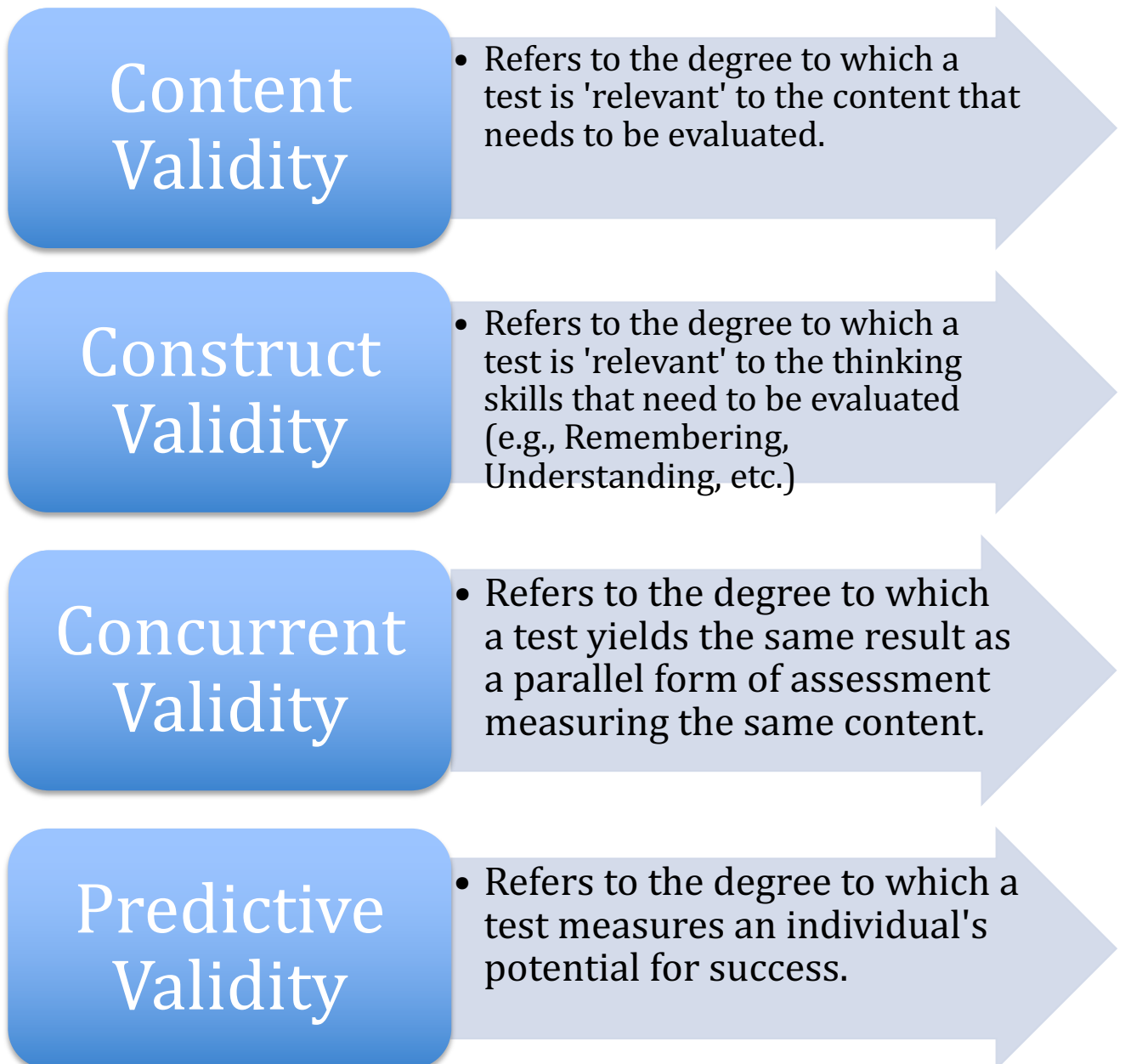
- **Example:** If a teacher teaches 50 objectives but then designs a test with only 5 questions, the test will have low validity because it was not relevant to what was taught.

A test is valid when the measurements obtained are 'reliable'.

- **Example:** If a teacher administers a test one week and gets a good result and administers a similar test on the same content the following week and gets a poor result, one or both tests may be invalid.

## HANDOUT 12a: Kinds of Validity

There are 4 Kinds of Validity that the classroom teacher should understand. The definition of these kinds of validity will help the teacher to assess to what degree a test is valid or not.



## Consider the Following Example:

*Mr. Sophea taught a unit on Mathematics with 4 lessons. He taught educational objectives that were spread across the Memory, Comprehension and Application levels. However, his test focused mainly on memory questions and the last 2 lessons of the unit.*



### Questions for Discussion

**Question 1:** Did Mr. Sophea's test have *Content Validity*?

**Question 2:** Did Mr. Sophea's test have *Construct Validity*?

**Question 3:** If Mr. Sophea administered a new test that covered all 4 lessons and included questions on Memory, Comprehension, and Application, do you think he would get a similar result? Why or why not? Be sure to phrase your answer in terms of the test's *Concurrent Validity*.

**Answers:**  
**Answer 1:** No, the test lacked 'Content Validity' because he omitted questions on the first two lessons of the unit.  
**Answer 2:** No, the test lacked Construct Validity because he did not include questions at the Comprehension and Application level.  
**Answer 3:** No, he would probably get a different result. Therefore, the previous test on the same unit probably lacks 'Concurrent Validity.'

## About Predictive Validity

### The Case of Albert Einstein

- In 1907, Albert Einstein applied for admission to the University of Bern as an Associate Professor.
- He was rejected . . .<sup>3</sup>
- Einstein went on to become the greatest scientist of the 20<sup>th</sup> Century.
- The evaluation standards of the university had very low predictive validity because their rejection of Einstein did not predict his later achievements and success.



<sup>3</sup> **Note:** In 1908, Einstein resubmitted his application to the University of Bern and was accepted. Nevertheless, his initial rejection seems puzzling for such a great scientist.

## An Example of a Test that Lacks ‘Content Validity’ (Optional)

Lesson	What a Teacher Actually Taught (A)	What the Test Evaluated (B)	A Students’ Actual Understanding (C)	How the Teacher’s Actual Test Represents the Student’s Understanding (D)
1	10%	30%	0%	0%
2	10%	10%	10%	10%
3	20%	20%	20%	20%
4	40%	20%	40%	20%
5	20%	20%	20%	20%
	100%	100%	??	??

### Assumptions:

- Column A indicates the amount of time that a teacher spent in teaching 5 lessons in a particular curricular unit.
- Let’s assume an excellent student achieved perfect understanding of all lessons except for Lesson 1 because he became very sick during that week and completely missed the lesson.
- Let’s assume that we have a special machine from the future that allows us to get a completely accurate depiction of the student’s understanding of the lessons (Column C)
- Let’s assume that the student gets perfect scores on the questions for Lessons 2 to 5 because he has ‘full’ understanding of these lessons; but let’s also assume that the student incorrectly answers all questions on Lesson 1 because he was absent (Column D)

### Questions:

- What is the ‘actual’ level of understanding of the student according to our machine from the future?
- What is the likely test result of the student according to the actual test that was administered?
- Does the likely test result (Column D) approximate the actual level of understanding of the student (Column C)?
- What can you conclude about the validity of this test?



## HANDOUT 12b: Reinforcement Exercise on Concepts of Validity

**Directions:** Read each of the short case study examples about test design and administration below and indicate the kind of validity that is implied in each case. Be prepared to discuss your responses in a plenary session. In some cases, more than one kind of validity may be implied.

### CASE STUDIES

1. Most conventional schools in the world tend to emphasize a limited number of intelligences in their curricular programming such as 'Verbal Intelligence' and 'Mathematical Intelligence.' If you are someone who has high 'Social Intelligence' or high 'Mechanical Intelligence', you will probably not do well on most of the tests administered at such schools even though you might be very successful in jobs that require these kinds of skills. What sort of assessment validity is implied in this case study?  

---
2. Most people know that Jack Ma, the Chinese billionaire who created *Alibaba*, was an academic failure during his time at university. What sort of assessment validity is implied by the evaluation that he experienced while at school?  

---
3. Some studies of educational achievement in Cambodia have found that there is often little relationship between the marks that students receive on their internal tests and the marks that they receive on externally administered examinations such as the Bac II Examination or the PISA test. What sort of validity is implied in this case study?  

---
4. Many Cambodian teachers prefer to use open-ended questions and essay questions, which require a great deal of time to answer, when designing their tests, even when they have a great many lessons to evaluate. What sort of validity might this habit affect in terms of student assessment?  

---
5. Cambodian teachers receive a great deal of pressure from development partners to change their teaching in a way so that they are teaching more higher order thinking skills such as creativity and evaluation. They are encouraged to use new methods of teaching such as 'project work' and 'problem-based learning'. However, when students are evaluated in their external examinations such as the Bac II Exam, they mainly encounter questions at the level of memory and understanding. What sort of validity is implied in this case study ?  

---

## **HANDOUT 13: Some Final Guidelines to Ensure Validity in Your Tests**

### **1. Be sure to test what you teach. .**

- If you taught memory, understanding, and application skills, be sure you have questions that cover all of these skill areas.
- If there is a lesson that you did not teach, do NOT include it on the test even though it may be in the textbook.

### **2. Cover all of the lessons in your test based on their priority and the amount of time you spent on them**

- If you taught 10 lessons, be sure that you test all 10, even if this means that you only sample the objectives from each lesson.
- When you have many educational objectives to assess, use question types that can be answered quickly so that you can cover as much content as possible in the time available.
- Remember that too many open-ended questions that require a lot of writing will constrain the *Content Validity* in the test because you will not be able to cover as much content as you would like to cover.

### **3. The lessons you spent the most time on should get the most questions or point weighting.**

### **4. Make sure that the conditions of assessment are the same in each classroom.**

- Different conditions from room to room will affect the *Concurrent Validity* of the test. For example, . . .
  - Make sure that you have taken equal precautions in all rooms to prevent cheating.
  - Make sure that the amount of time provided to students in each test room is the same.
  - Make sure that all invigilators have received the same instructions about how to administer the test.
  - Make sure that scoring guidelines are clear and consistent, especially if you are using open or subjective questions. One way to do this is to create a *model answer* for comparison purposes.

## HANDOUTS FOR PART 2

### **HANDOUT 14: Why Instructional Objectives are Important in Assessment**

#### **Writing Instructional Objectives is Like Planning a Road Trip**

**Trip:** Comparing educational objectives to a road map is a good way to understand their importance in the educational process. When one plans a road trip, one has to know one's ultimate destination before one can start the trip. That is, it makes no sense to start travelling if one has no idea where one is going.

→ Therefore, writing Instructional Objectives before one starts teaching and evaluating is akin to identifying one's destination when taking a trip.

#### **Identifying the Terminal Behaviors That Students Need to Exhibit**

**Exhibit:** One has to know what the ultimate behaviors that students should be able to exhibit are before we can start the teaching or assessment process. If teachers are not clear about what they want students to be able to do, it will nearly impossible to evaluate them. Thus, we should write instructional objectives by thinking about what we expect students to be able to do at the end of the lesson.

Writing Learning Objectives:



Beginning With the End in Mind

## Exercise: Planning a Trip with the Final Destination in Mind

***Imagine you need to take a trip. Between you and your final destination, there is a desert and a mountain range. What will you need to prepare and how will you use those preparations to get there.***



What Do You Need to Prepare?	How Will You Use It Along the Way?
<ul style="list-style-type: none"><li>•</li><li>•</li><li>•</li><li>•</li><li>•</li><li>•</li><li>•</li><li>•</li></ul>	<ul style="list-style-type: none"><li>•</li><li>•</li><li>•</li><li>•</li><li>•</li><li>•</li><li>•</li><li>•</li></ul>

## HANDOUT 15: The Role of Instructional Objectives in the Education Process

### Why is it important to use Educational Objectives?

Educational Objectives help educators to:

Guide  
Evaluation

- **Example:** Helps to guide questioning in the classroom, organize Tables of Specification, etc.

Plan  
Instruction  
Systematically

- **Example:** Helps to increase teachers' awareness of what to do when teaching to reach an explicit learning 'end point'

Provide  
Structure to  
Curriculum  
Development

- **Example:** Helps educators at national level to organize learning content in terms of both content and skills students must learn

## HANDOUT 16: Defining an Instructional Objective

**An Instructional Objective is defined as follows:**

*“ . . . . A clear and un-ambiguous description of the goals or changes in the student’s behavior that the teacher wishes to observe as a result of instruction.”*

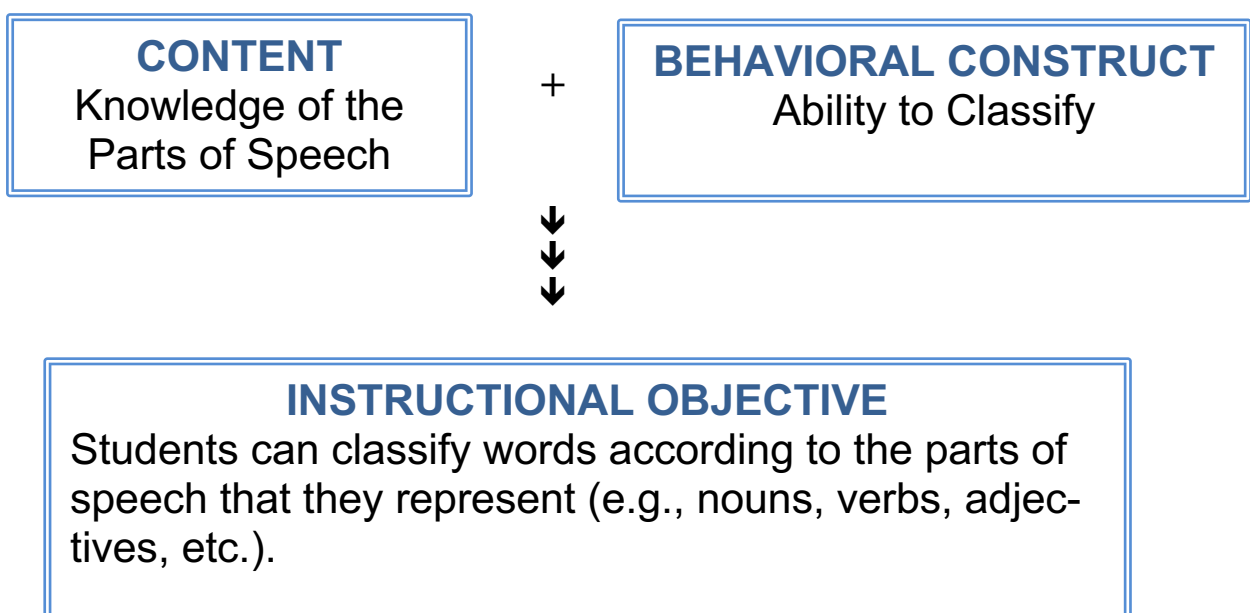


*Learning Objectives*

### Parts of an Objective:

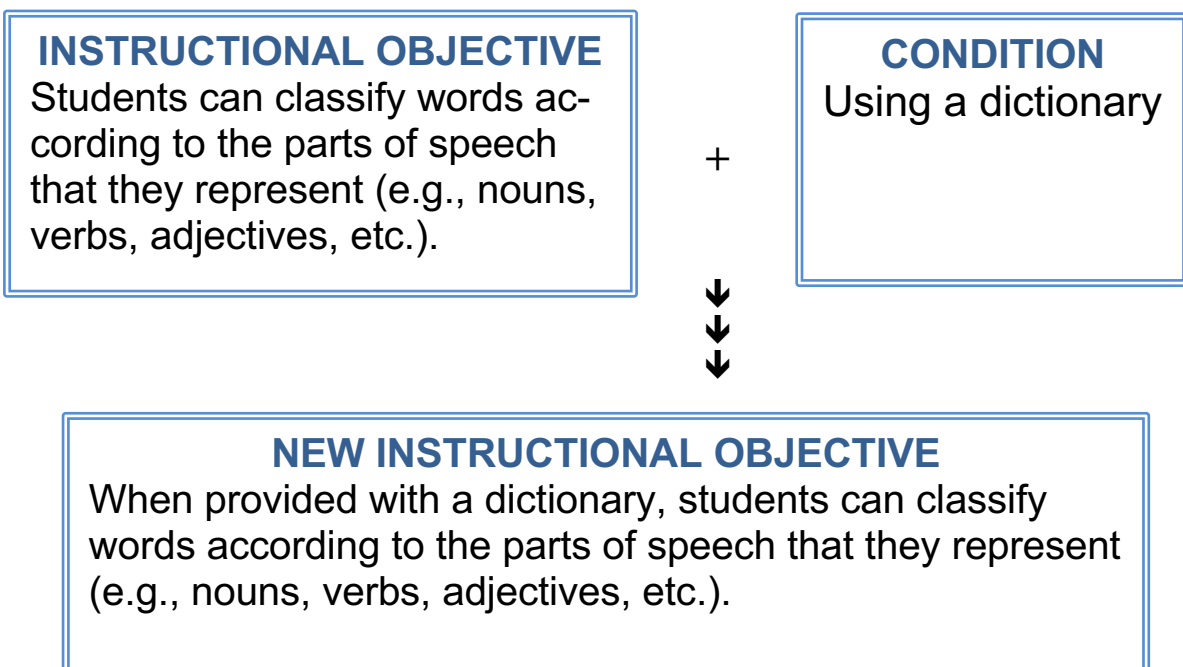
- An instructional should contain two or three components. These include the following:
  - The Content
  - The Behavioral Construct
  - The Condition(s) under which the learning should occur (OPTIONAL)

### Example:



## Including a Condition in the Objective . . .

- Sometimes, it is also possible to include a condition to further clarify the objective.
- The ‘condition’ often describes something that the teacher has to do that provides guidance about how the learning context is structured.
- This can include, . . .
  - Providing dictionaries for students to use
  - Watching a video or movie
  - Providing maps for a particular task
  - Setting up an experiment that students have to perform
  - Etc.
- **EXAMPLE . . .**



## However . . .

- Some educators like Gronlund (1985) advise not to use ‘conditions when writing instructional objectives because they reduce the flexibility of teachers in instruction.
- However, others encourage their use because they provide guidance to teachers about ‘how’ to teach the lesson.
- For this reason, it is up to the teacher whether they want to include conditions in their objectives.

### Exercise

**Directions:** Complete the following educational objectives using the conditions provided.

- When provided with a map, students can . . .
- Following an experiment on precipitates, students can . . .
- When given word problems relating to the multiplication and division of fractions, students can . . .
- After reading the story Thum Thiew, students can . . .
- When provided with a set of different atoms from an atomic toolkit, students can . . .



## HANDOUT 17: Kinds of Instructional Objectives

### What is a Good or Bad Objective?

- There are differences of opinion among educators regarding the form an objective should take.
- Consequently, it is sometimes difficult to make judgments about what is a 'good' or 'bad' objective.
- Our judgments will likely depend on the general educational philosophy that we hold.

### General & Specific Objectives

- One general view holds that instructional aims should be written in a 'general' form followed by multiple statements of observable learning outcomes.
- These observable learning outcomes are known as 'specific' objectives.
- Bloom's Taxonomy provides many useful examples of observable cognitive behaviors to use when writing objectives (see Handout 18).

## For example, . . .

### 1. *Students understand the meaning of written material (General Objective)*

1.1 Students can identify information explicitly stated in a passage.

1.1.1 Students can underline specific details in the passage as determined by the teacher (e.g., names, dates, etc.)

1.1.2 Students can select statements that best summarize the meaning of the passage.

1.1.3 Students can list facts that best support the major theme of the passage.

1.2 Students can summarize ideas in a passage.

1.2.1 Students can write a shortened version of the passage.

- In this example, the teacher starts with **general objectives** that use unobservable behaviors like ‘understand.’
- But then more **specific objectives** that use observable behaviors like **underline, select, list, and write** are added to clarify the general objective.

### On the other hand . . .

- Some educators feel that it takes too much time to write objectives in this amount of detail for a lesson plan.
- They suggest that it is only necessary to write the ‘specific’ objectives, as long as these state the observable behavior underlying the objective.
- This is especially true if the objective will be used for the classroom . . .



## HANDOUT 18a: Guidelines to Use When Writing Instructional Objectives

1. Always use verbs that demonstrate observable behaviors when writing 'specific objectives.' (e.g., determine, explain, summarize, write, list, etc.) (See Exercise below).
  - *Specifying observable behaviors makes it easier to assess whether the object has been achieved*
2. Avoid using words like *strengthen, encourage, support,* etc. in your objectives as these do not suggest any 'observable' cognitive behavior.
3. Never put two behavioral constructs into one objective. (e.g., *students can 'describe' and 'analyze' the characters of a novel*).
  - *Including two behavioral constructs makes it difficult to know if the objective has been achieved since one behavior may have been demonstrated but not the other.*
4. Ensure that your objectives include a mix of both lower and higher order thinking skills. Use Bloom's Taxonomy to help you know which objectives reflect higher order skills and which reflect lower order skills.
5. Classify each objective by the level that it implies in Bloom's Taxonomy (e.g., Remembering, Understanding, Applying, etc).



## HANDOUT 18b: Using Bloom’s Taxonomy to Develop and Classify Instructional Objectives

Bloom’s Taxonomy is the best tool to use to better understand the various thinking skills that an Instructional Objective may convey. There are 6 levels in the Taxonomy as described below.

LEVEL	DEFINITION	SAMPLE VERBS		SAMPLE OBJECTIVES
<b>REMEMBERING</b>	Student recalls or recognizes facts, concepts, and principles in the approximate form in which they were learned.	Write List Label	Name State Define	Students can <b>define</b> the 6 levels of Bloom's taxonomy of the cognitive domain.
<b>UNDERSTANDING</b>	Student translates, comprehends, or interprets information based on prior learning.	Explain Summarize Paraphrase	Describe Illustrate Give examples	Students can <b>explain</b> the purpose of Bloom's taxonomy of the cognitive domain.
<b>APPLYING</b>	Student can use one or more principles to solve a problem or task with a minimum of direction.	Use Compute Solve	Demonstrate Apply Construct	Students can <b>write</b> an instructional objective for each level of Bloom's taxonomy.
<b>ANALYZING</b>	Student distinguishes, classifies, and relates the assumptions, hypotheses, evidence, or structure of a statement or question.	Analyze Categorize Compare	Contrast Separate	Students can <b>compare</b> and <b>contrast</b> the cognitive and affective domains.
<b>EVALUATING</b>	Student appraises, assesses, or critiques on a basis of specific standards and criteria.	Judge Recommend	Critique Justify	Students can <b>judge</b> the effectiveness of writing objectives using Bloom's Taxonomy.
<b>CREATING</b>	Student originates, integrates, and combines ideas into a product, plan or proposal that is new to him or her.	Create Design Hypothesize	Invent Develop	Students can <b>design</b> a classification scheme for writing educational objectives that combines the cognitive, affective, and psychomotor domains.

## Exercise: Thinking Skill Classification Exercise

**Directions:** Review each of the verbs below and indicate what level of Bloom's Taxonomy that you think they apply to. One verb may be appropriate to more than one level in the taxonomy depending on how it is implemented. Please indicate all levels that apply to the verb provided.

English	Khmer	Taxonomy Level	English	Khmer	Taxonomy Level
Adapt			Explain		
Answer			Formulate		
Analyze			Identify		
Apply			Illustrate		
Arrange			Indicate		
Calculate			Interpret		
Categorize			Judge		
Clarify			Label		
Classify			List		
Combine			Match		
Compare			Measure		
Complete			Name		
Compose			Outline		
Contrast			Predict		
Create			Recite		
Critique			Rephrase		
Defend			Report		
Define			Select		
Demonstrate			Solve		
Describe			Specify		
Determine			State		
Diagram			Summarize		
Differentiate			Synthesize		
Distinguish			Tell		
Enumerate			Use		
Evaluate			Write		

## Exercise: Try This:

**Directions:** Below you will find some behavioral objectives for various subjects. Using Bloom's Taxonomy, identify the level of thinking at which you think the behavioral objective is taking place. Write **R** for Remembering; **U** for Understanding, **Ap** for Applying, **An** for Analyzing, **E** for Evaluating and **C** for Creating in the blank in front of each objective.

- \_\_\_\_\_ 1. Students can define the following terms in their own words providing suitable examples of each.  
a) planet    b) comet    c) star    d) satellite
- \_\_\_\_\_ 2. Students can state the capitals of all of the countries of South-east Asia.
- \_\_\_\_\_ 3. When given an ordered pair, student can plot the point that this pair represents on a graph.
- \_\_\_\_\_ 4. Students are able to determine the kinds of products made in different parts of Cambodia by consulting political and resource maps concurrently.
- \_\_\_\_\_ 5. Given a list of physical conditions, students can predict the changes that might occur in the state of water based on an understanding of its physical properties.
- \_\_\_\_\_ 6. Students can express an opinion regarding the merits and demerits of socialism and capitalism being sure to provide a justification for any position taken.
- \_\_\_\_\_ 7. Students can develop a project that demonstrates hydraulic principles in Physics.
- \_\_\_\_\_ 8. Students can construct a chart that summarizes the differences between Chinese and Indian civilization.
- \_\_\_\_\_ 9. When given a sentence that is grammatically incorrect, students can correct the sentence using rules of grammar that they have learned.
- \_\_\_\_\_ 10. Students can explain the chemical composition of a compound in words based on its chemical notation.

**Answers**  
1. U; 2. R; 3. Ap; 4. An; 5. Ap; 6. E; 7. C; 8. An; 9. Ap; 10. U

## Exercise: Try This:

Look at the subject areas below and expand the general learning area provided into a series of 'specific' behavioral objectives. Use the guidelines that were discussed earlier to make sure that your objectives are written properly. Write at least ONE Specific Objective for each content area provided.

- **English:** Vocabulary Words
- **Mathematics:** The idea of a set
- **Science:** The elements of the Periodic Table
- **Khmer:** The differences and similarities of characters in the story, Thum-Thiew
- **Geography:** Geographical maps
- **History:** The causes of World War II

## HANDOUT 19: Developing Tables of Specifications

### What is a Table of Specifications?

*A Table of Specifications is defined as a test blueprint that enables the development of a test, which is 'balanced' and 'relevant' in terms of the content and thinking skills taught.*

### Why Use a Table of Specifications?

#### Ensures Robust Validity

- The table helps to summarize content areas and thinking skills that have been taught so that tests accurately reflect those content areas and thinking skills.

#### Promotes Fairness

- The table helps to remind the teacher what he/she taught so that students experience a test that evaluates the things that they actually studied.

#### Rationalizes Accurate Weighting of Questions

- The table helps to ensure that areas taught for many hours receive more questions than areas that were taught for fewer hours.



## What Does a Table of Specifications Look Like?

Construct Content	Remembering	Understanding	Applying	Analyzing	Evaluating	Creating	Total
<b>Total</b>							

### How Does One Fill in a Table of Specifications?

- 1) Determine the content areas that you want to test. These can be lesson names or topics from one lesson.
- 2) Reflect on how you taught each content area. Did you ask students to simply remember things, understand things, apply principles, etc.?
- 3) Fill in the numbers in your table:
  - a. The numbers can be either the **number of questions** for each subject area/construct or the **number of points** on the test for each area.
  - b. If the numbers represent questions, all questions should be worth the same number of points.
  - c. If the teacher wishes to develop a test in which the questions have different point values, then the numbers in the Table of Specifications should represent 'points' and not the number of 'questions.'

4) Make totals for your content and constructs. The number of points or questions that you have indicated for each topic and construct should reflect the amount of time and emphasis that you placed on each topic/construct when teaching. If it does not, then there is a problem with the test's Content or Construct Validity.

### Case Study Exercises

**Case Study 1:** A teacher taught a unit with 5 lessons. Each lesson was equally important, and the teacher spent the same amount of time teaching each lesson. However, the teacher developed a test in which 80% of the questions were on the last lesson. Is there a problem with this test?

**Case Study 2:** A Biology Teacher taught several lessons on disease transmission focusing primarily on memorization of facts and concepts. But when she made her test to assess students, she included many questions at the level of understanding and application. Is there a problem with this test?

**Case Study 3:** A Physics Teacher taught three lessons on Motion and Force. One student demonstrated excellent understanding of the first two lessons but then fell sick and was absent while the teacher taught the third lesson. When the student took a test on the unit, she found that 90% of the questions were on the third lesson. Will this test accurately reflect the student's understanding of what was taught? Why or why not? What kind of validity will be in question?

## An Example of a Table of Specifications:

Subject: Khmer Language

Unit: 5

Grade Level: 10

Construct Content	Remembering	Understanding	Applying	Analyzing	Evaluating	Creating	Total
1) Vocabulary	5	5					10
2) Grammar		10		10			20
3) Poetry			10				10
4) Composition			10				10
<b>Total</b>	<b>5</b>	<b>15</b>	<b>20</b>	<b>10</b>			<b>50</b>

### Questions for Discussion

- 1) How many content areas are there in this test?
- 2) How many skills are measured in this test?
- 3) Which content area has the most emphasis? The least emphasis?
- 4) Which construct area has the most emphasis? The least emphasis?
- 5) If the numbers in the table represent the number of questions, how much is each question worth on a test with 100 points.

## HANDOUT 20: About Objective & Subjective Test Questions

**Explanation:** In Cambodia, questions are usually defined in terms of their format, that is whether they are **open** or **closed**. However, it is more customary to define questions in terms of HOW THEY ARE SCORED. We call such questions **Objective** or **Subjective** Questions.

### Objective Questions

**Definition:** Questions that are scored dichotomously, i.e., they are either completely right or completely wrong.

#### Kinds of Objective Questions:

- Fill in the Blank
- Multiple Choice
- Matching
- True or False
- Classification

### Subjective Questions

**Definition:** Questions where the scoring may vary from scorer to scorer; students may receive a full score, partial score, or no score.

#### Kinds of Subjective Questions

- Extended Response Essay
- Restricted Response Essay

## HANDOUT 21: How to Write Objective Test Questions Effectively

### *What is an Objective Question?*

*An Objective Question is a kind of question for which there is only ONE correct answer. When scored, the question must be evaluated as either completely right or completely wrong.*

## Characteristics of Objective Questions

### Format of Objective Questions

- Such questions have multiple formats including Fill-in the blank, Matching, True-False, and Multiple Choice.

### Scoring Objective Questions

- Scoring is done 'dichotomously', meaning that objective questions are marked completely right or completely wrong.
- There is only ONE Correct Answer

### Advantages

- Effective for covering a lot of content
- Students can answer quickly leading to time efficiencies
- Easy to correct with high inter-rater reliability
- Can be used to measure constructs at the level of Memory, Understanding, Application, and Analysis

### Disadvantages

- Cannot be used to measure creativity or synthesis level skills
- Cannot be used to measure process skills
- Difficult to write

## The Short Answer Question

### Different Varieties

- **The Question Variety**
  - *How many provinces are there in Cambodia? \_\_\_\_\_.*
- **The Completion Variety**
  - *In geography, positions of north, south, east, and west are expressed in degrees of \_\_\_\_\_ and \_\_\_\_\_.*
- **The Association Variety**
  - **Directions:** In the blanks provided, write the name of the Mathematical Laws of Equality that are demonstrated:
    - 1.  $3 \times 2 = 2 \times 3$  \_\_\_\_\_
    - 2.  $(2 + 3) + 5 = 2 + (3 + 5)$  \_\_\_\_\_
    - 3.  $6 \times (3 + 4) = (6 \times 3) + (6 \times 4)$  \_\_\_\_\_

### When to Use It

- The Short Answer Question is generally used to test Memory Level Questions only.

### Suggestions When Writing Short Answer Questions

- 1) Be sure that the question can be answered only by a unique word, phrase, or number. Look at the following example:
  - *The Cambodian Parliament meets in \_\_\_\_\_.*
  - *There are multiple possible answers to this question including January, Phnom Penh, or emergencies. How would you revise this question?*
- 2) Provide as many blanks or spaces as there are words in the answer. For example:
  - *Indochina is made up of the countries of \_\_\_\_\_, \_\_\_\_\_, and \_\_\_\_\_.*
- 3) Do not delete too many words from a statement to be completed. For example:
  - *The old \_\_\_\_\_ of Cambodia was \_\_\_\_\_.*

# The True False Question

## General Information

- The True False Question most typically consists of a declarative statement containing an assertion that the student must decide is true or false.
- Such statements are difficult to write because it is often hard to find statements that are unequivocally true or false. Sometimes, even experts disagree.
- In general, most teachers feel that True False Questions are easy to write when in fact the opposite is true.

## Different Varieties

- **The Basic Variety (True or False)**
  - *T F A number can be divided by '0'.*
  - *T F Air occupies space.*
  - *T F Battambang is the third largest city in Cambodia.*
- **The Correction Variety**
  - *T F Angkor Vat was built by Jayvaraman VII.*
  - \_\_\_\_\_
  - *In this variety, if the statement is incorrect, students must replace the underlined word in the blank provided below to make it correct.*
- **Right Wrong Variety**
  - *R W Running down the hill, the ball fell between my legs and tripped me.*
- **Fact Opinion Variety**
  - *F O The Angkorian Period was the most famous period of Cambodian history.*
  - *F O Less than half of Cambodians can speak English.*

## When to Use True False Questions

- The True False Question is useful to measure a wide range of thinking skills including Memory, Understanding, Application, and Analysis.
- On very important summative examinations, the use of True-False Questions should be avoided because of the high probability of guessing, which affects concurrent validity.

## Suggestions When Writing True False Questions

- 1) Do not make True False Questions too long as this introduces the irrelevant factor of reading ability into the process of answering.
- 2) Avoid using words known as 'specific determiners' in True False Questions. This includes words such as all, none, never, always, etc. Using such words makes it likely that such statements are false since there are always exceptions.
- 3) Do not use negatively worded True False statements as this will make the question confusing. For example:
  - *T F Oxygen is not one of the elements to be found in salt.*
- 4) Test only one point in each question. Do not have two points in a statement some of which are true while others are false. For example:
  - *T F Sound can travel through air, liquids, solids, and vacuums.*
  - *In this statement, some items are true while one is false.*
- 5) Do not use statements of opinion in True False Questions unless you attribute the source. For example:
  - *T F The Angkorian Period was the most famous period of history in Cambodian history.*
  - *T F According to many historians, the Angkorian Period was the most famous period of history in Cambodian history.*
- 6) Avoid systematic patterning of answers.
- 7) Use a somewhat larger number of false items than true items on a test.



# Multiple Choice Questions

## General Information

- The Multiple Choice Question is the most flexible of the objective question formats.
- It can be used to test any type of thinking skill from Memory to Analysis Level.
- It is superior to True False Questions in that its multiple response format minimizes the probability of correct responding due to guessing.
- One of its major advantages is that it allows assessment of complex thinking skills without requiring the skill of writing. This greatly enhances the ability of a test to enhance its content and construct validity.

## Basic Format of Multiple Choice Questions

- **The Stem**
- The stem of a multiple choice question is typically worded as a direct question or an incomplete sentence. For example:
  - *What is the most populated province in Cambodia?*
    - a) Battambang    b) Takeo    c) Phnom Penh    d) Kandal
  - *The most populated province in Cambodia is:*
    - a) Battambang    b) Takeo    c) Phnom Penh    d) Kandal
- Whatever form the stem takes, it **MUST** be able to elicit appropriate responses from students. Look at the following example of a faulty stem:
  - *Food Chains*
    - a) are combinations of proteins, carbohydrates and fats.
    - b) are illustrations which show the final destination of farm products.
    - c) are diagrams that help us to understand how prey and predators are related.
    - d) are illustrations that show the relationship between plants and animals.

## Basic Format of a Multiple Choice Question (Cont.)

- **The Responses**
- The second major part of the Multiple Choice Question refers to the 'responses' that are placed under the stem.
- There are normally 4 or 5 possible responses in a Multiple Choice Question.
- Using less than 4 responses is not advised because this adversely affects the *reliability* of the question.
- Responses can be indicated by letters or numbers. Where possible, responses should be arranged in a logical order where one exists (e.g., alphabetical, lowest to highest, etc.).
- In most cases, it is advisable to arrange responses vertically instead of horizontally (except in cases of very short responses). For example:
  - *Why has Africa often been called the Dark Continent?*
    - *a) Many of the people of Africa have dark skin b) Very little was known about Africa for a long time c) The jungles of Africa are dark and forbidding d) Most of Africa is covered with black earth soils*
  - *Why has Africa often been called the Dark Continent?*
    - *a) Many of the people of Africa have dark skin*
    - *b) Very little was known about Africa for a long time*
    - *c) The jungles of Africa are dark and forbidding*
    - *d) Most of Africa is covered with black earth soils*
- Clearly, the second form of the question is much easier to read.

## Basic Format of a Multiple Choice Question (Cont.)

- **Punctuation of Multiple Choice Questions**
- Where the stem of a multiple choice question is an incomplete sentence, there is usually no punctuation at the end of the stem. Each response, however, should have appropriate punctuation at its end.
- Where the stem is a question, it should always have a question mark (?) at the end.
- The responses of questions having incomplete sentences as stems should begin with lower case letters. If the stem is a question, responses should begin with upper case letters.

## Different Varieties

- **The Correct Answer Variety**
- This is the most commonly used kind of a Multiple Choice Question. In such questions, there is only ONE completely correct response while all of the other responses are completely wrong. For example:
  - *My mother took 5,000 Riels to the market. She bought 800 riels of fish and 2,400 Riels of pork. How much did she have left?*
    - *a) 1,500 Riels*
    - *b) 1,800 Riels*
    - *c) 1,850 Riels*
    - *d) 1,870 Riels*

## Different Varieties (Cont.)

- **Best Answer Variety**
- In this kind of question, one or more responses may have some element of truth in it. However, students must choose the question that is the MOST correct.
  - *Which of the following things best symbolizes Cambodia?*
    - a) *The Independence Monument*
    - b) *Angkor Vat*
    - c) *The Royal Palace*
    - d) *The Sugar Palm Tree*
- In the example, all of the responses are correct to some degree but only one is the MOST correct.
- **The Negative Answer Variety**
- In this variety of question, all of the responses are correct except one. Students must choose the response that is NOT correct. For example:
  - *All of the following statements are characteristics of corn EXCEPT*
    - a) *Corn is a seasonal crop.*
    - b) *Corn is a cash crop.*
    - c) *Corn is a self-fertilizing crop.*
    - d) *Corn powder can be used to make cakes.*

## Different Varieties (Cont.)

- **The Multiple Answer Variety**
- *In this variety of question, students must choose a response or combination of responses that are correct. For example:*
- *Which of the following picture(s) represent a vertebrate?*



I



II



III



IV



V

- a) *I only*
- b) *II, III, and IV*
- c) *II, IV, and V*
- d) *I and III*

## Suggestions When Writing Multiple Choice Questions

- 1) Make sure that the Stem of each question contains a clearly defined problem (see earlier example).
- 2) Do not use more words than necessary to make your meaning clear. For example:
  - **(Weak)**
  - *Phnom Penh*
    - *a) is located at the intersection of two rivers.*
    - *b) is located in Kandal Province.*
    - *c) is located in Siem Reap Province.*
    - *d) is located on Tonle Sap Lake*
  - **(Better)**
  - *Phnom Penh is located*
    - *a) at the intersection of two rivers.*
    - *b) in Kandal Province.*
    - *c) in Siem Reap Province.*
    - *d) on Tonle Sap Lake.*
- 3) Make sure that your responses are parallel in grammatical structure and that the wording is appropriate to the stem. For example:
  - **(Weak)**
  - *Why do we give vaccinations to children against small pox?*
    - *a) Because some children have the disease*
    - *b) To prevent them from getting the disease.*
    - *c) For experimental purposes.*
    - *d) Stop the spreading of germs.*
  - **(Better)**
  - *We need to give vaccinations to children against small pox to:*
    - *a) cure children who already have the disease.*
    - *b) prevent them from getting the disease.*
    - *c) experiment on methods to prevent the disease.*
    - *d) stop the spread of germs that cause the disease.*

## Suggestions When Writing Multiple Choice Questions (Cont.)

- 4) Do not allow your responses to overlap so that one response includes one or more other responses. For example:
  - *Butterflies are a member of which of the following animal groups?*
    - a) *Invertebrates*
    - b) *One-celled animals*
    - c) *Insects*
    - d) *Annelida*
- 5) Try to keep all responses the same length as there is a tendency for teachers to make the longest response the correct response.
- 6) Avoid the use of responses such as 'None of the above' or 'All of the above.'
- 7) Vary the position of the correct response from question to question.
- 8) Do not allow the stem to provide an irrelevant clue as to the correct response. This hurts the overall validity of the examination. Look at the following examples:
  - **(Weak)**
    - *What major factors cause the seasons to change on the earth?*
      - a) *The rotation of the earth.*
      - b) *The wind from the oceans*
      - c) *The revolution of the earth around the sun and its tilted axis.*
      - d) *The distance of the earth from the sun*
  - **(Weak)**
    - *What is an equation?*
      - a) *A mathematical expression where one side is equal to the other side.*
      - b) *A number sentence with at least two members on either side.*
      - c) *Two number sentences with similar value.*
      - d) *Any number sentence that does not contain the signs (<) and (>).*

## The Matching Question

### General Information

- The Matching Question is a variation of the Multiple Choice Question.
- The Matching Question consists of items each of which has the same set of alternative responses. Look at the following example:
- **Directions:** Choose the name of the country on the right that corresponds to the name of the famous building on the left. Write the letter of the famous building in front of the name of the country in the space provided. A country may be used more than once or not at all.
  - \_\_\_ 1. *The Great Wall of China*     a) *France*
  - \_\_\_ 2. *The Parthenon*                     b) *China*
  - \_\_\_ 3. *The Pyramids*                         c) *Egypt*
  - \_\_\_ 4. *Stonehenge*                             d) *Greece*
  - \_\_\_ 5. *The Sphinx*                             e) *England*
- One of the major advantages of the Matching Question is its efficiency. If the above exercise had been written as a multiple choice question, it would have taken up considerably more space.

### When Is It Used

- The Matching Question is generally limited to assessing Memory Level skills.



## Suggestions When Writing the Matching Question

- 1) Write very precise directions when developing a Matching Question. Do not simply say, 'Match the following.'
- 2) Always use the longer of the elements to be matched and shorter elements as the responses. This helps to reduce the amount of time students need for reading the question.
- 3) Use short lists of phrases to be matched. The elements to be matched should never exceed 10 in number.
- 4) Try to use more responses than phrases to be matched.
- 5) Arrange the responses in a logical order (e.g., alphabetical, chronological, etc.) if such an order exists.
- 6) Use items and responses that are similar. Do not mix apples and oranges. Here is an example of a faulty question:
  - \_\_\_ 1. *Date of India's independence*     a) *Nile River*
  - \_\_\_ 2. *Largest river in the world*     b) *Amazon River*
  - \_\_\_ 3. *Longest river in North America*     c) *Mississippi River*
  - \_\_\_ 4. *Longest river in the world*     d) *1947*
  - \_\_\_ 5. *Famous African explorer*     e) *Livingstone*
- 7) Never spread a Matching Question over 2 pages as this causes students to lose time in answering.
- 8) Unless required by the objectives one is evaluating, do not make wide use of the Matching Question in your test because it is only effective in assessing Memory Level objectives.

## The Classification Question

### General Information

- The Classification Question is similar in format to the Matching Question.
- It has one major advantage over the Matching Question in that it can be used to measure higher level thinking skills including Application and Analysis.

### Format of the Classification Question

- The format of the Classification Question requires students to use two or more classifying principles to classify a set of factors or statements. For example:
  - **Example 1:**
  - **Directions:** Items 1 to 5 consist of living and non-living things. You must choose the grouping describing these items by placing the letter 'a' or 'b' in the blank provided to the left of the number items.
    - \_\_\_ 1. Rock
    - \_\_\_ 2. Water
    - \_\_\_ 3. Grass
    - \_\_\_ 4. Car
    - \_\_\_ 5. Rabbit
  - a. Living Thing
  - b. Non-living Thing

## HANDOUT 22: Defining Subjective Test Questions and Understanding Their Limitations

### *What is a Subjective (Essay) Question?*

*Subjective Questions generally take the form of what is known in English as an Essay Question. An Essay is a 'free response' question that requires students to produce a written response in sentence or paragraph form, rather than to select the correct response from a number of alternatives or to generate a short word or phrase. Subjective Questions are not scored dichotomously and point awards for a student's response may vary from scorer to scorer.*

*There are TWO TYPES of Essay Questions that depend on the amount of freedom given to the student to answer the question:*

#### Extended Response Essay Question

- In this type of essay question, few, if any limits are placed on the student in terms of the selection of pertinent information, the organization of the answer, or the integration of ideas. This lack of structure introduces scoring difficulties because there will be little consistency among students in the length or quality of their responses. However, this type of question also allows a teacher to find out how well students perform tasks at the Evaluation and Creation level.
- **Example:** Evaluate the impact of Cambodia's contact with Western countries on its subsequent development.

#### Restricted Response Essay Question (Short Paragraph)

- In this type of essay question, specific guidelines are provided to limit both the content and form of student responses. This greater structure makes marking easier, but does not provide a measure of students' ability to evaluate and create information. Restricted Response type questions are best for measuring learning outcomes at the Level of Understanding, Application, and Analysis.
- **Example:** Analyze the cause of the Fall of Angkor in terms of the following Factors: (a) Political Factors; (b) Social Factors; and (c) Economic Factors.

## The Need for Caution When Using Essay Questions

### Sampling Coverage

- A test made up entirely of essay questions presents problems in terms of how well the test can cover or sample the course content.
- Thus, 'Sampling Reliability' is low.
- Limit the use of Essay Questions only to those topics that can not be evaluated with objective questions.

### Time

- Because Essay Questions take so long to answer, they offer limited coverage of course material.
- This negatively affects Sampling Reliability.

### Sampling Reliability

- If you want to use Essay Questions exclusively on a test, you may increase Sampling Reliability by using several shorter Essay Questions rather than 'Extended Response Essay Questions.'

### Directing Words

- Directing Words (e.g., Summarize, Compare, Explain, etc.) are very important in writing Essay Questions.
- Teachers must choose directing words carefully and ensure that students understand these words well.

### Setting Test Expectations

- Essay Questions should provide students with clear expectations.
- These expectations include: 1) Criteria for Evaluation; 2) Time Limits; 3) Marking Values; and 4) Length of Answers.

## HANDOUT 23: The Use of Directing Words When Writing Essay Questions

### Using 'Directing Words' in Essay Questions

#### *Directing Words*

- *The Directing Word in an Essay Question tells the student what to do:*  
**Explain . . . . Summarize . . . . Compare . . . . Etc.**
- *If students do not understand what the directing word tells them to do, their answers will suffer in quality.*
- *It is very important that teachers choose directing words carefully and ensure that their students know what these words mean. This may require pre-teaching or providing definitions of these words on the test itself.*
- *It is advised that the word 'Discuss' never be used as a directing word because it is too vague.*

Teachers need to be careful in their selection of Directing Words when writing Essay Questions. Some common examples of Directing Words and their meaning are provided in the table below.

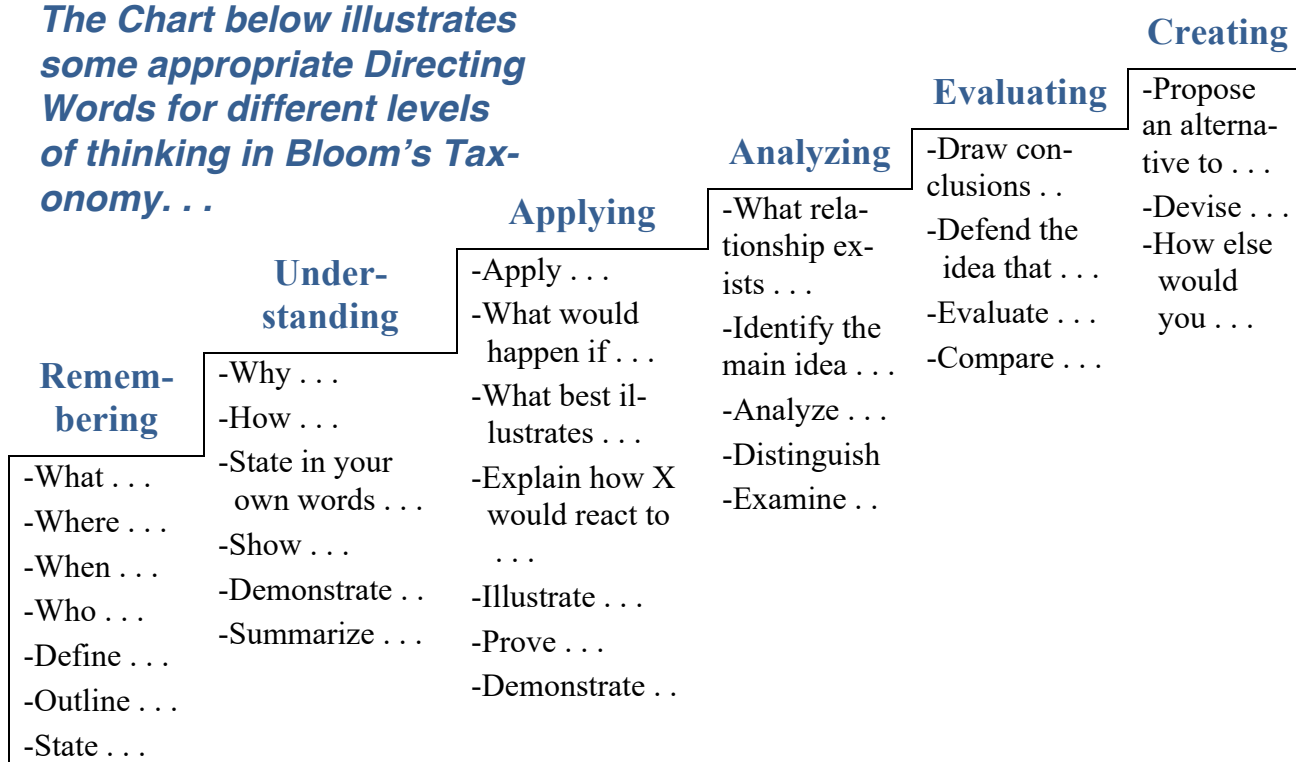
#### *Defining Some Useful Directing Words*

<b>Contrast</b>	<i>Bring out the points of difference.</i>	<b>Illustrate</b>	<i>Use a picture, diagram, chart, or concrete example to clarify a point.</i>
<b>Compare</b>	<i>Bring out points of similarity and difference.</i>	<b>Interpret</b>	<i>Make plain, show your thinking about something.</i>
<b>Criticize</b>	<i>State your opinion of the correctness or merits of an issue</i>	<b>Justify</b>	<i>Show good reasons for something</i>

<b>Define</b>	<i>Give the meaning of a word or concept by placing it in the class to which it belongs and describing it is different from other members in the same class.</i>	<b>Outline</b>	<i>Give, in order, the main points about something; sketch in general terms.</i>
<b>Describe</b>	<i>Give a picture of something.</i>	<b>Prove</b>	<i>Establish the truth of something by giving factual information.</i>
<b>Evaluate</b>	<i>Give the good and bad points of something.</i>	<b>Summarize</b>	<i>Briefly review the main points of something</i>
<b>Explain</b>	<i>Make something clear, Interpret something</i>	<b>Trace</b>	<i>Give a description of the development of something</i>

## Directing Words and Bloom's Taxonomy

*The Chart below illustrates some appropriate Directing Words for different levels of thinking in Bloom's Taxonomy. . .*



## Some Useful Guidelines in Choosing Directing Words . . . .

- The diagram above lists some Directing Words and the levels of thinking they generally call for. Steps may easily overlap; for example, Application must include Memory and Comprehension.
- It is important to remember that just changing the Directing Word in an Essay Question does not automatically raise the level of the thinking required to answer the question.
- Once again, even though it is possible to formulate Essay Questions at the Memory and Understanding Level, teachers are urged to only use Essay Questions at the Application Level or higher. Memory and Understanding skills are better assessed by Objective Questions.

## Exercise: Try writing some Essay Questions . . .

**Directions:** Consider the famous folktale, *Tum Tiew*. In the space provided below, write an essay question to assess students' understanding of *Tum Tiew* at each of the levels of thinking described earlier. Consider your choice of Directing Words as you write a question for each level of thinking.

MEMORY: \_\_\_\_\_

\_\_\_\_\_

UNDERSTANDING: \_\_\_\_\_

\_\_\_\_\_

APPLYING: \_\_\_\_\_

\_\_\_\_\_

ANALYZING: \_\_\_\_\_

\_\_\_\_\_

EVALUATING: \_\_\_\_\_

\_\_\_\_\_

CREATING: \_\_\_\_\_

\_\_\_\_\_



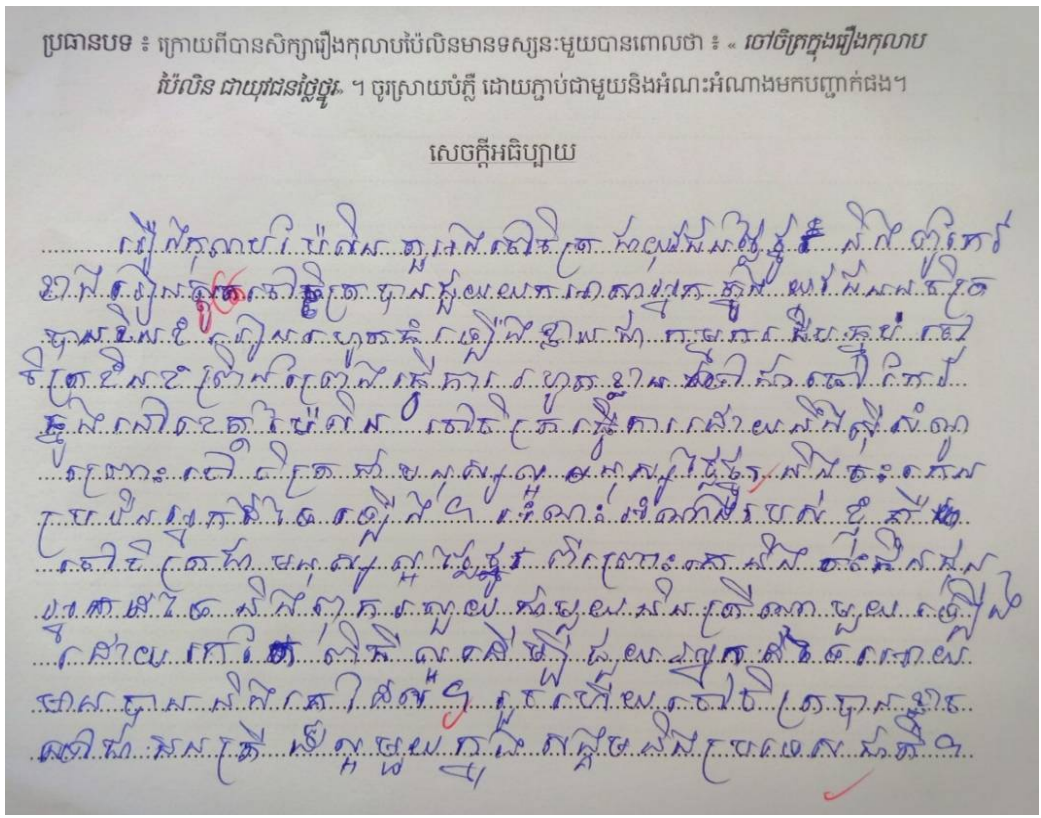
# HANDOUT 24: Challenges in Scoring Essay Questions

## Reliability in scoring Essay Questions is a serious and persistent problem!

Research shows that Essay Questions are scored differently by different teachers and that the same teacher will score an Essay Question differently at different times.

In one study, 100 teachers scored an Essay Question written by a Grade 12 student. Their task was to assign a numerical mark and then to judge what grade they thought the student was in. The numerical marks they assigned ranged from 60% to 98% and their estimated grade placements ranged from Grade 5 to second year university!

**Try This:** Look at the Essay Question Response written by a student in Grade 9. Read the student response and give a score on a scale of 1 to 10 based on the criteria that you usually use to score Essay Question Responses.



## Follow-up and Questions for Discussion

After giving your score, make a comparison with the scores given by other participants.

- How great was the variation between scores?
- Where do the scores tend to cluster? Why do you think that they cluster there?
- What could be done to reduce any variation that was observed?

<b>Points Awarded</b>	<b>How Many Participants Gave This Score?</b>
<b>1</b>	
<b>2</b>	
<b>3</b>	
<b>4</b>	
<b>5</b>	
<b>6</b>	
<b>7</b>	
<b>8</b>	
<b>9</b>	
<b>10</b>	

## Two General Approaches for Marking Essay Questions

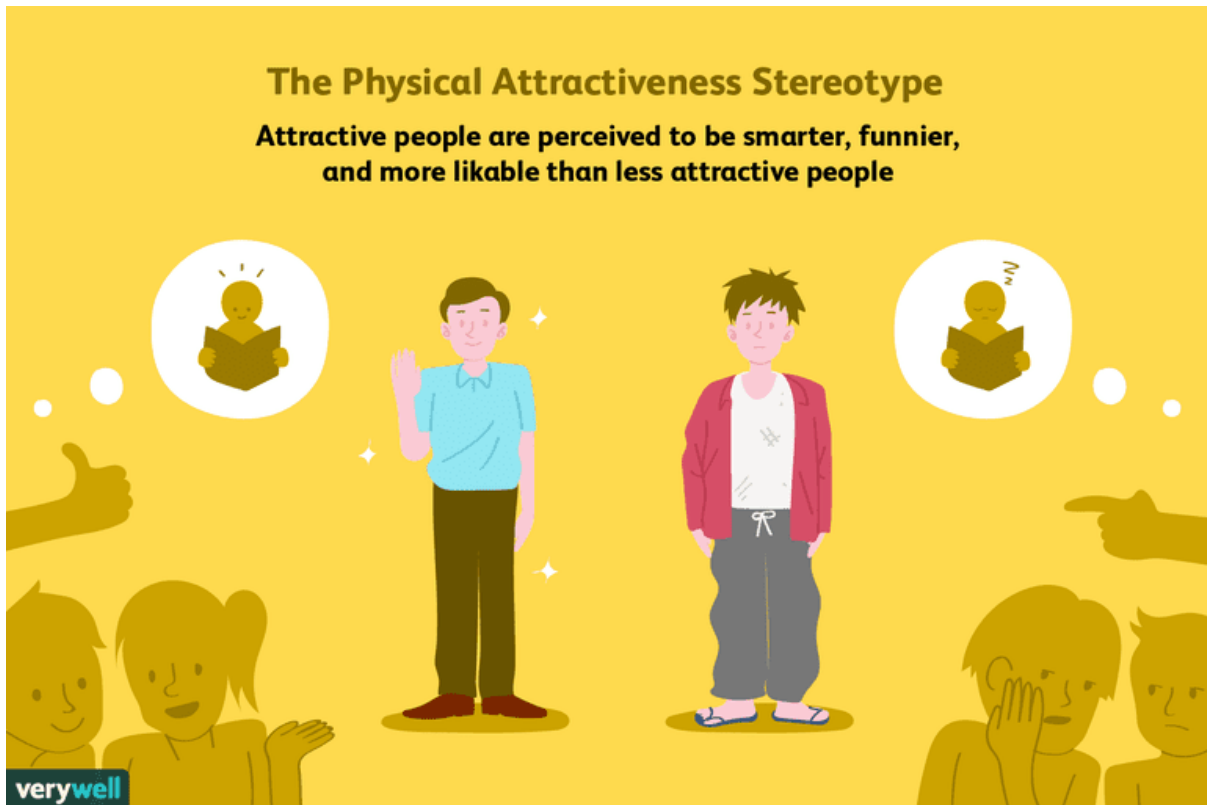
### Analytical Method (Point Score Method)

- A 'model' or 'ideal' answer is prepared beforehand and broken down into specific segments, each of which is assigned a point value. In addition to knowledge content, things like 'effectiveness of expression,' 'support of statements,' 'logical organization' are specified and assigned point values. This method uses a kind of 'checklist' for scoring, which introduces some objectivity into the process.
- **Advantages:**
  - Increases the reliability of the scoring
  - Helps control for 'halo' effect
  - Makes it easier for the teacher to discuss and justify marks with students and parents
- **Disadvantages**
  - Can be laborious and time-consuming to prepare the checklist to score the responses
- **Recommended for:**
  - Extended Essay Questions

### Global Method (Holistic Rating Method)

- Under this method, the 'model' or 'ideal' answer is not broken down into specific points, but serves as a standard for the 'best' answer. The teacher then selects from among the students' papers, samples that will serve as 'anchor points' for a set of rating categories such as Excellent, Above Average, Average, Below Average, and Poor. The teacher reads each student's answer and forms a general impression and then assigns it to one of the rating categories.
- **Advantages**
  - Simpler and quicker to use than the Analytical Method
- **Disadvantages**
  - More subjective than the Analytical Method
  - No Clear justification for the assigned grade
  - No specific feedback to students about areas of weak understanding
- **Recommended for:**
  - Shorter Essay Questions, which are likely to elicit detailed and uniformly structured answers

# The 'Halo Effect' & Scoring Essay Questions



## What Is the Halo Effect?

The halo effect is a type of [cognitive bias](#) in which our overall impression of a person influences how we feel and think about his or her character. Essentially, your overall impression of a person ("He is nice!") impacts your evaluations of that person's specific traits ("He is also smart!").

One great example of the halo effect in action is our overall impression of movie stars. Since we perceive them as attractive, successful, and often likable, we also tend to see them as intelligent, kind, and funny.

When correcting Essay Questions, teachers are often swayed in their assessments by the quality of handwriting, the sophisticated use of language, and neatness. In such cases, students often get good marks even if their answers lack substance.

## HANDOUT 25: Overview of Essay Questions

Factor	Characteristics
How easy to design?	<ul style="list-style-type: none"> <li>• Essay Questions are relatively easy to write</li> </ul>
Level of Thinking Measured?	<ul style="list-style-type: none"> <li>• Restricted Response Essay Questions are appropriate for assessing Comprehension, Application, and Analysis.</li> <li>• Extended Response Essay Questions are appropriate for Evaluation and Creation.</li> </ul>
How Efficient to Cover Content?	<ul style="list-style-type: none"> <li>• Because Essay Questions take a long time to answer, the amount of content that they can be used to cover is highly limited.</li> </ul>
Impact of Guessing?	<ul style="list-style-type: none"> <li>• Guessing is not an issue in Essay Question Responses. However, Essay Questions are susceptible to the Halo Effect when scoring.</li> </ul>
Dangers of Irrelevant Clues that influence responding?	<ul style="list-style-type: none"> <li>• Essay Questions rarely give irrelevant clues to students to help them to respond.</li> </ul>
Scoring?	<ul style="list-style-type: none"> <li>• Essays are very difficult and time-consuming to score and suffer from very low reliability.</li> </ul>

# HANDOUTS FOR PART 3

## HANDOUT 26: About Item Analysis

### BASIC FACTS

### SHEET 1



#### *About Item Analysis*

- Definition of Item Analysis
- Purpose of Item Analysis
- Advantages of Using Item Analysis

## Definition, Purpose, & Advantages

### **Definition of Item**

**Analysis:** *The set of procedures by which we demonstrate how effectively a given test question functions*

### **Purpose of Item Anal-**

**ysis:** *By enabling teachers to identify the good questions and the bad questions on a test, item analysis can be an important too in developing valid tests in any educational setting.*

### **Advantages of Item**

- Analysis:**
1. *Allows teachers to construct more valid tests by keeping good questions and discarding or revising bad questions*
  2. *By facilitating the creation of a Question Bank, Item Analysis saves time in making tests*
  3. *Helps teachers to become better question writers*



### *What is a Good Question?*

- Very difficult?
- Very easy?
- Moderate difficulty?
- Does it have discriminating power?

## Information Provided by Item Analysis

### Suppose . . . .

- *What if 100% of the students answering a question got it wrong. Is this a good question?*
- *What if 100% of the students answering a question got it right? Is this a good question?*

**Answer:** Usually when one gets these extremes in responding patterns, the question needs revision.

### How difficult should a question be?

- *For most tests (e.g., summative tests), questions should be of moderate difficulty*
- *If 26% to 75% of the students answer a question correctly, we say it is of moderate difficulty*
- *Anything more or less than that is considered very easy or very difficult.*

### Suppose . . . . .

- *Suppose that there were 5 students in a class who got nearly all the questions on a test right. They got the highest marks in the class*
- *But on Question X, all of these students got the wrong answer.*
- *Was Question X a good question?*
- **Answer:** *No, Question X is not a good question because it did not help us to discriminate the strong students from the weak ones*



### *Steps in Conducting Item Analysis*

- Ordering test papers
- Forming upper and lower groups
- Tabulation

## Steps in Conducting Item Analysis

**Step 1:** Take all the test papers and rank them in order from the highest to the lowest scoring students.

**Step 2:** Form an Upper Group and a Lower Group by multiplying the total number of students by 27.5%. Then count this number of test papers from the top of the pile and the bottom of the pile. This is your Upper and Lower Group

**Step 3:** Tabulate the scoring patterns for each individual question for the Upper Group and again for the Lower Group. Use the **Item Data Tabulation Form** attached

**Step 4:** Transfer the data from the tabulation sheets onto a summary form for each question. Use the form provided entitled, **Item Analysis Score Card**

**Step 5:** After compiling the data as described above, you are now ready to calculate the following key statistics:

- Index of Difficulty
- Index of Discrimination



## Item Data Tabulation Form

Question No: \_\_\_\_\_  
Date: \_\_\_\_\_

Group: \_\_\_\_\_  
Size: \_\_\_\_\_

Question Type: \_\_\_\_\_  
Grade Level: \_\_\_\_\_

No	a	b	c	d	e
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					

No	a	b	c	d	e
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					

No	a	b	c	d	e
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					

<i>Tabulation Summary Area</i>	
Response	Number
a	
b	
c	
d	
e	
Omit	

### Item Analysis Score Card

Section:		Difficulty Index:		
Question No:		Discrimination Index:		
Item Stem:				
<i>Distrac-tors</i>	<i>Upper Group</i>	<i>Lower Group</i>	<i>Total</i>	
A				
B				
C				
D				
E				
Context (Picture/Diagram)				

Section:		Difficulty Index:		
Question No:		Discrimination Index:		
Item Stem:				
<i>Distrac-tors</i>	<i>Upper Group</i>	<i>Lower Group</i>	<i>Total</i>	
A				
B				
C				
D				
E				
Context (Picture/Diagram)				



### Analyzing Item Analysis Data

- Index of Difficulty

## Calculating & Interpreting Item Difficulty

**In order to Calculate the Difficulty Index, use the following formula:**

$$\frac{N_u + N_l}{N_{u+l}}$$

**Where:**

$N_u$  = Upper group students responding correctly

$N_l$  = Lower group students responding correctly

$N_{u+l}$  = Total students in both groups

### **EXAMPLE**

*5 out of 8 students in an upper group got a question correct*

*2 out of 8 students in the lower group got the same question right*

$$\begin{aligned} & \frac{5 + 2}{16} \\ & = \frac{7}{16} \end{aligned}$$

$$= 0.44$$

$$= 44\%$$

**Interpreting the Index of Difficulty**

*75% and higher – Easy*

*25% to 74% - Moderate Difficulty*

*24% or Less - High Difficulty*



## Analyzing Item Analysis Data

- Index of Discrimination

## Calculating & Interpreting Item Discrimination

In order to Calculate the Discrimination Index, use the following formula:

$$\frac{N_u - N_l}{N_g}$$

**Where:**

$N_u$  = Upper group students responding correctly

$N_l$  = Lower group students responding correctly

$N_g$  = Total students in each group

### EXAMPLE

5 out of 8 students in an upper group got a question correct

2 out of 8 students in the lower group got the same question right

$$\begin{aligned} & \frac{5 - 2}{8} \\ &= \frac{3}{8} \\ &= 0.38 \end{aligned}$$

### Interpreting the Index of Discrimination

0.40 or greater – High Discrimination

0.20 – 0.39 - Moderate Discrimination

0.19 or Less - Low Discrimination



### *Exercise*

- Checking your knowledge
- Look at the Item Analysis data provided and answer the questions below

## Analyzing Question Data

### **Question 1:**

1. *How many students in the upper group?*
2. *What is the correct answer to this question?*
3. *Was the question easy or difficult?*
4. *Which distractor worked better, A or C?*
5. *How would you describe the discriminating power of this question?*

### **Question 2:**

1. *How many students are there in the lower group?*
2. *Can you calculate the difficulty of this question?*
3. *Can you calculate the discrimination of this question?*
4. *How many students did not answer the question?*
5. *Would you keep this question for your test? Why or why not?*

### **ANSWERS**

### Item Analysis Score Card

Section: 1	Difficulty Index:	0.31
Question No: 1	Discrimination Index:	0.38

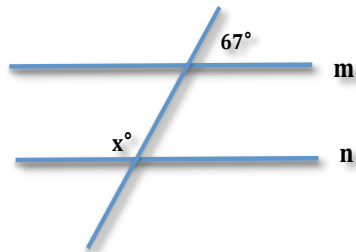
Item Stem: Given that lines  $m$  and  $n$  are parallel in the diagram below, then  $\angle x$  must be equal to:

Distractors	Upper Group	Lower Group	Total
A	9	17	35
B*	24	11	
C	1	1	
D	1	2	
Omit	0	4	

Context (Picture/Diagram)

Given that lines  $m$  and  $n$  are parallel in the diagram below, then  $\angle x$  must be equal to:

- A.  $67^\circ$
- B.  $113^\circ$
- C.  $117^\circ$
- D.  $137^\circ$



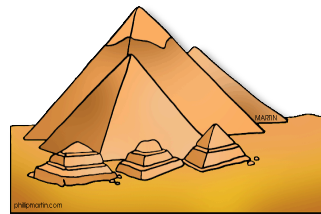
Section: 2	Difficulty Index:	50%
Question No: 2	Discrimination Index:	0.24

Item Stem: Where would you find this monument?

Distractors	Upper Group	Lower Group	Total
A	6	8	49
B*	47	25	
C	0	7	
D	0	7	
Omit	2	2	

Context (Picture/Diagram)

Where would you find this monument?



- A. Greece
- B. Egypt
- C. United States
- D. England

## HANDOUTS FOR PART 4

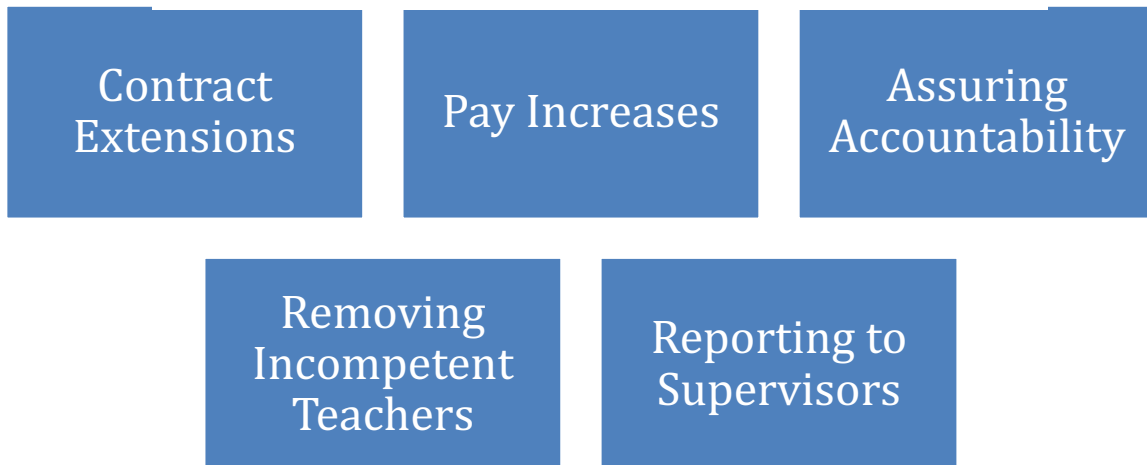
### HANDOUT 27: Knowing Why You Are Evaluating a Teacher

Teacher Evaluation is often undermined by many of the same problems that we see in student evaluation. Perhaps the most important of these problems is that evaluators often do not know clearly 'why' they are doing an evaluation. They may simply be doing the evaluation because it is mandated by the Ministry.

In such cases, evaluators have probably not asked themselves what the **PURPOSE** of the evaluation is and what **DECISIONS** need to be made with the information generated by the evaluation.

The following is a list of traditional purposes of teacher evaluation that you may have cited in your group discussions:

#### Common Traditional Purposes of Teacher Evaluation



**Question:** Do you think that teachers have a positive view of the various Purposes of Evaluation cited above?

**Answer:** The answer to this question is probably '**No.**' When teachers are hostile to evaluation, it is unlikely that it will have much impact on their teaching practice.

**What Is Missing from the Purpose List Above?** What seems to be missing from this list of purposes is the need to:

*'help teachers to improve their teaching practice.'*

**This should be one of the most important purposes of Teacher Evaluation in any school. However, this purpose is often overlooked, or evaluators naively believe that an evaluation to which teachers are hostile will result in improved practice.**

### *Opinion about Teacher Evaluation in Cambodia*

Most of the teacher evaluation that occurs in Cambodia seems to focus on 'traditional' evaluation purposes such as quality assurance rather than helping teachers to improve their practice.

#### **Exercise:**

### **What Do YOU Think about Teacher Evaluation in Cambodia?**

**Directions:** In your group, discuss the opinion stated above and indicate whether you agree or disagree. Be sure to provide justifications for whatever answer you arrive at in your group. Write down your thoughts on a sheet of poster paper and be prepared to discuss them with large group.

We agree/disagree with the above opinion because:

- \_\_\_\_\_  
\_\_\_\_\_
- \_\_\_\_\_  
\_\_\_\_\_
- \_\_\_\_\_  
\_\_\_\_\_
- \_\_\_\_\_  
\_\_\_\_\_



## HANDOUT 28: Different Kinds of Teacher Evaluation

The differences between the kinds of evaluation that were discussed earlier in PART 1 of this document also apply to Teacher Evaluation. Thus, there is **Summative Teacher Evaluation** and **Formative Teacher Evaluation**.

Questions for Small Group Discussion	
What kind of evaluation do you think 'helping teachers to improve their practice' is?	What kind of evaluation do you think the evaluation purposes listed in the boxes in Handout 27 refer to?
If you said, 'Formative Evaluation,' you are correct.	If you said, 'Summative Evaluation,' you are correct.

### Did everyone in your group agree with this classification?

Discuss and resolve any differences in classification that occurred using the definitions and guidelines provided below.

Kinds of Teacher Evaluation (Characteristics)	
Formative	Summative
<b>Purpose:</b> Teacher Growth	<b>Purpose:</b> Accountability; Quality Assurance
<b>Operational Focus:</b> Processes (e.g., What is being done, how can it be improved?)	<b>Operational Focus:</b> Products (e.g., What was done, how well was it done)
<b>Timing:</b> Beginning and Middle of the School Year	<b>Timing:</b> End of the School Year.
<b>Common Instruments:</b> Narrative descriptions of observed behaviors, Open-ended questions as part of a reflection process, etc.	<b>Common Instruments:</b> Checklists, Scored Criteria
<b>Common Uses of Information:</b> Feedback to teachers, Information for reflection, Teacher professional planning.	<b>Common Uses of Information:</b> Contract extension, Pay raises, Inspector reports to Ministry, etc.

## Opinions about Mixing Formative & Summative Evaluation

*Mixing formative and summative evaluation in the evaluation of teachers is a problem. It is quite common to find teacher evaluation systems with elements of both Formative and Summative Evaluation mixed together. For example, ‘using classroom observation data for reporting both to the teacher and to supervisors at district and provincial level’ is an example of mixed evaluation. Far from leading to growth and improvement, dual evaluation systems often demotivate teachers and create resistance and resentment of the assessment process.*

### Exercise:

#### Does Cambodia Have a Dual Teacher Evaluation System?

**Directions:** In your group, discuss the problems raised in the Opinion given above and form your own opinion about whether Dual Evaluation Systems can work well. Then, discuss whether Cambodia uses a Dual Teacher Evaluation System or keeps the different kinds of teacher evaluation separate. Do you think Cambodia’s Teacher Evaluation System works well and, in particular, does it promote Teacher Growth? Give examples to justify your conclusions and your own assessments of Cambodia’s Teacher Evaluation System. Formulate your answers on a piece of poster paper provided by the Facilitator.

<p><b>Question:</b> Do you agree that Mixed Teacher Evaluation Systems are a problem? Explain why or why not.</p>	<hr/> <hr/> <hr/> <hr/> <hr/>
<p><b>Question:</b> Does Cambodia have a Mixed Evaluation System? If you said ‘yes,’ why do you think so?</p>	<hr/> <hr/> <hr/> <hr/> <hr/>
<p><b>Question:</b> Does Cambodia’s Teacher Evaluation System promote Teacher Growth? Justify your conclusion.</p>	<hr/> <hr/> <hr/> <hr/> <hr/>

## HANDOUT 29a: Know What You Want to Evaluate When You Assess a Teacher

Teacher Evaluators need to know what **skill areas** they should be assessing. Many institutions have made comprehensive lists of **teacher competencies** and your school or institution may wish to make its own list. Creating a **Teacher Skill Framework** will help you to identify appropriate teacher skills that are the target of your evaluation so that everyone (both teachers & evaluators) is clear about what is being assessed.

The framework provided below is an example of a Teacher Skill Framework. You can use this framework as it is or adapt it to your own purposes. It is up to you.



Having a framework like this will help you to increase the validity of your evaluation and ensure that your assessment tools are relevant.

### Sample Teacher Skill Framework<sup>4</sup>

No.	Skill Areas
<b>A.</b>	<b>INSTRUCTIONAL PLANNING</b>
	1. Plans lessons and units with objectives.
	2. Plans instruction to achieve objectives.
<b>B.</b>	<b>INSTRUCTIONAL DELIVERY</b>
	3. Carries out instructional plans.
	4. Uses an effective instructional process.
	5. Uses teaching methods that respond to learners' needs.
<b>C.</b>	<b>CLASSROOM MANAGEMENT</b>
	6. Organizes instructional learning time.
	7. Uses space, equipment, and materials to support instruction.
	8. Establishes classroom rules and routines that promote instruction.

<sup>4</sup> Adapted from: Manning, Renfro C. (1988) *The Teacher Evaluation Handbook: Step by Step Techniques and Forms for Improving Instruction*, Englewood Cliffs, NJ: Prentice Hall.

No.	Skill Areas
	9. Manages learner behaviors effectively.
<b>D.</b>	<b>TEACHER-LEARNER INTERACTION</b>
	10. Communicates high expectations for learning.
	11. Engages learners in instruction.
	12. Uses of principles of motivation to engage students in learning.
	13. Demonstrates exemplary communication skills.
	14. Uses principles of reinforcement
	15. Uses principles of retention and transfer of understanding to new lessons.
<b>E.</b>	<b>SUBJECT CONTENT</b>
	16. Shows command of subject matter.
	17. Communicates major concepts and principles of subject matter.
<b>F.</b>	<b>EVALUATION PRACTICES</b>
	18. Evaluates learner performance.
	19. Uses evaluation results to improve learning.
<b>G.</b>	<b>PROFESSIONAL RESPONSIBILITIES</b>
	20. Engages in professional growth activities that relate to classroom performance.
	21. Demonstrates dependability in professional duties.
	22. Works cooperatively in bringing about the success of the school program.
<b>H.</b>	<b>PROFESSOINAL RELATIONSHIPS</b>
	23. Maintains an effective working relationship with staff at all levels.
	24. Maintains a relationship with students that is conducive to learning.
	25. Maintains a relationship with parents that promotes effective communication.

**Exercise:**  
**Improving Your Understanding of Teacher Skills**

**Directions:** In your group, choose a teacher skill area from the list provided and try to identify sub-skills that help to better define the skill area. An example is done for you to give you some idea of how to do the exercise.

**Example:**

<b>B.</b>	<b>INSTRUCTIONAL DELIVERY (Skill Area)</b>
	<b>3. Carries out instructional plans. (Teaching Skill)</b>
	<ul style="list-style-type: none"> <li>a. Teaches to established objectives in the lesson plan.</li> <li>b. Follows the planned instructional process and subject content.</li> <li>c. Uses the planned learning activities and materials.</li> <li>d. Adapts educational plans if unexpected situation occurs</li> </ul>
	<b>4. Uses an effective instructional process. (Teaching Skill)</b>
	<ul style="list-style-type: none"> <li>a. Establishes the focus of the lesson.</li> <li>b. Reviews concepts and skills previously learned.</li> <li>c. Presents clearly the new concepts to be learned.</li> <li>d. Assists learners in practicing new concepts and skills.</li> <li>e. Provides learners with an opportunity for independent practice.</li> <li>f. Gives feedback to learners on attainment of concepts/skills.</li> </ul>

Now, pick a teaching skill from the list provided and identify suitable subskills that you think exemplify that skill. You may choose one or two teaching skills to expand on, depending on the time available.

	<b>Skill Area:</b> _____
	<b>Teaching Skill:</b> _____
	<ul style="list-style-type: none"> <li>a. _____</li> <li>b. _____</li> <li>c. _____</li> <li>d. _____</li> </ul>
	<b>Teaching Skill:</b> _____
	<ul style="list-style-type: none"> <li>a. _____</li> <li>b. _____</li> <li>c. _____</li> <li>d. _____</li> </ul>

When you are done, present your analysis to the large group and discuss/justify the subskills that you have identified.

If another group chose the same skill area, discuss the similarities and differences in the analyses of each group. Synthesize the two lists to form a unified list, if time permits.



## HANDOUT 29b: Variations That May Occur in What You Evaluate

### Variations in the Focus of Evaluation according to the Purpose

As noted earlier, Formative Evaluation is used for purposes of promoting Teacher Growth whereas Summative Evaluation focuses on Accountability and Minimum Competencies in teaching.

As a result, it is likely that Growth-focused Evaluation will be much more flexible and based on teacher needs that are determined in a consensual and non-threatening manner through conferencing, personal planning, and private discussions. A school or institution may also set priority areas of educational change to develop the school, such as helping all teachers to use technology in their teaching or adopting a specific methodology such as Problem-based Learning.

On the other hand, Summative Evaluation is likely to be more fixed, non-consensual, and based on pre-determined lists of indicators that are received from a central office.

#### Variations in the Focus of Evaluation Content by Purpose

Evaluation Type/ Purpose	Content Focus	Processes for Setting Content	Flexibility in Setting Focus	Frequency
Formative Evaluation <i>(Growth Purpose)</i>	Any Skill or Professional Area Agreed by the Teacher or the School as a Whole	Conferencing Teacher Planning School Committee Setting Teaching Priorities	High	Throughout the Year
Summative Evaluation <i>(Accountability Purpose)</i>	Minimum Competency Check lists of sub-skills	Set by a Central Office	Low	Once every 3 or 4 Years End of Yr.

## Variations in the Focus of Evaluation by Teacher Group

Different groups of teachers will have different evaluation needs and the focus of your evaluation may, therefore, vary accordingly. For example, **new teachers** who are just starting their careers may have more general and basic needs in terms of improving their teaching practice. In such cases, the Teacher Skill Framework shown in **Handout 29a** is more likely to play a role in setting the content focus of the evaluation. If a new teacher is still on probation, more summative types of evaluation may also be needed to determine whether they have achieved minimum competency in teaching.

On the other hand, older and more **experienced teachers** are more likely to expect greater input into the process of setting the focus of the evaluation. The focus of the evaluation will depend largely on areas where the teacher feels he/she needs the most assistance. In such cases, the focus may be narrower and more pinpointed on just a few subskill areas.

### How to Determine the Focus of Your Evaluation

There are different ways that an evaluator might use to set the focus of an evaluation depending on teachers' status and the purpose of the evaluation. The table below summarizes the process.

Teacher Status	Growth Purpose	Accountability Purpose
Experienced Teacher	<ul style="list-style-type: none"> <li>Review the Teacher's career path plan to determine areas of interest.</li> <li>Organize a conference with the teacher to determine areas of interest for growth.</li> <li>Review a school program that has set priorities for changes that it expects from teachers.</li> </ul>	<ul style="list-style-type: none"> <li>Observe the teacher using minimum competency lists of teacher skills.</li> <li>Review records of attendance, participation in school planning, interactions with parents and students, and other professional behaviors based on standardized sets of criteria.</li> </ul>
New Teacher	<ul style="list-style-type: none"> <li>Observe the teacher to determine areas of need.</li> <li>Conference teachers to determine areas of interest for growth.</li> </ul>	<ul style="list-style-type: none"> <li>Observe the teacher using minimum competency lists of teacher skills.</li> </ul>



## Exercise: Identifying the Challenges in Setting the Focus of a Teacher Evaluation

**Directions:** Based on the information provided earlier, discuss in your group the challenges for setting the content focus of a teacher evaluation for each kind of teacher and for each of the purposes given below. Write these down on a piece of poster paper using the table format provided. An example is provided to get you started. Once you have identified the challenges, try to think about possible solutions. Be prepared to present your thoughts on your analysis to the large group.

Teacher Status	Growth Purpose	Accountability Purpose
Experienced Teacher	<p>What are likely to be the challenges to setting the focus of a teacher evaluation?</p> <ul style="list-style-type: none"> <li>• <u>Teacher is not motivated.</u></li> <li>• _____</li> <li>• _____</li> <li>• _____</li> </ul> <p>What could be the solutions?</p> <ul style="list-style-type: none"> <li>• <u>Peer Pressure.</u></li> <li>• _____</li> <li>• _____</li> <li>• _____</li> </ul>	<p>What are likely to be the challenges to setting the focus of a teacher evaluation?</p> <ul style="list-style-type: none"> <li>• _____</li> <li>• _____</li> <li>• _____</li> <li>• _____</li> </ul> <p>What could be the solutions?</p> <ul style="list-style-type: none"> <li>• _____</li> <li>• _____</li> <li>• _____</li> <li>• _____</li> </ul>
New Teacher	<p>What are likely to be the challenges to setting the focus of a teacher evaluation?</p> <ul style="list-style-type: none"> <li>• _____</li> <li>• _____</li> <li>• _____</li> <li>• _____</li> </ul> <p>What could be the solutions?</p> <ul style="list-style-type: none"> <li>• _____</li> <li>• _____</li> <li>• _____</li> <li>• _____</li> </ul>	<p>What are likely to be the challenges to setting the focus of a teacher evaluation?</p> <ul style="list-style-type: none"> <li>• _____</li> <li>• _____</li> <li>• _____</li> <li>• _____</li> </ul> <p>What could be the solutions?</p> <ul style="list-style-type: none"> <li>• _____</li> <li>• _____</li> <li>• _____</li> <li>• _____</li> </ul>

## HANDOUT 30: Who Should Do Teacher Evaluation?

**Characteristics of a Teacher Evaluator:** Identifying the personnel at a school or institution who are responsible for Teacher Evaluation needs to be done carefully. If not done carefully, the Teacher Evaluation System will not work effectively. Formative Teacher Evaluation requires that the evaluator has the following characteristics:

*Teacher Evaluators who seek to help teachers 'grow' must have strong skills in building relationships and earning teachers' trust.*



**Assigning Personnel in a Dual Evaluation System:** Identifying Teacher Evaluation personnel can be problematic if a school or institution has a Dual Evaluation System, as described earlier. You will remember that when the same person is doing both Formative and Summative Evaluation, there is likely to be a level of distrust among many teachers that undermines the purpose of promoting 'growth' or improvement in practice.

**Key Recommendation:** Administrators should try to separate the functions of Formative and Summative Teacher Evaluation in a school or institution and ensure that different people are executing each function.

Here are some additional suggestions about creating a division of labor at a school to maintain the separation of teacher evaluation functions:

## Suitable Division of Labor for Teacher Evaluation Functions

Evaluation Type	Suitable Personnel
Formative Teacher Evaluation	<ul style="list-style-type: none"> <li>• Mentor (School-based)</li> <li>• Teacher Peers such as Technical Group Leaders</li> </ul> <p><b>Explanation:</b> These individuals do not have a position of power over the teacher and can cultivate a collegial and non-threatening relationship with the teacher based on trust. Communications between Mentors/Teacher Peers remain confidential and are not reported to superiors. Such conditions of interaction help to build good communication between Mentors/Teacher Peers and teachers.</p>
Summative Teacher Evaluation	<ul style="list-style-type: none"> <li>• School Director</li> <li>• School Vice Director</li> <li>• DOE/POE/MoEYS Inspector</li> </ul> <p><b>Explanation:</b> These individuals have the responsibility of ensuring teacher accountability or quality assurance (as opposed to growth) in the school or institution. They are the direct supervisors of teachers and, therefore, are well-positioned to ensure that teachers are meeting a minimum standard of competence and performance. Although teachers may not look favorably on these evaluations, the goal of the evaluation is not to promote improvement but to certify that a teacher is teaching at a minimum standard or higher.</p>
<p><b>When we consider the very different functions required of personnel engaged in Formative and Summative Teacher Evaluation, it is easier to see why these functions should be kept ‘separate.’</b></p>	

### Exercise:

## The Feasibility of Separating Teacher Evaluation Functions

**Directions:** Discuss in your group the need to have separated teacher evaluation functions as described above. On a piece of poster paper, write out your ideas in response to the following questions:

Questions	Responses
Is it feasible to have separated teacher evaluation functions in a school or institution?	
What do you expect the challenges to be in separating these teacher evaluation functions?	
How can these challenges be effectively addressed?	

## HANDOUT 31: Guidelines for Organizing Teacher Conferences

### What is the Teacher Conference?

One of the primary mechanisms used to facilitate teacher evaluation is the individualized teacher conference. Conferences usually occur one-on-one between the teacher and the evaluator. Teacher conferences are highly structured and occur for clearly defined purposes. There must be a clear agenda for the conference and they are usually organized as pre-conferences (before the evaluation begins) and post-conferences (after data has been collected).

### How Do Conferences Differ between Formative & Summative Settings?

There are important differences in how conferences are organized depending on whether the purpose is formative or summative. The table below summarizes some of these differences:

#### ***Teacher Conference Guidelines***

<b>Evaluation Setting</b>	<b>Pre-Conference</b> <i>(Things to be discussed)</i>	<b>Post-Conference</b> <i>(Things to be discussed)</i>
Formative	<ul style="list-style-type: none"> <li>• Communicate the <b>Purpose</b> of the evaluation (Growth).</li> <li>• Set the <b>goals</b> of evaluation.</li> <li>• Determine areas of interest that the teacher wishes to develop, so that the <b>content</b> of the evaluation (i.e., what is observed) is agreeable to both teacher and evaluator.</li> <li>• Reach a consensus with the teacher on how <b>data collection</b> will occur (e.g., observations, team teaching, etc.).</li> </ul>	<ul style="list-style-type: none"> <li>• The post-conference is for <b>feedback</b> on an observation or general progress towards improvement. Evaluators are urged to avoid identifying 'good' points and 'bad' points as part of the talk. Use a Socratic method where the teacher reflects on his or her own practice and self-criticizes. The evaluator's job is to ask the right questions</li> <li>• Discuss <b>next steps</b> for the change process.</li> </ul>
Summative	<ul style="list-style-type: none"> <li>• Communicate the <b>Purpose</b> of the evaluation (Accountability)</li> <li>• Set the <b>goals</b> of evaluation.</li> <li>• Share with the teacher what</li> </ul>	<ul style="list-style-type: none"> <li>• The post-conference in a summative evaluation is also an opportunity for the evaluator to provide <b>feedback</b> on observed performance.</li> </ul>

Evaluation Setting	Pre-Conference <i>(Things to be discussed)</i>	Post-Conference <i>(Things to be discussed)</i>
	<p><b>performance indicators</b> will be focused on.</p> <ul style="list-style-type: none"> <li>• Inform the teacher how <b>data collection</b> for the evaluation will occur (e.g., observations).</li> </ul>	<p>However, the feedback is usually less nuanced than in a formative evaluation where the evaluator has to consider the effect on the relationship between teacher and evaluator if criticism is too harsh.</p> <ul style="list-style-type: none"> <li>• Inform the teacher of any <b>decisions</b> made regarding contracts, pay increases, etc.</li> </ul>

### Exercise: Role Play *(Optional)*

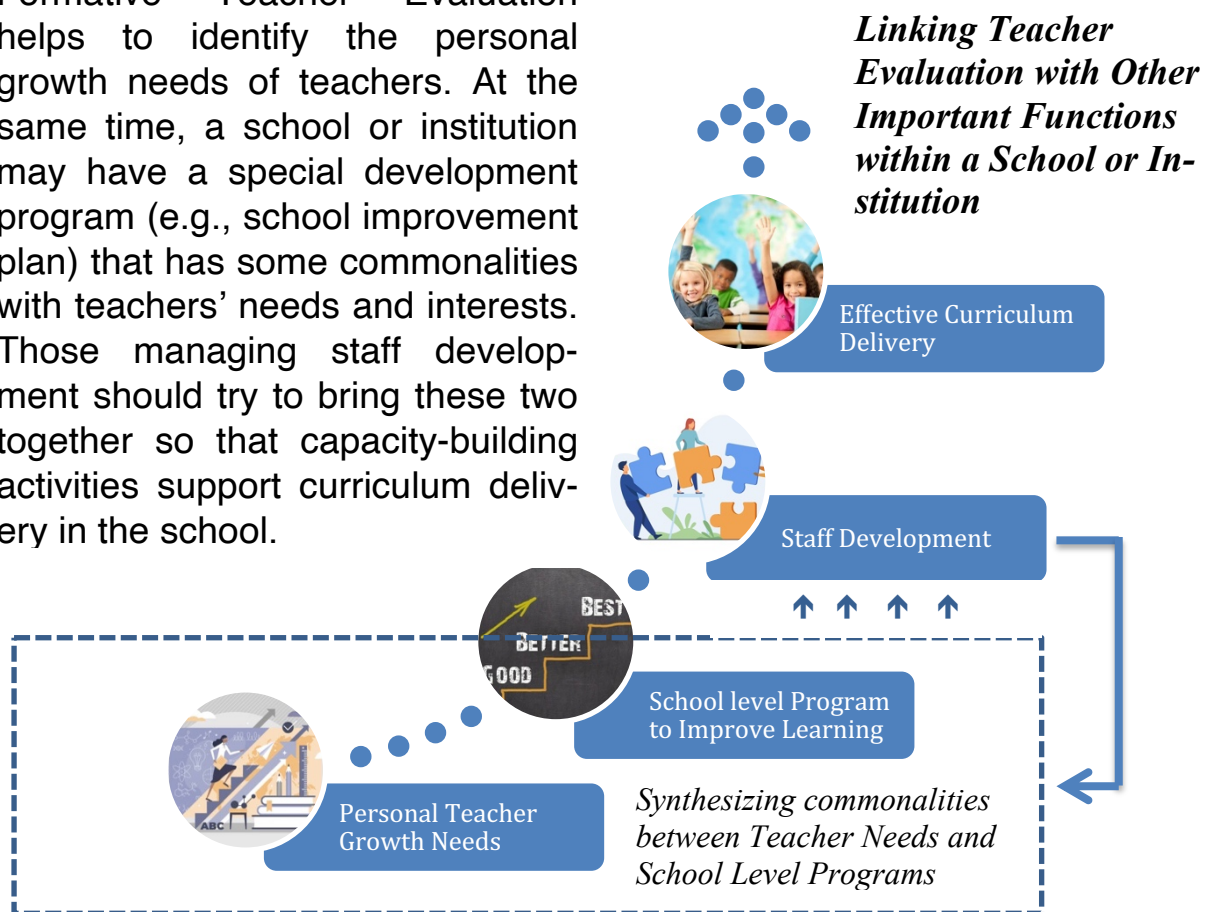
**Directions:** In your group, try to organize a role play in which your group organizes a pre-conference **or** a post-conference for the large group to observe. The role play may use guidelines for either a formative or summative evaluation conference. In your role play, discuss the things that are appropriate for the kind of conference that you are trying to simulate (e.g., goals, feedback, content focus, etc.). Also, be sure that the way you discuss things reflects whether you are a mentor/peer trying to preserve a relationship while at the same time helping someone to grow or an administrator whose main purpose is ensuring accountability for performance. Each role play should be about 10 to 12 minutes. When a role play is completed, the whole group may engage in a short discussion about how well guidelines seem to have been followed.



## HANDOUT 32: Linking Your Teacher Evaluation System to Staff Development

A common problem in many school systems is that staff development systems, teacher evaluation systems, and school improvement programs all operate separately. It would be better if these functions demonstrated interdependence, especially since the information generated by teacher evaluation can help to better inform the content focus of staff development. These recommended interlinkages are illustrated in the diagram below.

Formative Teacher Evaluation helps to identify the personal growth needs of teachers. At the same time, a school or institution may have a special development program (e.g., school improvement plan) that has some commonalities with teachers' needs and interests. Those managing staff development should try to bring these two together so that capacity-building activities support curriculum delivery in the school.



When a school or institution can set priorities for staff development based on a combination of improved understanding of teachers' needs as well as priorities set out in a general school improvement program, the relevance and effectiveness of staff development activities will increase dramatically.

An effective staff development program will in turn create a positive feedback loop with formative efforts to meet teachers' personal needs that are identified through the evaluation process as well as realize the school improvement program.

## Linking Formative Teacher Evaluation and Staff Development: A Concrete Example

(A)	(B)	Explanation
<p style="text-align: center;"><b>Formative Teacher Evaluations identified the following Teacher Needs:</b></p> <ul style="list-style-type: none"> <li>• Additional support in organizing group work.</li> <li>• Strengthening Classroom Management</li> <li>• Extensive support for using ICT in teaching</li> <li>• Using the library more effectively in teaching</li> <li>• Understanding of Problem-based Learning</li> </ul>	+	<p style="text-align: center;"><b>The School Management Committee developed a School Improvement Plan that put a strong focus on the following priorities:</b></p> <ul style="list-style-type: none"> <li>• Helping the school to invest heavily in promoting ICT in Education</li> <li>• Strengthening the use of student-centered learning</li> <li>• Improving physical environments at the school.</li> <li>• Promoting the establishment of student clubs.</li> </ul>
		<p>A number of Technical Grade Leaders who were also peer evaluators summarized the results of formative evaluations and found a number of common themes in teachers' needs (Box A)</p> <p>TGLs reported these needs to the SMC, which had identified 4 priority areas in the School Improvement. Two of these areas overlapped with the formative needs analysis completed by TGLs. These findings then provided guidance to staff development activities.</p>

**Commonalities between the Findings of Formative Teacher Evaluation and the SIP included the following:**

- Improved teaching practices involving student-centered learning, especially Cooperative Learning and PBL
- Using ICT in education

**School Managers and Technical Grade Leaders organized a Staff Development Workplan that included the following activities:**

- A foundational workshop for teachers on PBL and Cooperative Learning (5 days)
- 2 Reflection Workshops (1 day each)
- Opportunities for team teaching with TGLs using new techniques
- One exposure visit to see a New Generation School where such methods are more commonly used.
- Training every Saturday for teachers on basic ICT techniques.
- Providing 50% matching funds from school funds to help teachers buy a used laptop.

**General Discussion Questions**

Read the example provided above and try to answer the following questions.

1. Who in this school assessed the teachers?
2. How many need categories did assessors identify?
3. Through what process did the SMC identify general priorities to improve the school?
4. What were the commonalities between the SIP and teacher assessment?
5. What staff development activities were identified to link the SIP and Teacher Evaluation?



## HANDOUT 33: Bringing the Elements of Effective Teacher Evaluation Together

**Formative Teacher Evaluation Systems:** When a school or institution sets up a Formative Teacher Evaluation system, there are some key considerations that it should consider in organizing the system to make sure that it functions effectively. We have already discussed several of these elements in previous sessions. Review the elements provided in the diagram below to promote the effective organization of a Formative Teacher Evaluation System.

### Key Elements for an Effective Formative Teacher Evaluation System

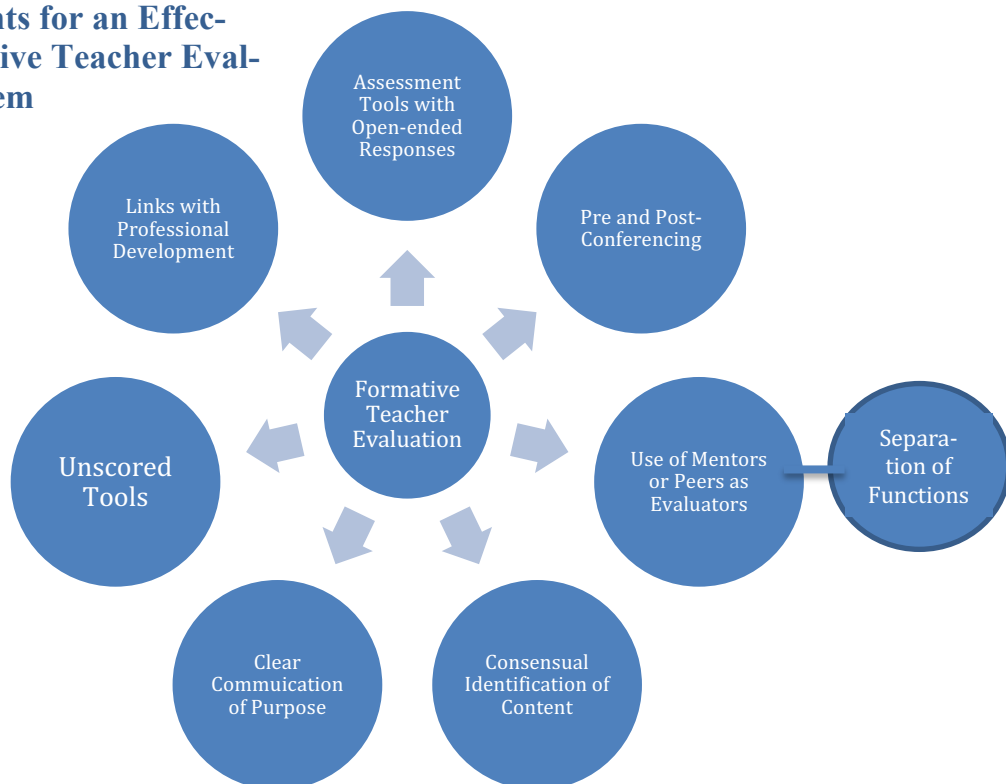


Chart Explanation	
<b>Assessment Tools with Open-ended Responses:</b>	Make sure that the assessment tools used by evaluators provide maximum flexibility to capture insights about what has been observed from a teacher's practice. This suggests the use of narrative responses. Avoid the use of checklists.
<b>Unscored Tools:</b>	Assessment tools should avoid the use of scored responses, which gives the impression that the results will be used for a summative purpose. Since assessments will take a narrative form,

	scoring is not appropriate.
<b><i>Clear Communication of Purpose</i></b>	The purpose of the evaluation should be clear to teachers from the outset so that they do not feel threatened by the process, which will affect motivation and receptiveness to change. The purpose in this case is to facilitate teacher 'growth.'
<b><i>Consensual Identification of Content</i></b>	The identification of the skills or behaviors to be evaluated should be agreed with the teacher beforehand and these should be perceived by the teacher as important to promote their growth. The skill areas identified should also be consistent with any priorities in school-level programs designed to improve education at the school (e.g., using ICT in education).
<b><i>Pre- and Post-Conferencing</i></b>	Organizing individualized pre- and post-conferences with the teacher are essential elements of the evaluation for several reasons. Pre-conferences give an opportunity to communicate the 'purpose' of the evaluation to the teacher and ensure that the focus of skill areas to be evaluated are agreed in a consensual manner, as described above. The post-conference provides the opportunity for the evaluator to provide constructive feedback to the teacher and plan for the next steps in the change process.
<b><i>Use of Mentors or Peers as Evaluators</i></b>	The power relations of those conducting formative teacher evaluation should not be such that they make the teacher feel threatened. This suggests the use of peers (e.g., Technical Grade Leaders) or school-based mentors (if available) as the evaluators. It is often difficult for administrators to play the role of a formative evaluator effectively. Make sure that those doing summative teacher evaluation are not also doing formative teacher evaluation (i.e., separation of functions).
<b><i>Links with Continuous Professional Development</i></b>	The Formative Evaluation System should be closely linked with CPD Systems in a school or institution. The identification of teacher interests to promote their growth can also help guide the formulation of CPD activities such as INSET Workshop Content, Exposure Visits, etc.

**Summative Teacher Evaluation Systems:** Most schools or institutions may likely already have Summative Teacher Evaluation Systems in place, as these serve more ‘traditional’ and administrative purposes such as contract extensions, pay increases, tenure, etc. Nevertheless, there are also some key elements for schools and institutions to consider that will help make the organization of such systems as effective as possible. We have already discussed several of these elements in previous sessions. Review the elements provided in the diagram below to promote the effective functioning of a Summative Teacher Evaluation System.

**Key Elements for an Effective Summative Teacher Evaluation System**

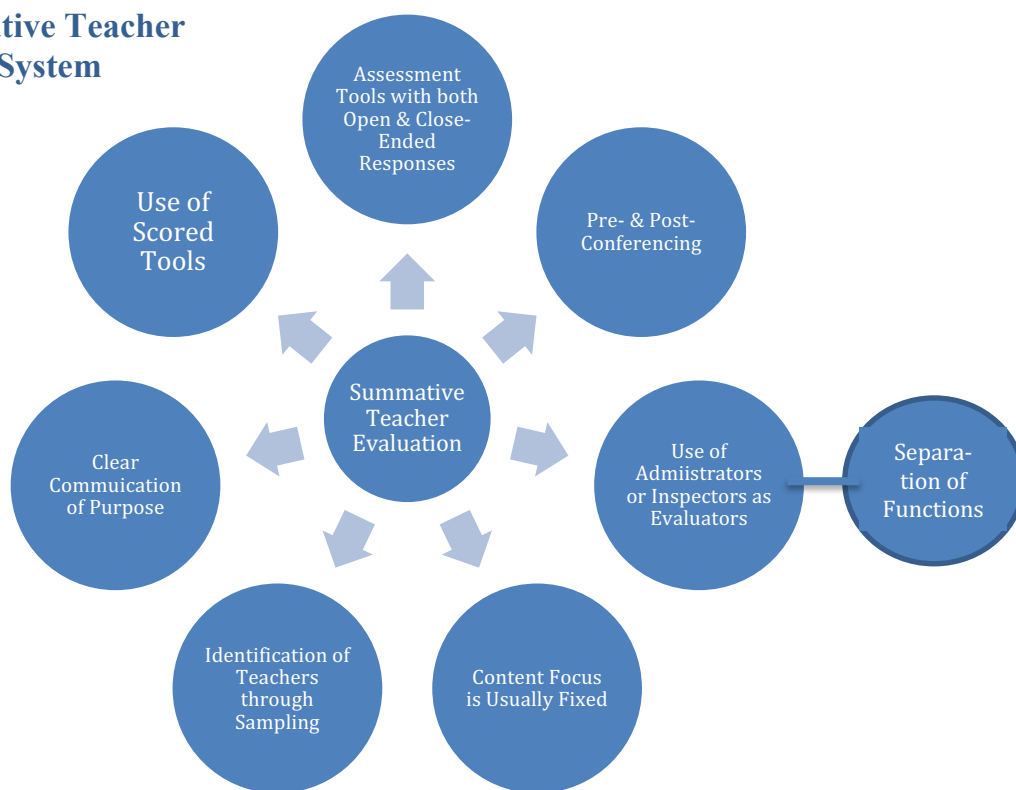


Chart Explanation	
<b>Assessment Tools with Open- and Close-ended Responses:</b>	Assessment tools for Summative Teacher Evaluation may already be established by MoEYS. Usually these tend to use checklists and other close-ended response questions. Open-ended questions that require narrative responses may also be included leading to a combination of both question types on summative assessment tools.
<b>Use of Scored</b>	Summative Assessment tools are also generally

<b>Chart Explanation</b>	
<b><i>Tools:</i></b>	scored to determine cut-off points for minimum competency in teaching.
<b><i>Clear Communication of Purpose</i></b>	The purpose of the evaluation should be clear to teachers from the outset so that they understand the purpose of the evaluation (i.e., accountability) and how the information will be used. Since summative evaluations often make teachers feel threatened, efforts should be made to limit such feelings of discomfort.
<b><i>Identification of Teachers through Sampling</i></b>	Summative Teacher Evaluation is generally applied to teachers whose performance seems borderline to the school director. The director may collect this data from multiple sources such as 'walking around supervision,' short classroom visits, statistical review (e.g., attendance), feedback from parents, etc. If this data indicates low performance, the teacher may be identified for summative evaluation. On the other hand, central office inspectors may have their own system for teacher identification, which may be random or based on fixed time cycles for observation (e.g., once every 5 years).
<b><i>Content Focus is Usually Fixed</i></b>	The identification of the skills or behaviors to be evaluated tend to be pre-determined by central offices. A negotiation process to reach consensual agreement on the content of the evaluation is not generally observed, since the purpose is to determine minimum competency as defined by MoEYS.
<b><i>Pre- and Post-Conferencing</i></b>	Organizing individualized pre- and post-conferences with the teacher are essential elements of the evaluation as they are for Formative Teacher Evaluation. But the focus of the conferences is a bit different. Pre-conferences give an opportunity to communicate the 'purpose' of the evaluation (i.e., accountability) to the teacher and how the data will be used. The manner in which the data is to be collected may also be shared (e.g., classroom observations, profile review, etc.). The post-conference provides the opportunity for the evaluator to provide constructive feedback to

Chart Explanation	
	the teacher and the decisions to be made regarding contract extension, tenure, etc.
<b><i>Use of Administrators or Inspectors as Evaluators</i></b>	Summative Teacher Evaluations whose purpose is 'accountability' are usually carried out by Administrators (e.g., school director or vice director) or Inspectors. Delegating this function to Administrator/Inspectors in this way is most logical because they are the ones who will make the key decisions regarding contract extension, tenure, etc. Once again, evaluation roles should be separate from those doing formative evaluation.



**Any Questions?**

**Exercise:**  
**The Feasibility of Setting Up a Teacher Evaluation System  
 at Your School or Institution**

**Directions:** In your group, review the guidelines for setting up Formative and Summative Teacher Evaluation Guidelines and answer the questions provided below to the best of your ability. Record your responses on a piece of poster paper and be prepared to present your ideas to the large group.

Questions	Responses
In your view, what are the biggest differences in the organization of Formative and Summative Evaluation Systems.	
Are you in agreement with the guidelines provided in this handout? If there are any specific guidelines you do not agree with, what are they and why do you disagree?	
What are likely to be the key challenges in setting up a teacher evaluation system at your school or institution that follows these guidelines? How might you resolve these challenges?	<p><b>Challenges:</b></p> <ul style="list-style-type: none"> <li>•</li> <li>•</li> <li>•</li> <li>•</li> <li>•</li> </ul> <p><b>Solutions:</b></p> <ul style="list-style-type: none"> <li>•</li> <li>•</li> <li>•</li> <li>•</li> </ul>

## 5. ASSESSMENT TERM GLOSSARY

### English-Khmer

English Term	Khmer Translation	Meaning
Analytical Scoring Method		Analytic scoring is a method of scoring used for essay questions in which a fixed number of points is set for particular criteria such as the degree to which an answer addresses the question, writing style, grammar and other criteria deemed important by the teacher. Analytical scoring helps to strengthen the reliability of scoring an essay question.
Assessment		The process of appraising something or someone; the act of assigning a numerical value or rank to the quality, value or importance of something without making a final judgement.
Behavioral Construct		Behavioral constructs in assessment refer to skills that are exemplified by certain behaviors, e.g., the skill of remembering may be exemplified by the behavior of recitation.
Bloom's Taxonomy (Cognitive Domain)		Bloom's Taxonomy is a classification system used to define and distinguish different levels of human cognition—i.e., thinking, learning, and understanding
Classification Question		This is an objective question that requires students to assess an object, statement, or concept using specific statements of principle.
Cognitive Bias		A cognitive bias is a systematic error in thinking that occurs when people are processing and interpreting information in the world around them.
Concurrent Validity		This is the degree to which a test corresponds to an external criterion that is known concurrently (i.e., occurring at the same time). For example, two separate test scores for the same student on the same topic will show concurrent validity (if the scores are similar) or the lack of it (if the scores are very different).
Construct Validity		This is the degree to which a test actually assesses the skills that it was intended to measure (e.g., Remembering, Understanding, etc.).
Content Validity		This is the degree to which a test actually assesses the content areas that it was intended to measure.

<b>English Term</b>	<b>Khmer Translation</b>	<b>Meaning</b>
Criterion-referenced Testing		Criterion-referenced tests compare a person's knowledge or skills against a predetermined standard, learning goal, performance level, or other criterion.
Diagnostic Evaluation		Diagnostic evaluation means to diagnose or discover students' difficulties while learning and assess the problem accurately.
Dichotomous Scoring		A scoring approach in which the response to an item or task is always scored as either correct or incorrect, regardless of the task's inherent complexity or any indication of partial knowledge or understanding in the response.
Difficulty Index		This is a measure of a question's level of difficulty used in Item Analysis.
Discrimination Index		This is a measure of a question's ability to distinguish between high and low achieving students.
Essay Question		This is an examination question that requires an answer in the form of one or more sentences, paragraph, or short composition.
Evaluation		A process that focuses on making a judgment about values, numbers or performance of someone or something.
Extended Response Essay		An extended response essay question is an open-ended question that begins with some type of prompt. These questions allow students to write a response that arrives at a conclusion based on their specific knowledge of the topic.
Formative Evaluation		The goal of formative assessment is to monitor student learning to provide ongoing feedback that can be used by teachers to improve their teaching.
General Objective		A General Objective is an instructional objective that defines performance standards in terms of thinking skills that may not be observable using such words as 'know' and 'understand.'
Global Scoring Method (also known as Holistic Scoring)		Global scoring is a method of scoring essay questions in which the teacher provides an examinee with a single global score regarding the quality of examinee work (i.e., performance) as a whole. This scoring method is in contrast to Analytical Scoring, which breaks a score down into component parts based on discrete performance criteria.



English Term	Khmer Translation	Meaning
Halo Effect		The halo effect is a type of <u>cognitive bias</u> in which our overall impression of a person influences how we feel and think about their character. This bias particularly comes into play when correcting Essay Questions where nice handwriting and good grammar may sway a teacher to give higher marks than a student deserves.
Instructional Objective		An instructional objective is a statement that will describe what the learner will be able to do after completing the instruction.
Inter-rater Reliability		inter-rater reliability is the degree of agreement among independent observers or scorers who rate, code, or assess the same phenomenon or question.
Item Analysis		Item analysis is a process which examines student responses to individual test items (questions) in order to assess the quality of those items and of the test as a whole.
Matching Question		This is a kind of objective question in which students pair items in a prompts column to items in the answers column.
Mean		A mean scale score is the average performance of a group of students on an assessment.
Multiple Choice Question		Multiple choice questions are commonly used objective questions which provide respondents with multiple answer options of which only ONE is correct.
Norm-referenced Testing		Norm-referenced tests are designed to compare and rank test takers in relation to one another. They provide information on whether pupils performed better or worse than a hypothetical average student.
Normal Distribution		A Normal Distribution is a <b>probability distribution of scores that is symmetric about the mean</b> , showing that data near the mean are more frequent in occurrence than data far from the mean.
Objective Questions		Objective test questions are those that require a specific answer. An objective question usually has only one potential correct answer (although there may be some room for answers that are close), and they leave no room for different opinions or disagreements between scorers.

English Term	Khmer Translation	Meaning
Percentile		A percentile is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations fall. ... For example, if a score is at the 86th percentile, where 86 is the percentile rank, it is equal to the value below which 86% of the observations may be found.
Predictive Validity		Predictive validity refers to how likely it is for test scores to predict future job performance or success in life.
Question Bank		A question bank is a collection of analyzed questions that is stored for repeated use.
Question Clue		A word or phrase in a question that unintentionally helps a student to answer the question correctly without really understanding the information tested.
Reliability		Reliability refers to the degree to which scores from a particular test are consistent from one use of the test to the next.
Restricted Response Essay		This is a kind of open-ended question that limits the content and response to be given by a student when answering the question.
Short Answer Question		Short Answer Questions are generally open-ended questions that require students to construct a very specific response where there is only one correct answer.
Specific Objective		A Specific Objective is an instructional objective that defines performance standards in terms of thinking skills that must be 'observable' using such as words as 'state,' 'explain,' 'solve,' and other observable behaviors.
Subjective Question		Subjective questions are open-ended questions that require answers in the form of lengthy written explanations and where the scoring of answers may vary between scorers.
Summative Evaluation		Summative evaluation involves making final judgments about a student's degree of learning at the end of a course or academic program.
Table of Specifications		A table of specifications (TOS) is a chart that teachers and test developers use in item writing. It ensures that the test de-

English Term	Khmer Translation	Meaning
		veloped assesses the content taught and the appropriate learning levels at which the content was taught (e.g., remembering, understanding, etc.).
True False Question		A true or false question is an objective question that consists of a statement that requires students to indicate whether it is true or false.

## Annex 1: Sample Teacher Skills Framework<sup>5</sup>

No.	Skill Areas
<b>A.</b>	<b>INSTRUCTIONAL PLANNING</b>
	1. Plans lessons and units with objectives. <ul style="list-style-type: none"> <li>a. States objectives for learner outcomes.</li> <li>b. Includes objectives that meet the varied needs of learners.</li> </ul>
	2. Plans instruction to achieve objectives. <ul style="list-style-type: none"> <li>a. Matches the instructional process and subject content with stated objectives.</li> <li>b. Identified learner activities and materials for carrying out stated objectives.</li> </ul>
<b>B.</b>	<b>INSTRUCTIONAL DELIVERY</b>
	3. Carries out instructional plans. <ul style="list-style-type: none"> <li>a. Teaches to established objectives in the lesson plan.</li> <li>b. Follows the planned instructional process and subject content.</li> <li>c. Uses the planned learning activities and materials.</li> <li>d. Adapts educational plans if unexpected situation occurs</li> </ul>
	4. Uses an effective instructional process. <ul style="list-style-type: none"> <li>a. Establishes the focus of the lesson.</li> <li>b. Reviews concepts and skills previously learned.</li> <li>c. Presents clearly the new concepts to be learned.</li> <li>d. Assists learners in practicing new concepts and skills.</li> <li>e. Provides learners with an opportunity for independent practice.</li> <li>f. Gives feedback to learners on attainment of concepts/skills.</li> </ul>
	5. Uses teaching methods that respond to learners' needs. <ul style="list-style-type: none"> <li>a. Teaches at varying cognitive levels to meet the varied learning needs of students.</li> <li>b. Accommodates varying learner styles.</li> <li>c. Paces lesson to meet learner needs.</li> <li>d. Uses examples related to learner experiences.</li> <li>e. Varies student groupings to meet learner needs.</li> </ul>
<b>C.</b>	<b>CLASSROOM MANAGEMENT</b>
	6. Organizes instructional learning time. <ul style="list-style-type: none"> <li>a. Systematizes routine procedures and tasks.</li> <li>b. Teaches scheduled class/subjects according to the timetable.</li> <li>c. Maximizes the availability of learner equipment and materials during instruction.</li> </ul>

<sup>5</sup> Adapted from: Manning, Renfro C. (1988) *The Teacher Evaluation Handbook: Step by Step Techniques and Forms for Improving Instruction*, Englewood Cliffs, NJ: Prentice Hall.

No.	Skill Areas
	7. Uses space, equipment, and materials to support instruction. <ol style="list-style-type: none"> <li>a. Arranges the learning area to achieve stated objectives.</li> <li>b. Uses equipment and materials effectively.</li> <li>c. Makes materials and equipment readily accessible to learners.</li> </ol>
	8. Establishes classroom rules and routines that promote instruction. <ol style="list-style-type: none"> <li>a. Uses rules and/or routines at the developmental level of learners.</li> <li>b. Applies rules and/or routines consistently.</li> </ol>
	9. Manages learner behaviors effectively. <ol style="list-style-type: none"> <li>a. Communicates behavioral expectations to learners.</li> <li>b. Maintains a common purpose among learners and the teachers.</li> <li>c. Monitors learners' behaviors.</li> <li>d. Corrects inappropriate classroom behavior.</li> </ol>
<b>D. TEACHER-LEARNER INTERACTION</b>	
	10. Communicates high expectations for learning. <ol style="list-style-type: none"> <li>a. Expects all learners to succeed.</li> <li>b. Provides all learners with an opportunity to learn.</li> <li>c. Holds learners accountable for learning.</li> </ol>
	11. Engages learners in instruction. <ol style="list-style-type: none"> <li>a. Uses questions to elicit learner responses.</li> <li>b. Encourages learners' questions and contributions to lessons.</li> <li>c. Responds constructively to learner questions and contributions.</li> <li>d. Provides activities that promote interaction among learners.</li> </ol>
	12. Uses of principles of motivation to engage students in learning. <ol style="list-style-type: none"> <li>a. Rewards learners' success.</li> <li>b. Creates a pleasant learning climate.</li> <li>c. Varies instructional activities to match learners' interests.</li> <li>d. Creates an appropriate level of concern.</li> <li>e. Makes sure students know the results of their efforts.</li> <li>f. Takes advantage of intrinsic and extrinsic awards.</li> </ol>
	13. Demonstrates exemplary communication skills. <ol style="list-style-type: none"> <li>a. Speaks and writes clearly.</li> <li>b. Uses oral and written language correctly.</li> <li>c. Gives clear, concise directions and explanations.</li> <li>d. Matches communication to meet the purpose of instruction and learners' needs.</li> </ol>
	14. Uses principles of reinforcement. <ol style="list-style-type: none"> <li>a. Uses positive reinforcement to increase desired behavior.</li> <li>b. Uses negative reinforcement (e.g., escape/avoidance) tech-</li> </ol>

No.	Skill Areas
	<p>nique to increase desired behavior.</p> <p>c. Uses punishment when necessary to suppress undesired behaviors.</p> <p>d. Uses extinction to decrease undesirable behavior.</p>
15.	<p>Uses principles of retention and transfer of understanding to new lessons.</p> <p>a. Promotes instruction with meanings that relate to students' lives and experiences.</p> <p>b. Makes use of feeling tones in teaching.</p> <p>c. Teaches in a way so that students can use what they previously learned for improved understanding of new concepts.</p> <p>d. Uses a schedule of teaching practice for new and previously learned content.</p> <p>e. Makes sure that students have learned the most basic information in a lesson well.</p> <p>f. Uses association to make students aware of meanings that apply to new information.</p> <p>g. Uses similarity to teacher students new concepts and skills.</p> <p>h. Uses the identification of critical elements to promote retention and transfer of understanding.</p>
<b>E.</b>	<b>SUBJECT CONTENT</b>
16.	<p>Shows command of subject matter.</p> <p>a. Demonstrates mastery of subject knowledge and skills.</p> <p>b. Presents information that is accurate and up-to-date.</p> <p>c. Communicates information without bias (e.g., cultural, political, etc.).</p> <p>d. Uses a variety of subject matter resources.</p>
17.	<p>Communicates major concepts and principles of subject matter.</p> <p>a. Organizes subject matter into meaningful lesson(s).</p> <p>b. Relates specific lesson topics to major subject matter concepts and generalizations.</p> <p>c. Incorporates recognized (or required) curricular guidelines into lessons.</p>
<b>F.</b>	<b>EVALUATION PRACTICES</b>
18.	<p>Evaluates learner performance.</p> <p>a. Assesses prior learning.</p> <p>b. Monitors on-going performance of learners.</p> <p>c. Encourages learners to evaluate their own performance.</p> <p>d. Evaluates learner's achievement of stated objectives.</p> <p>e. Uses a variety of evaluation techniques.</p>
19.	<p>Uses evaluation results to improve learning.</p> <p>a. Uses evaluation to give learners timely feedback on their per-</p>

No.	Skill Areas
	<p>formance.</p> <p>b. Uses evaluation to diagnose learning difficulties.</p> <p>c. Uses evaluation to plan/adapt instruction according to the needs identified.</p>
<b>G.</b>	<b>PROFESSIONAL RESPONSIBILITIES</b>
	<p>20. Engages in professional growth activities that relate to classroom performance.</p> <p>a. Pursues knowledge about current thinking trends, and practices in education.</p> <p>b. Teacher is flexible their approach to teaching.</p>
	<p>21. Demonstrates dependability in professional duties.</p> <p>a. Fulfills assigned tasks.</p> <p>b. Is prompt and punctual in attendance of meetings, class, and the performance of tasks.</p>
	<p>22. Works cooperatively in bringing about the success of the school program.</p> <p>a. Cooperates with others to bring about the success of the school learning program.</p> <p>b. Exhibits a professional responsibility of the physical and materials resources that the instructional program of the school.</p> <p>c. Helps to communicate the purposes of the school's instructional program to the local community.</p>
<b>H.</b>	<b>PROFESSOINAL RELATIONSHIPS</b>
	<p>23. Maintains an effective working relationship with staff at all levels.</p> <p>a. Respects needs and feelings of his/her colleagues.</p> <p>b. Maintains a positive relationship with all school personnel.</p>
	<p>24. Maintains a relationship with students that is conducive to learning.</p> <p>a. Maintains a supportive and positive relationship with students.</p> <p>b. Respects the confidentiality of student information.</p>
	<p>25. Maintains a relationship with parents that promotes effective communication.</p> <p>a. Maintains a good relationship with parents.</p> <p>b. Is available for conferences with parents.</p> <p>c. Shows diplomacy and respect in communication.</p>