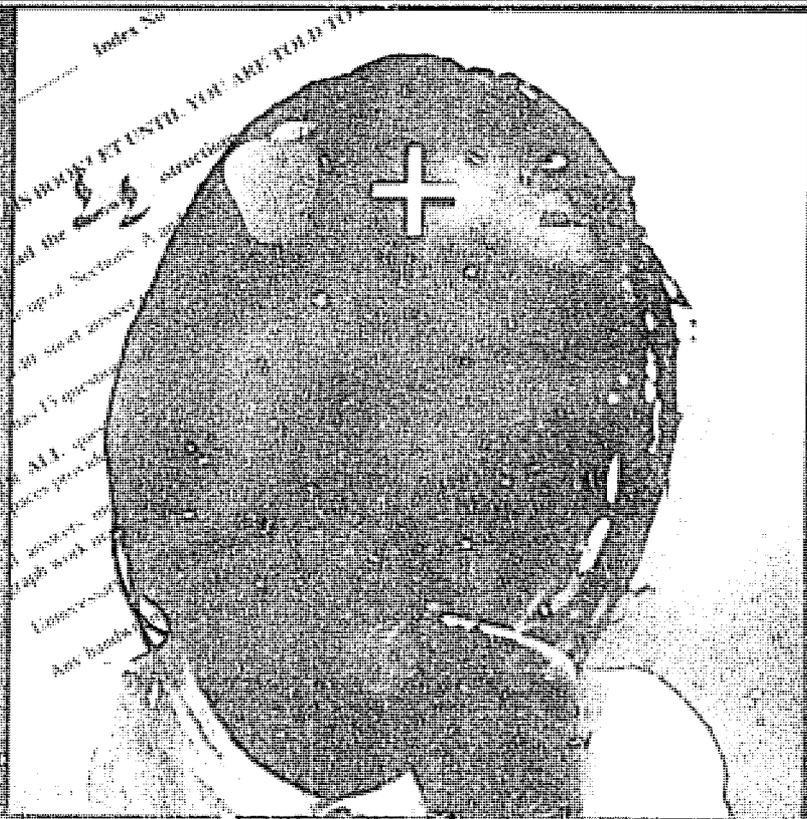
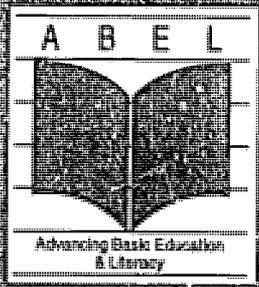


PN-ABS-465 ISBN 90004



Testing to Learn Learning to Test

A Policymaker's Guide to Better Educational Testing



This document was supported by the Agency for International Development (USAID), Office of Research and Development/Education in Washington, D.C. The book and executive summary were prepared by the Academy for Educational Development (AED) as part of its Advancing Basic Education and Literacy (ABEL) Project, Contract No. DPE-5823-7-02-9032-01.

The views expressed in this publication are those of the author and do not necessarily reflect the views of the organizations that supported this effort.

Photo Credits:

Sonjai Reynolds, cover, p. 16

John Allison/Peter Arnold, Inc., ©1992, p. 10

Jean-Luc Ray, Aga Khan Foundation, pp. 11, 20, 24

Joanne Capner, p. 33

Printed on recycled paper 

*Testing to Learn . . .
. . . Learning to Test*

A Policymaker's Guide to
Better Educational Testing

Joanne Capper, Ed.D.

Academy for Educational Development
March 1994

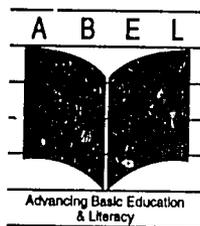


TABLE OF CONTENTS

Using Testing as a Policy Tool to Improve Teaching and Learning	6
Principles of Good Teaching and Learning: Implications for Testing	8
Examinations and National Assessments	12
Designing Tests for Better Teaching and Learning	17
Performance Tasks	17
Open-ended Questions	22
Multiple-choice Questions	23
Increasing the Validity and Reliability of Essay Tests	25
Communicating with Teachers about Testing and Teaching	29
A Call to Action: Fourteen Steps to a Better Testing System	32
Glossary	34
Endnotes	34
Ordering Information	35

ACKNOWLEDGMENTS

Testing to Learn—Learning to Test describes a new view of the role of examinations and national assessments in relation to the central acts of education—teaching and learning. It is based on dramatic changes occurring worldwide in the design and use of testing systems and addresses the technical, social, pedagogical and policy dimensions of the interaction between testing and teaching. This publication represents one facet of the Project ABEL (Advancing Basic Education and Literacy) mandate to describe activities “that work” in the improvement of basic education throughout the world.

This book was begun as an applied research effort under the USAID/Research and Development/Education Office-funded Project ABEL. Under the general project oversight of James Hoxeng, Cognizant Technical Officer, Project ABEL, through the Academy for Educational Development, supported the development of this summary and companion book to highlight concrete experiences and actual practices that could improve testing systems throughout the developing world. A preliminary version of this document was supported by USAID/Pakistan under the Primary Education Development (PED) Program.

A number of people contributed to this effort. The original idea for the book arose from Joanne Capper’s 19 years of experience in testing and evaluation at the U.S., state, national, and international levels. Significant contributions to the conceptual development of the text have been made by a number of U.S. and international testing experts and policymakers, including Ben Makau of Research and Evaluation Associates, Kenya; Peliwe Lolwana of the Independent Examinations Board, South Africa; Anthony Somerset of the United Kingdom; Fuad Abou Hatab of the National Council for Educational Examinations and Evaluation, Egypt; David Ongom and Cyprian Cele of the Uganda National Examinations Board, Uganda; Sar Biland Khan of the Ontario Institute for Studies in Education, Canada; W. James Popham of IOX Assessment Associates, United States; Anthony Nitko of the University of Pittsburgh, United States; David Carroll of the United Kingdom; Karma El-Hassan of the American University of Beirut, Lebanon; Ype H. Poortinga of Tilburg University, The Netherlands; Ash Hartwell of USAID, Africa Bureau, Washington, D.C.; Annie Myeni of the United Nations, New York; Mary Rauner of USAID, Washington, D.C.; and Marilyn Hatch of Virginia. Editorial and technical writing assistance at AED was provided by Sonjai Reynolds, Barbara O’Grady, Franci Hays and John Hatch. Production assistance at AED was provided by Virginia Manta, Leni Nikolov, Tim Lukens and Latina Butler. The executive summary was edited by Lynn Cook and translated into French by Catherine Utz. Proofreading of the French translation was done by Ambassador Alan Lukens. Chroma Design and Communications provided design services. The cover was designed by Judy Zatsick.

We trust that this work will contribute to the dialogue about reforming examinations and designing national assessments in the service of high-quality teaching and learning.

FOREWORD

Testing to Learn—Learning to Test was written in response to a growing recognition, in both developed and developing countries, that what is tested is what is taught in schools. Testing has assumed increasing importance worldwide because education has become the preeminent means for social mobility, an engine for economic growth, and the mechanism for inculcating a sense of national unity. Testing, as the “gateway” mechanism for students, and as a means for monitoring educational quality and efficiency of a nation’s schools, has increased in importance as education competes with other sectors for scarce public resources.

This summary and the companion book describe how examinations and national assessments can be used to encourage more pedagogically sound teaching and learning. Based on recent research and classroom experience, **Testing to Learn—Learning to Test** provides concrete examples of new ways to measure student learning and describes how to develop, score and interpret tests to ensure that they are valid, reliable and fair to all children. It offers techniques and guidelines to increase the involvement of parents, teachers and students in the use of tests to motivate and improve the educational system—not just judge it. These examples and guidelines are offered to educational policymakers, recognizing the many constraints within which educational systems in developing countries must operate.

Although the primary focus of the book is on formal systems of schooling, the discussions and examples, particularly those related to classroom-level testing and teacher training, are equally applicable to non-formal activities.

We hope that this work will reawaken educational policymakers to the importance of testing, and prompt a thoughtful reassessment of examinations and the contribution which national assessment systems can make to promoting educational quality. We hope that from this reassessment will emerge new approaches to the involvement of government, parents, teachers, and students in education and its consequences.

Kurt D. Moses
Vice President and Director
Project ABEL
Academy for Educational Development

USING TESTING AS A POLICY TOOL TO IMPROVE TEACHING AND LEARNING

One education ministry reported that “the system is so exam-ridden that the entire teaching and learning process is geared to passing the exam and getting the good marks needed for entry to higher levels of education.”

Better tests mean better teaching; better teaching means better learning. A well-designed testing system can spearhead educational improvement, but a poorly designed system can sabotage the most dedicated efforts to improve instructional quality. Examinations and national assessments convey powerful messages to teachers, parents, and students about what is important to learn and how it should be taught. This summary provides educational policymakers in developing countries with information on how testing can be used as a powerful and cost-effective tool to improve educational quality.

Tests* can have an enormous impact on what is taught and learned in classrooms, especially when the results of those tests are used to make important decisions such as who gets certified or who gets selected for the next level of schooling. Evidence has shown that teachers and students devote great amounts of time to teaching and learning those topics they expect will be on tests.

The higher the stakes associated with tests,** the more time and attention will be devoted to preparing for them. If the tests do not measure important and meaningful content, skills and knowledge, then precious time and resources are wasted and national development is undermined.

The Way Something Is Tested Is How It Gets Taught and Learned

Educators also have learned that how something is tested influences how it gets taught and learned. If a test focuses on factual knowledge, then teachers will teach factual knowledge; if a test measures reasoning, analysis, and solving real-life problems, then teachers are more likely to teach students to reason, analyze, and solve problems.

Educators and test developers have begun to redesign tests to influence teaching in ways that are consistent with good teaching practice and with new thinking about what students should know to succeed in the modern world.

* The term “test” or “testing” is used to refer to both examinations and assessments. When referring only to examinations, the term “examination” or “exam” will be used; when referring to assessments, the term “assessment” or “national assessment” will be used.

** High-stakes tests have major consequences, such as promotion, selection, or scholarships. Low-stakes tests have minor consequences. National assessments are considered to be low stakes because they generally measure achievement on a sampling basis rather than achievement of individual students.

Because tests are associated with who does and does not get access to opportunities for a better life, they are highly visible and often controversial. Secrecy often surrounds tests, and some people believe, often correctly, that tests are designed to “trick” the test taker. While the changes recommended here will not eliminate the need to allocate opportunities, they will, if implemented properly, ensure that examinations do not distort the process of education. Instead, tests can be designed to enhance teaching and learning.

This Executive Summary and the complete book describe the characteristics of a high-quality testing system and the steps needed to ensure that teachers, students and parents are comfortable with and supportive of change. The role of the educational policymaker is to initiate the process of testing reform and to pave the way for the needed steps to occur in a

thoughtful and thorough manner. These basic steps, which are described in this summary, include:

- ensuring that curricula, textbooks, and tests are consistent with principles of good teaching and learning;
- revising examinations to be consistent with a criterion-referenced approach to testing;
- establishing a national assessment system;
- communicating with teachers and parents about changes in the testing program, and using tests and test results to create a dialogue about teaching and learning;
- setting up mechanisms for ensuring a reliable process for scoring essays, performance tasks, and open-ended questions.

A study of Form 1 science classes found that, “in the observed classrooms, completion of the worksheets with the ‘correct’ answers has become the goal of every science lesson despite a postulated student-centered inquiry approach. The race against time required to complete all the worksheets prior to the government examination has led to considerable cutting of the hands-on activities...”

What Does This Mean for Educational Policymakers?

Policymakers can harness the power that tests have to contribute to better education for all students. Effective use of this power requires that policymakers:

- understand the principles of good teaching, which have become clearer because of recent research on how people think, learn, and solve problems;
- understand some of the policy and technical aspects of testing;
- ensure that certain critical steps are taken to design and use a technically accurate, educationally sound, and intellectually thoughtful testing program.

PRINCIPLES OF GOOD TEACHING AND LEARNING: IMPLICATIONS FOR TESTING

If a test is to promote better teaching and learning, it must be consistent with principles of good teaching and learning. These principles should guide test development as well as the design of curricula, textbooks and teacher training.

It is useful for policymakers to understand these principles so they can ensure that the instructional components of the educational system are designed in ways that promote effective teaching and learning.

Current Practice May Promote Superficial Learning

Numerous studies have shown that much of the learning that occurs in classrooms around the world is superficial. Facts, rules and formulas are memorized, but the bits and pieces of knowledge learned often are not connected into a coherent framework that would allow students to make sense of them and to use them in new situations they may encounter. Remembering, understanding and using knowledge are the essence of education. It is a waste of time and scarce resources to teach students facts or skills which they are unable to access and use when appropriate.

Instruction which relies primarily on teaching children to acquire facts and which neglects to teach them the relationships among and use of those facts fails to impart real understanding. Such learning is viewed by students and teachers as “getting through the book” and “passing the test” rather than

knowing, thinking, understanding, and using. Superficial learning contributes to students’ lack of interest in school and, in turn, to higher rates of repetition and dropping out.

Tests may contribute to superficial learning and also may give inaccurate impressions of how well something has been learned. This happens when a test consists of questions which simply require that students recall facts. For example, students may be able to respond correctly to such questions as, “What was the most important discovery that marked the beginning of the New Stone Age?” or “In which African country was writing on papyrus practiced?” The more important aspects of understanding, integrating and using this knowledge are not addressed in such questions.

Some factual knowledge is important for students to learn. But when a test relies solely, or primarily, on such questions, teachers have little incentive to teach students to understand the significance, meaning or use of such knowledge. Such questions do not require that students apply their knowledge or demonstrate that they see how one piece of knowledge is connected to another; they do not require that students see the patterns and relationships among various pieces of knowledge which combine to form concepts—the powerhouses of knowing, thinking and doing.

Although deep-level understanding and learning to think, learn and solve problems are the ultimate aims of schooling, evidence abounds that we have established procedures and systems that actually promote superficial learning and undermine more meaningful, in-depth learning. We do this by:

- overburdening the curriculum with too many topics;
- writing textbooks that cover topics superficially;
- developing tests that measure and reinforce rote memorization and superficial learning.

Confronted with such constraints, even the most dedicated and capable teacher is forced to shortchange meaningful learning. Under these conditions, a teacher's measure of success is to cover the book and make sure that students pass the test. There is little time to make sure that students really understand and use what they are learning, especially if these things are not measured on the examinations.

Many countries and education agencies are now redesigning examinations to measure deep-level understanding, higher-order thinking and the application of learning to real-life situations. They are doing this in the belief that these new tests will promote similar changes in classrooms—and there is evidence that they are.

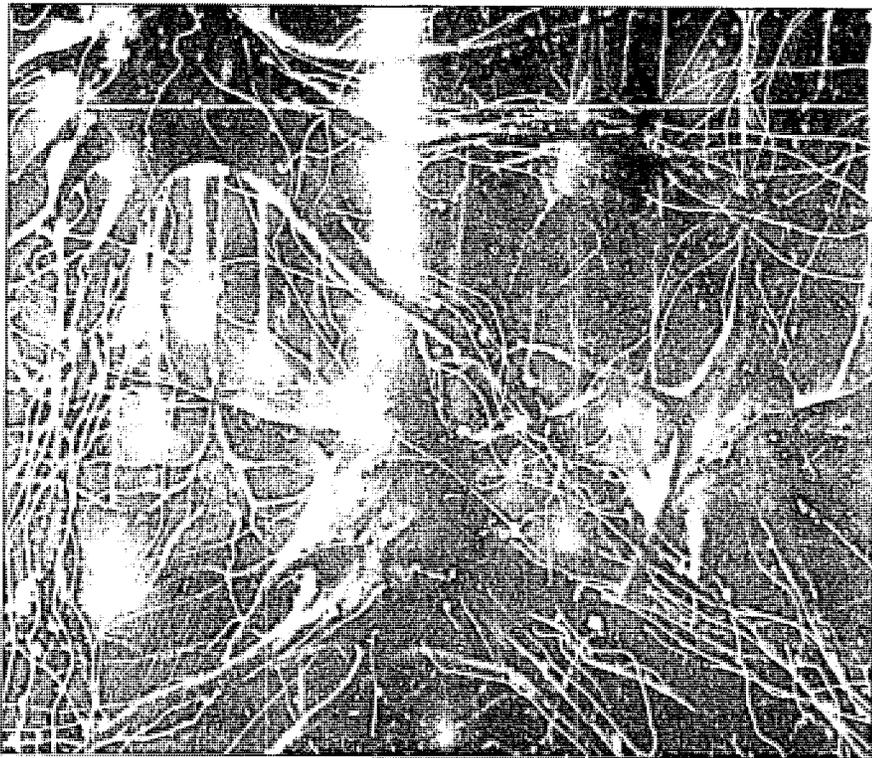
The basic principles of good teaching described below have been found to foster meaningful and useful learning. The description of principles is followed by several examples of clever and innovative test items and performance tasks which illustrate how the principles have been used in tests.

Principle 1: Ensure That Instruction Is Coherent

A critical factor in helping students to understand and make meaning is coherence. Coherence is connectedness. It is the glue that holds the bits and pieces of information or ideas together and prevents them from being fragmented and random. Coherence helps students remember and understand what they learn.

Much of the instruction that students receive is incoherent to them—it doesn't make sense. They may know the bits and pieces of what they learn (e.g., the earth rotates on its axis), but they often are unable to access and use that knowledge when appropriate (e.g., they cannot relate the earth's rotation on its axis to the position of the sun in the sky). If information does not make sense or if it is not coherent, then students cannot remember it or use it.

The notion of coherence centers around the need for key ideas or themes. These themes serve as the anchors upon which students can connect the various bits and pieces of knowledge, skills, concepts and ideas which are required for meaningful learning to take place. Themes allow for a richer set of connections in the student's mind than do less central or less coherent ideas. For example, the theme of "cycles" addresses many important ideas in science, history, geography, and other subjects: the cycle of a day and year, the digestive cycle, the cycle of human development, the water cycle, etc. Each of these cycles involves change and repetition. At various grade levels, students can analyze what changes occur in a cycle, what things repeat in a cycle, what physical or psychological elements contribute to the cycles, etc.



Drawing of the pathways through the brain.

Principle 2: Connect New Knowledge to What Students Already Know

Revolutionary new findings in brain research support what scientists and learning theorists have suspected for some time: the more connections one can make in relation to a topic, the more likely one is to remember and use that knowledge. The drawing above of a section of the brain shows the pathways through the brain which transmit information and thoughts. It also shows that some of the pathways are thick, while others are thin and wispy. Scientists now know that the thick pathways (neurotransmitters) represent a rich and interconnected web of experience and knowledge, and the thin, wispy transmitters represent superficial and unconnected knowledge--bits and pieces.

This means that schooling should be structured to allow for and encourage multiple connections for important ideas. This involves designing curricula, textbooks, instruction and tests in ways that help students see how new information connects to what they already know, to their lives outside of school, and to what they have learned previously in that subject and in other subjects.

Principle 3: Cover Topics at a Deep Level Rather than a Superficial Level

Simply exposing students to information will not cause them to understand or use that information. Instead, students must be provided with experiences which allow them to learn at a much deeper level than is typically provided for in most curricula and textbooks. Learning at a deep level requires that more time be devoted to important topics and that students be given opportunities to learn these topics in a variety of ways. This also means that fewer topics are covered.

Principle 4: Provide Students with Opportunities for Active Learning

To expand the number of connections students make, they should be given a variety of ways to learn (e.g., talking, debating, acting, building models, discussing connections to other topics and subjects, writing stories and reports). The more senses that are activated in learning, the greater the number of connections between existing knowledge and new knowledge.

Principle 5: Use Real-World Tasks

An important characteristic of having an ability is its transferability; that is, students must be able to use their abilities in a variety of appropriate situations. For example, if students can only subtract when they encounter a number problem (e.g., $49 - 36 = 13$) but not when they purchase something at the market, then the skill is of little use. It is important that teachers, texts, and examinations all emphasize applying or transferring newly learned knowledge to a variety of situations.

Principle 6: Make Students Aware

Students are more likely to access and use what they have learned if they are aware of what they know and do not know. Writing or talking about something brings thoughts to a conscious level and allows them to become objects of reflection. Because most misconceptions and oversimplifications are tacit, speaking and writing provide a way to bring them to light and control or correct them.

What Does This Mean for Educational Policymakers?

Policymakers may want to encourage a review of the curriculum, textbooks and tests to ensure that they are consistent with principles of good teaching and learning. Curricula, textbooks and tests should be designed to:

- use central ideas or themes to facilitate coherence in knowledge acquisition and use.
- provide learners with opportunities which help to establish meaningful connections between new and prior knowledge.
- reduce the number of topics and subjects covered and provide more time to study important topics in depth.
- provide more opportunities for students to practice applying what they are learning to real-life types of situations.
- provide students with opportunities to talk and write about content.



When children are allowed to be active learners, they are more likely to remember and use what they learn. Performance tests can help to promote active learning.

EXAMINATIONS AND NATIONAL ASSESSMENTS

Most developing countries have been using examinations to make decisions regarding certification, selection and promotion. However, few countries measure student achievement to assess and improve the effectiveness of their educational systems.

A recent report to the Board on

International Comparative Studies in Education noted:

The statistics that countries regularly gather...fail to cover some of the most fundamental aspects of education. Most resources have been dedicated to...

Figure 1: Educational Assessment

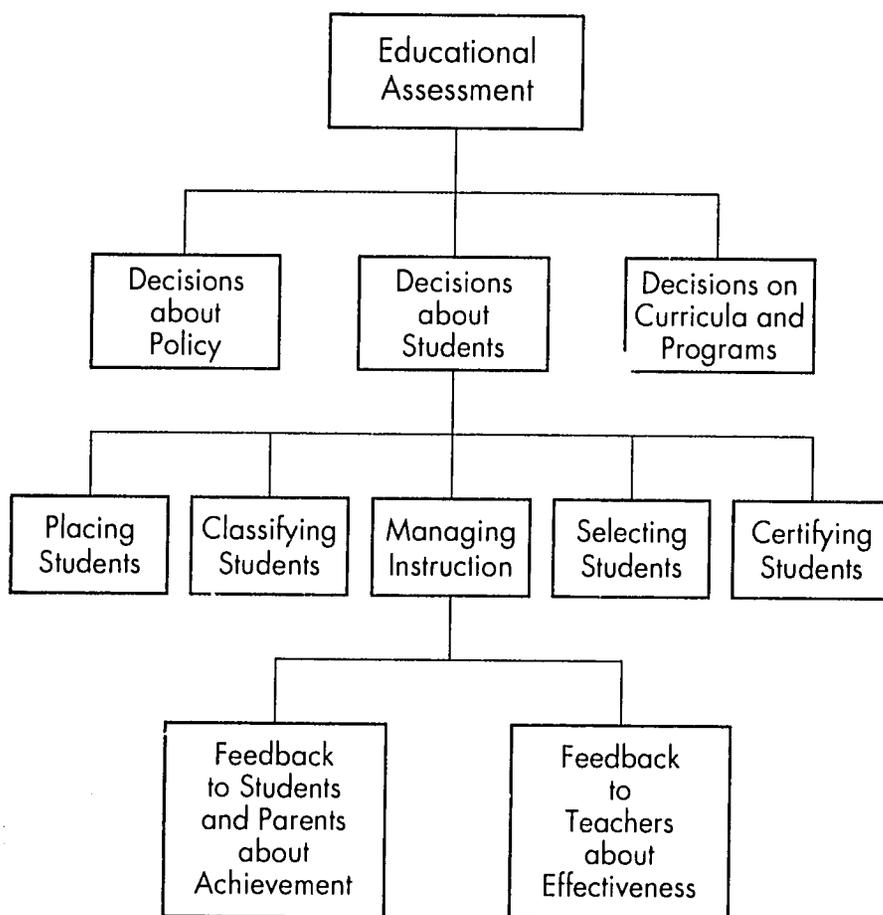


Figure 1: Examples of educational decisions supported by educational assessment. Adapted from A. Nitko (1993)

counting the inputs. Governments have...given virtually no attention to documenting how schools function or what students learn. Therefore much of the data that could help determine where scarce educational resources ought to be allocated, or how effectively they are being used, is simply not on hand.³

Increasingly, developing nations are beginning to recognize the importance of collecting data on student achievement for use in assessing their systems and making decisions about how to improve them (see Figure 1). Donor agencies are encouraging and supporting such efforts.

Although there are a number of differences between examinations used to make selection and certification decisions, and tests used to assess curricula, programs, and educational policy, there are a number of similarities in the ways that each should be designed and used to have a positive influence on teaching and learning. Figure 2 highlights the main differences between examinations used for selection and certification, and tests used for assessing the effectiveness of

the educational system. In addition, a brief description of the differences between norm-referenced and criterion-referenced tests appears below. Both examinations and national assessments should be developed using criterion-referenced testing technology.

Norm-referenced and Criterion-referenced Tests

Assessments and examinations are either norm-referenced or criterion-referenced. With norm-referenced tests, the intent is to compare student performance on the test with that of a norm group of students, rather than to determine how proficient a student is in a particular subject or skill.

The main problem with making normative comparisons is that they do not indicate what students know or do not know. Norm-referenced tests generally only report how well one student, or one group of students, performed in comparison with the norm group of students. A norm-referenced test gives teachers and policymakers very little information that can be useful for improving schools. That is the main reason

Option: Use Examination Results for National Assessment Purposes.

Ideally, a country would have separate tests for public examinations and national assessments. However, most developing countries have limited resources and may not feel they can afford to operate two separate testing systems. Since it is so important to have student-achievement data for use in making decisions about how to improve the educational system, one option is to use examination results for national assessment purposes. Strategies for doing this are described in the full book.

Figure 2: Differences Between Examinations and National Assessments

	Examinations	National Assessments
Administered to	Individual students	A sample of students
Level of report	Individual student	School, district, region, nation
Frequency	Every year	Every 2 or 3 years per subject
Subjects tested	All in curriculum	Math, science, reading, writing
Administered at	End of schooling cycle	Usually one grade per cycle (e.g., 4, 8, and 11)

educators have turned to criterion-referenced tests.

Criterion-referenced tests (CRT's) are used to determine whether an individual has learned specific skills or knowledge (e.g., can the student name foods that would alleviate a particular nutritional deficiency?). CRT's provide detailed descriptions of what the test is measuring, thereby giving clear information about what students do or do not know. The test results are useful for:

- teachers who want to improve their instruction,
- policymakers who want to know how to improve schools,
- headmasters and inspectors who need to give teachers assistance, and
- developers of curricula and textbooks who need to determine whether the curriculum and textbooks are doing the job intended.

The descriptions of what the criterion-referenced test is measuring are called item, or domain, specifications. The specifications describe the skills, concepts, or knowledge to be measured, including allowable content, level of difficulty, and test-item format. Specifications are essential when preparing questions for both examinations and national assessments.

- They help to ensure that a test is measuring what is *intended* to be measured.
- They serve as a guide for item (or question) writers so that all individuals writing items to measure a particular skill or concept will write the items in the same way and at the same level of difficulty.

- Simplified versions of specifications can be used to inform teachers and students about what will be tested and how it will be tested.

- They allow banks (or collections) of test items to be produced and used across situations and time, while ensuring that all the questions that measure a particular ability, skill or knowledge measure it in the same way and at approximately the same level of difficulty each time the test is administered.

A sample set of specifications for measuring the skill of telling time is provided on the next page. These specifications would be given to item writers to ensure that all the items written to measure this skill would be similar. Writing specifications involves making a number of important decisions that can influence both testing and teaching.

SAMPLE SPECIFICATIONS⁴

Primary Mathematics

Grade: Two

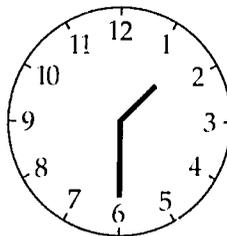
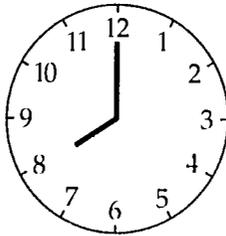
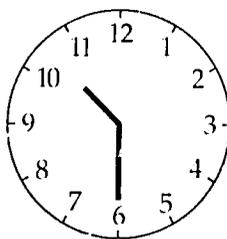
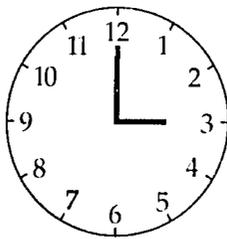
Skill: Tell time in hours and half-hours.

General Description: Students will see drawings of four clock faces with Arabic numerals 1 through 12 with the time indicated at five-minute interval markers. Students will write the correct time in the space provided.

Sample Item:

Directions: *What time does it show on the clocks?*

Write the correct time under the drawings of the clocks in this form: 7:00.



Even for a specification as simple as telling time the decisions are numerous and must be considered in light of what is reasonable for students to know given their age, grade, the curriculum, and the textbooks. For example:

■ *Should children in grade two know how to tell time to the hour, the half-hour, or the minute?*

■ *What size and shape clocks are likely to be used or seen by children of that age?*

■ *Should the test use clocks with digital or Arabic numerals?*

■ *Should there be lines on the clock to indicate one- or five-minute intervals?*

■ *Should the pictures of the clocks used on the test have all 12 numbers or only these at three- or six-hour intervals?*

What Will the Student be Given?

1. Each clock face shall be a 3 cm. diameter circle.
2. Clock faces shall be placed 1.5 cm. apart in two horizontal rows.
3. Arabic numerals 1–12 shall be placed inside the circle next to the correct 5-minute interval marker.
4. One-minute intervals shall not be displayed.
5. Arabic numerals shall be .25 cm. in height.
6. The hour hand shall be .67 cm. long. The minute hand shall be 1 cm. long.
7. No clock face shall show 12 noon or 12 midnight.

How the Student Will Respond

1. The student shall write the correct times and include a colon mark between the hour and minute numerals on the line under the clock face.

What Does This Mean for Educational Policymakers?

The policymaker's role is to ensure that:

- a criterion-referenced approach is used for examinations and national assessments;
- specifications are written for each of the skills and knowledge measured on the tests. This means that testing staff have received training in writing specifications and that they are allocated adequate time to write high-quality specifications. If this step is rushed, then it is likely that the questions which are written will be inadequate;
- curriculum staff and teachers are involved in writing specifications. They are the individuals in the system who have the greatest knowledge of what is important for students to know and how it should be taught and tested;
- specifications and questions written to match the specifications are reviewed carefully and revised based on the review and on field-test results.

Effective learning helps children see the connections between new knowledge and what they already know.



DESIGNING TESTS FOR BETTER TEACHING AND LEARNING

There are many ways to measure students' knowledge and abilities. Each way has advantages and disadvantages. The trade-offs tend to be between the types of questions that encourage better teaching and learning and those that are less costly to administer and score.

Because each type of question format has advantages and disadvantages, most examinations should include some of each format. However, the emphasis here is on how to use tests to influence better teaching and learning, so the formats most likely to do so—performance tasks and open-ended questions—are highlighted here.

Performance Tasks

In performance tasks the student actually performs with knowledge, rather than merely recalling or recognizing other people's knowledge. A performance is a complex and important whole activity for which the student practices (e.g., collecting rocks and classifying them, or interviewing family and community members about a topic and presenting the findings in a report).

Advantages: The primary benefit of performance tasks is that they require students to integrate skills, concepts and knowledge that they are learning and apply these in a real-life situation. Students who engage in performance tasks are more likely to understand what they are learning at a deeper level and to remember and apply what they have learned in new situations.

Performance tasks also are usually

more interesting and motivating for students.

Disadvantages: Performance tasks are more time-consuming, costly and complex to administer and score. Because of these constraints, their use in examinations is likely to be limited. In addition, the use of performance tasks in tests will require that teachers be provided with additional training and assistance in learning how to teach such tasks.

Reading and Writing Performance Tasks

On the next page is an example of an integrated reading and writing performance task designed for primary students. The full task, not shown here, contains four sections:

1. Students read a short story
2. Students individually answer questions about what they have read, including their feelings about the story
3. Students discuss the story in small groups
4. Students reflect on their reading and writing based, in part, on the discussions that occurred in the small groups.

Each of these steps reflects recent research findings about how children learn to read. They illustrate that understanding grows as children interact more with a text. Understanding is not an all-or-nothing experience; instead, it evolves as one devotes more time and thought to the task.

In studying these performance tasks, imagine how classroom teaching might change if teachers were given examples like these in expectation of preparing for an examination.

My thoughts about what I am reading. *(This helps students see that reading is a process of making meaning which emerges as one progresses through the text).*

SAMPLE PERFORMANCE TASK Integrated English-Language Arts⁵

Primary Level

“Stone Fox”

Getting Ready to Read

Here is some information about the story you will read. It will help you understand what is happening.

Little Willy and his dog Searchlight are in a dog sled race. So are Stone Fox and his five beautiful Samoyeds (a strong breed of dog with a thick white coat.) The prize in the race is \$500.

Little Willy hopes to win the money to save his grandfather’s farm. His grandfather is very ill. Stone Fox hopes to win the \$500 to buy back land for his Indian tribe, the Shoshone.

This part of the story begins in the middle of the race. Read to see what happens.

As they approached the farmhouse, Little Willy thought he saw someone in Grandfather’s bedroom window. It was difficult to see with only one good eye. The someone was a man. With a full beard.

It couldn’t be. But it was! It was Grandfather!

Little Willy was so excited he couldn’t think straight. He started to stop the sled, but Grandfather indicated no, waving him on. “Of course,” Little Willy said to himself. “I must finish the race. I haven’t won yet.”

“Go, Searchlight!” Little Willy shrieked.

“Go, girl!”

Grandfather was better. Tears of joy rolled down Little Willy’s smiling face. Everything was going to be all right.

And then Stone Fox made his move.

He went from fifth place to fourth. Then to third. Then to second.

Until only Little Willy remained.

The lead Samoyed passed Little Willy and pulled up even with Searchlight. Then it was a nose ahead. But that was all. Then the Samoyed regained the lead. Then Searchlight poured on the steam.

The crowd cheered madly when they saw Little Willy come into view at the far end of Main Street, and even more madly when they saw Stone Fox was right on his tail.

“Go, Searchlight! Go!” Little Willy cried.

Searchlight gave it everything she had.

She was a hundred feet from the finish line when her heart burst. She died instantly. There was no suffering.

The sled and Little Willy tumbled over her, slid along the snow for a while, then came to a stop about ten feet from the finish line.

The crowd became deathly silent.

Stone Fox brought his sled to a stop alongside Little Willy. He stood tall in the icy wind and looked down at the young challenger, and at the dog that lay limp in his arms.

“Is she dead, Mr. Stone Fox? Is she dead?” Little Willy asked looking up at Stone Fox with his one good eye.

Stone Fox knelt down and put a hand on Searchlight’s chest. He felt no heartbeat. He looked at Little Willy and the boy understood.

Little Willy squeezed Searchlight with all his might. “You did real good, girl. Real good. I am proud of you. You rest now. Just rest.” Little Willy began to brush the snow off Searchlight’s back.

Stone Fox stood up slowly.

No one spoke. No one moved. All eyes were on the Indian, the one called Stone Fox, the one who had never lost a race, and who now had another victory within his grasp. But Stone Fox did nothing. He just stood there. Like a mountain.

After You Have Read the Story:

1. This is only part of the story *Stone Fox*. After reading this part, what questions do you have about the story?
2. What feelings do you have about the story?
3. How do you think the story will end?
4. Choose a line from the story that you especially liked. Write the line here.

Notice how these questions connect the story to the student's existing knowledge, thoughts, and ideas, and how they view students' ideas as credible and important.

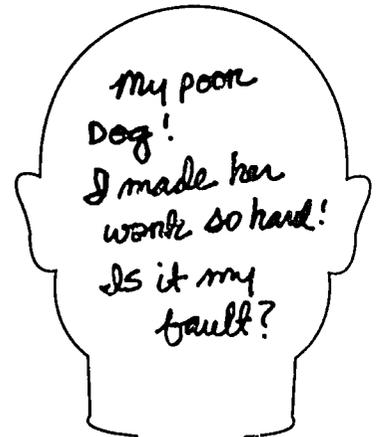
Why did you choose this line?

5. Read these lines from the story again:

Open Mind of Little Willy

Little Willy squeezed Searchlight with all his might "You did real good, girl. Real good. I am proud of you. You rest now. Just rest."

At the right is a drawing of Little Willy's head. If you could look into the mind of Little Willy, what thoughts would you see? Draw or write the thoughts inside his head.



Reflecting on Your Reading and Writing

The teachers who will read your test want to know how you read and write. They are interested in what helps you understand what you read. Anything you want to say about your work on this test will be helpful to them.

Questions such as these encourage students to think about their own reading processes.

1. What helps you understand what you read?
2. How did working in your group help you understand the reading?
3. How did working in your group help you get ready to write?
4. How do you think you did on this test?

Primary-level science lab activities do not require a formal laboratory or expensive equipment. One primary school in Jamaica, located in a very poor neighborhood, used the tree in the school yard as a laboratory. Students brought jars from home to collect specimens from the tree. The specimens included insects, insect eggs, small plants, fungi, etc. The children then observed, classified, and recorded their specimens. This inexpensive and easy assignment engaged students in doing science as scientists would do it. Observing, classifying, and recording are the foundations of the scientific process and important activities for primary school children.

Science Performance Tasks

Many education agencies are beginning to use performance tasks to measure students' understanding of the process of science. Science is about finding out, and students need to know how to do that. Unfortunately, most instruction in science at the primary level tends to focus on having students read from their texts and memorize scientific terminology. To encourage teachers to involve students in hands-on science tasks—that is, actually doing science—a science test should include tasks that allow students to demonstrate their skills in laboratory techniques and in using scientific thinking processes.

For example, students at the primary level can learn to classify and observe—two basic and essential elements of the scientific process. Hands-on science tests at this level

might involve students in putting seeds and beans into categories and explaining why they selected those categories, or measuring and recording various objects, such as their hands, feet, or pulse rate before and after exercise.

Hands-on science assessment is likely to be the most costly and complex form of performance testing, but probably the most essential. Most primary teachers are deficient in their understanding of science and many avoid teaching science at all, let alone hands-on science. Evidence has shown repeatedly that if students do not learn to like science by the end of primary school, they will probably be lost to it forever. If hands-on science is not tested, it will probably not be taught. On the next page is a sample performance task in science for use at the primary level.

Including hands-on science tasks on examinations or national assessments will encourage teachers to involve students in actually *doing* science in the classroom.



SCIENCE PERFORMANCE TASK—GRADE 6

ROCKS

Directions: You are a geology student. Your teacher has asked you to investigate the properties of some rocks you found on a field trip.

1. In front of you are three rocks, "1," "2," and "3." You will need to perform four tests on each rock. Use the materials on the table to answer the following questions.

Record your answers on the chart below.

- Does the rock have holes?
- Does the rock fizz when you gently place it in vinegar? (Leave each rock in vinegar for at least 30 seconds.)
- Does the rock sink or float when you put it in water?
- Does the rock have bands or stripes?

	Holes or no Holes	Reaction to Vinegar Fizz or no Fizz	Sink or Float	Stripes or no Stripes
Rock 1				
Rock 2				
Rock 3				

2. You may use the following information to help you name rocks "1," "2," and "3."

Pumice—Has holes, does not fizz, sometimes floats, no stripes

Calcite—No holes, usually fizzes, sinks, no stripes

Gneiss—No holes, does not fizz, sinks, has stripes

- a. Rock "1" is _____
- b. Rock "2" is _____
- c. Rock "3" is _____

3. In "BAG B" you have a mystery rock labeled "4." You will need to perform the same four tests on Rock 4. Take the rock out of its bag. Do the four tests on the rock. Record your answers on the chart below. Put the rock back in its bag. Use the materials on the table to answer the following questions.

- Does the rock have holes?
- Does the rock fizz when you gently place it in vinegar? (Leave each rock in vinegar for at least 30 seconds.)
- Does the rock sink or float when you put it in water?
- Does the rock have bands or stripes?

	Holes or no Holes	Reaction to Vinegar Fizz or no Fizz	Sink or Float	Stripes or no Stripes
Rock 4				

4. Which rock is mystery Rock "4" like? Check one box.

Rock "1"

Rock "2"

Rock "3"

Explain your answer based upon the results of your tests.

Open-ended Questions

Open-ended questions are another way to measure students' thinking and problem-solving abilities. They generally measure more complex, real-life applications of skills and knowledge and often attempt to get at the strategies students use when they solve a problem. In open-ended assessment tasks, students are asked to:

- restructure information rather than simply recall and reproduce it;
- understand and use information in new and unfamiliar contexts;
- explain “why” and “how” rather than just state a result of some arithmetic or algebraic manipulation;

- integrate and connect their conceptual understanding as they observe, reason, experiment, interpret, make decisions and draw conclusions in situations they encounter within and outside of school;

- demonstrate persistence, imagination, and creativity; and

- approach problems in novel ways.

The following open-ended questions measure a number of mathematical, analytical, problem-solving, and writing skills. These skills are all essential to performing well at higher levels of learning and in many professions, and are more interesting tasks than checking a response from four options.

SAMPLE OPEN-ENDED PROBLEMS⁷

Example 1—Train Ride

A friend of yours, who just moved to India, must ride the train to her aunt's house each week. The train ride costs 50 rupees. Your friend must have exact notes and must use only one-, five-, and ten-rupee notes.

Your friend does not yet understand our money, and does not know how to count our money.

Help your friend find the right rupee notes to give to the fare collector. Draw and write something on a whole sheet of paper that can help her. She needs to see which combinations of rupee notes can be used to pay for the 50-rupee train ride.

Be sure to organize your paper so it is clear and helpful for your friend.

Example 2—Survey of Smoking

The social studies class of Mombako Secondary School surveyed 100 of the school's 800 students about their smoking habits. The results of the survey follow:

38	Never smoked
11	Current smoker who has smoked less than one year
24	Current smoker who has smoked more than one year
18	Quit smoking less than one year ago
<u>9</u>	Quit smoking more than one year ago
100	Total

Write a short article that could appear in a school newspaper about the results of the survey. Include five statements that show interpretations or conclusions you derive from the survey data.

In this problem students must:

- identify important elements of the problem, such as various denominations of money;
- explore possible combinations that yield 50 rupees and determine that all possible combinations have been accounted for;
- show combinations by using illustrations, tables, or charts; and
- organize the information so that it is most helpful to the friend.

In this problem the students must:

- analyze and interpret the data;
- draw conclusions from the data; and
- communicate their interpretations and conclusions to their peers in a style consistent with the style of the school newspaper.

Multiple-choice Questions

Multiple-choice questions differ from the other types described above in that the student *selects* the correct answer from among several options—typically options “A” through “D” or “True” and “False.” In the other types of questions, the students *construct* their responses, in writing, speaking, or perhaps by creating a piece of art or music.

Example of a Multiple-choice Item

To fly from Cairo to Jakarta, the most direct route would be:

- A. Due east
- B. Due south
- C. North, then west
- D. East, then north

Multiple-choice question types can have a dramatic effect on what happens in the classroom. If the examination primarily contains items which have students selecting from among options, then students are likely to spend much of their time preparing for the test with worksheets in which they select the correct answer. This type of behavior is less likely to help students make the extensive web of mental connections which help them to understand and use what they learn in school. However, multiple-choice questions are a very efficient way to measure a large body of knowledge and are likely to be an integral part of any test developed and used for large-scale testing.

Advantages

- Because multiple-choice questions take little time to answer, a test can measure a broader range of content than is possible with a test which relies solely on essay items or performance tasks.
- They are less costly to score. If specially prepared forms are used, they can be machine scored, allowing thousands of answer sheets to be scored in a very short period of time.
- They are an efficient way to measure recall of factual knowledge and some skills.

Disadvantages

- It is more difficult to design multiple-choice questions which measure higher levels of thinking and problem solving.
- It is more difficult to design questions which measure more complex, real-life types of skills and thinking.
- They take more time to develop because of the need to construct four or five response choices.
- Multiple-choice tests promote multiple-choice teaching—that is, teaching where students are always looking for the one right answer.
- There is a high chance of being able to get the correct answer by guessing, which is not the case with performance tasks or essays. If a multiple-choice question has four options, the student has a 25 percent chance of guessing the item correctly.

What Does This Mean for Educational Policymakers?

Obviously, performance tasks and open-ended questions will increase the cost of the testing program, and policymakers will be the ones responsible for requesting funds to support the increased costs. However, the costs must be considered in light of the impact that these types of tasks would have on instruction—if educators can produce tests that energize rather than undermine active learning in the classroom, then the overall cost-effectiveness is greater.

A number of education systems involve teachers in developing and scoring performance tasks, and these teachers claim that it is the best staff development they have ever experienced. Thus, some of the funds allocated to staff development could be incorporated into the testing budget.

Performance tasks and open-ended questions can involve students in restructuring information rather than simply recalling it – in explaining “why” and “how” rather than just “what.”



INCREASING THE VALIDITY AND RELIABILITY OF ESSAY TESTS

Another test format commonly used is the essay. Essay tests are superior to multiple-choice tests for measuring a student's ability to synthesize, organize and analyze subject-matter knowledge, and they are the best way to measure writing competence. Unfortunately, many countries have testing practices which make these tests invalid, unreliable and unfair.

Four steps can be taken to reduce the subjectivity and increase the validity, reliability and comparability of essay tests and the decisions which are made based on the results of the tests. These steps are also essential for developing

and scoring performance tasks and open-ended questions.

Step 1: Write Specifications

Specifications describe the details of the writing test, such as the type of writing, the length, audience, and specific skills tested. Specifications should be written by testing specialists with input from subject-matter specialists, curriculum developers, and teachers. The specifications, such as those below, then would be given to item writers who would write a collection of questions, or items, to match the specifications.

EXAMPLE OF AN ESSAY TASK SPECIFICATION*

(The sample prompt on the next page was written to match these specifications.)

Essay Context	Students should be given a historical context for writing their response that specifies time and historical period.
Audience	Students should be told that their audience is a particular person in the same target historical period. The person to whom they should write must be ignorant of the information in the texts for some plausible reason, such as living abroad or returning from a long trip, to heighten the believability of the task.
Intellectual Task	The student needs to prepare an explanation of the dispute or topic in the text selection(s). This explanation requires the students to understand the viewpoints expressed, to compare and contrast perspectives.
Directions	Directions must stress the need to use knowledge the student has acquired about history outside the text as well as in the provided texts.
Administration	Students should have the text selections available to them as they write, and may be asked to complete the task in one class period, or be given a chance to revise their work. In the latter case, students should turn in their work at the close of each period.
Scoring Scheme	Essays should be scored in terms of the five scoring dimensions described in the scoring guidelines. (see page 27.)

Step 2: Write a Clear, Detailed Prompt

The prompt is the information given to students. It tells them what to write about, how long the writing sample should be, who the audience is, and may specify other criteria on which the students will be judged.

Notice how the second prompt below provides a context, specifies the audience to whom the student should write, and clarifies the student's task.

Step 3: Write Guidelines for Scoring Students' Responses

Scoring rubrics are guidelines for scoring students' responses to the

prompt. Without such guidelines, it is likely that if we give a student's examination to one marker, the student might pass, and if we give the examination to another marker, the student might fail. Such a marking process renders what may be an otherwise perfectly good test unreliable—and reliability is an essential element of a good test.

For example, one marker may place great emphasis on neatness in handwriting and proper punctuation; another may be looking for a creative, perhaps somewhat poetic response; and another may be concerned with determining that the student learned the facts or content covered in the course. Scoring guidelines are essential

EXAMPLE OF AN INADEQUATE PROMPT

Write a paragraph on one of the following:

- A. An apple.
- B. A good deed.
- C. The computer.

EXAMPLE OF AN ADEQUATE PROMPT

History Writing Prompt

Remember 12 June 1964, when, at age 12, you and your father arrived early at the Rivonia Trial courtroom to catch a glimpse of seven Black South Africans on trial for treason. You watched as Nelson Mandela denounced state-supported discrimination against Blacks, and vowed to "take up again, the struggle as best I can..." You watched as he is pronounced guilty of treason and given a sentence of life in prison.

Now, 29 years after that historic trial, you have house guests visiting from Australia. The son of your guest asks you to explain this "struggle" and why a man convicted of crimes against the state has been seen on television shaking hands with the President.

Write an essay in which you explain the most important ideas and issues your guest's son should understand. Your essay should be based on two sources: the general concepts and specific facts you know about South African history, especially what you know about apartheid, and what you learned from yesterday's readings. Be sure to show the relationships among your ideas and facts.

(Students were provided with the text of a speech by Nelson Mandela and an article from a South African newspaper which reported on the history of the African National Congress and the Rivonia Trial. These texts are not printed here.)

for ensuring consistency in scoring and for making valid decisions based on students' performance on the test.

The scoring guidelines provided below were prepared for the specifications and history prompt shown earlier and have been abbreviated here. They are provided in detail in the complete book.

As you read them, think about the kind of test preparation that would go on if secondary students and teachers were given these criteria early in the academic year. Even if students learned

nothing else in their writing classes, to perform well on a version of this essay test they would have had to master a number of important writing, analysis and communication skills and gained an understanding of historical content knowledge.

SAMPLE SCORING GUIDELINES^a

1. General Impression—How well does the student know and understand this historical content? A score of 5 is for the highest level of understanding and 0 is for no response.

2. Use of Prior, Relevant Knowledge—How well does the student incorporate relevant concrete information that is not mentioned in the speeches into his or her essay? This may include pieces of legislation, court decisions, names of people, places, events and general information about the period.

3. Number of Principles or Concepts—How many different social studies concepts or principles does the student use with comprehension? A *concept* is an abstract, general notion, such as “inflation.” It does not refer to particular events or objects, but instead represents features common to a category of events or objects. It must be clear that the student is using a term conceptually, not just as a label. It should be evident that the student understands the concept and means to discuss it. The concept should not simply be mentioned within a quotation from the text with no indication that the student grasps the concept.

A *principle* is a rule or belief used to justify an action or judgment, as in the statement “Apartheid is immoral,” where “morality” serves as a justifying principle.

4. Argumentation—How well does the student organize historical knowledge to make a convincing argument? A score of 5 is the highest level of coherent and cohesive argumentation; a score of 0 means no response. Essays scoring at the highest level will provide adequate evidence to support and justify interpretive stances and a chain of logical argumentation or analysis. A paper that lacks coherence or logical flow should not earn the highest score.

5. Misconceptions—This is a measure of the amount of incorrect information, or the number of misconceptions or misinterpretations in the essay. (A high score on Misconceptions indicates few or no misconceptions.)

Possible types of errors include a) factual errors such as incorrect names or dates, b) misconceptions about the historical period, c) misunderstanding or misinterpretation of the text of the debates.

Step 4: Train Markers to Work in Groups

A number of countries develop brief scoring criteria and then give a stack of students' test papers to individuals who do the marking in their offices or at home. The markers never have the opportunity to discuss the criteria with other markers. This approach is likely to result in a high degree of inconsistency in the scores assigned by the various markers and is also likely to contribute to incidences of cheating in situations where that is a problem.

Instead, bringing markers together to score in groups and providing them with training will increase the reliability of the scoring process significantly. Training helps to ensure that all

markers are interpreting and applying the evaluation criteria in a uniform way.

Training for markers has become somewhat standardized and involves introducing markers to the scoring guides, allowing them to practice applying the criteria to a set of papers, and conducting a trainer-led discussion of the features of each paper that result in the paper's classification or grade.

Using teachers as markers has the advantage of engaging teachers in discussions about what good writing is as well as what and how they should be teaching writing. Several evaluations have found that this results in substantial improvement in the teaching of writing.

What Does This Mean for Educational Policymakers?

- Ensure that testing staff receive training in writing specifications and scoring guidelines for essay questions, performance tasks and open-ended questions.
- Provide teachers with information about the criteria which will be used in marking. Teachers can then use these criteria with their students to articulate standards for quality writing and quality learning.
- Allocate funds to involve teachers, headmasters, curriculum developers and others in the marking of essays, performance tasks and open-ended questions.
- Ensure that markers are trained to use scoring guides and to score consistently.
- Make markers available to share with their colleagues what they have learned about students' errors and misconceptions and to provide suggestions for successful instructional strategies they may have learned through the scoring process.

COMMUNICATING WITH TEACHERS ABOUT TESTING AND TEACHING

Testing can be used as an opportunity to carry on a dialogue with teachers, parents, and other educators about teaching and learning. A testing system provides one of the few opportunities policymakers have to focus public attention on what children are learning and what it is that facilitates or hinders their learning. There are two stages in the testing process where this dialogue

should occur. The first is before testing begins, and the second is after the tests have been administered and the results analyzed. The following is a selection from one education agency's attempt to communicate with teachers about what will be tested, why it is important to test, how it will be tested, and how the results will be used.

EXAMPLE OF ONE EDUCATION AGENCY'S EFFORT TO COMMUNICATE WITH TEACHERS ABOUT THE NEW WRITING ASSESSMENT¹⁰ (Only part of this example is provided here)

The Purposes of the New Writing Assessment

The purposes of this assessment are to provide information to teachers about the current status of students' writing and to begin monitoring writing ability over time. This assessment was designed to influence instruction rather than merely to reflect it. Some of the other purposes are to:

- encourage more and different types of writing in classrooms,
- provide information to help teachers strengthen their writing programs,
- stimulate writing throughout the curriculum,
- provide staff development in writing instruction and evaluation,
- monitor progress of writing achievement, and
- encourage schools and districts to develop a systematic program for improving the quality of writing.

Types of Writing Tested Next Year

The assessment administered next year will focus on three types of writing:

1. **Problem Solving**—In this kind of writing, the writer analyzes a specific problem and suggests ways to solve the problem. Suggested strategies for solving the problem should be backed up with arguments.
2. **Speculation about Causes and Effects**—In this kind of writing the writer presents his or her views about the causes or results of an event, trend, or phenomenon.
3. **Imaginative Writing**—In this kind of writing, the writer tries to create an experience for the reader, making the experience as vivid and real as possible through detail and descriptive use of language.

Reporting and Using Test Results

Reporting test results is another opportunity that policymakers have to communicate with teachers and the broader education community about the substance of education. If a testing system is designed and implemented properly, test results can provide policymakers and other educators with solid, reliable and regular evidence of what is and is not working. Test results reported on a regular basis provide a wealth of valuable data which can be used by different groups of educators throughout the system to analyze and

guide efforts to improve teaching and learning.

Reporting Results to Teachers

Kenya's Examinations Council has done an excellent job of reporting test results in a way that helps teachers solve instructional problems. The council does this through an annual newsletter which is highly readable and free of testing or statistical jargon. These newsletters are comparable to a set of teachers' guides, but have the added benefit of being directly tied to evidence of student performance.

What Does This Mean for Educational Policymakers?

The policymaker should ensure that test results get used by all relevant audiences. Here are some steps to ensure that they are reported and used:

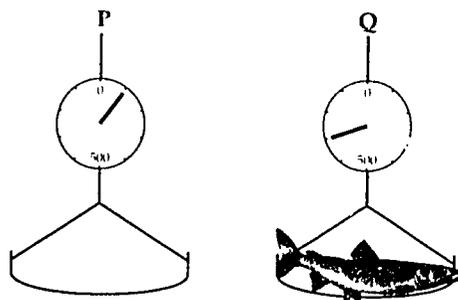
1. Meet with the testing staff to consider the various audiences for test reports and how to address each audience.
2. Encourage the testing staff to meet with the various audiences to identify their information needs.
3. Budget for the preparation, printing and dissemination of reports of test results.
4. Allocate sufficient and appropriate staff to the testing unit to ensure that reports of results are disseminated on a timely basis. Appropriate staff are those who understand testing, teaching and learning.
5. Make available funds to bring various groups of educators together to discuss the results and to prepare plans for improvement.
6. Meet with the press to report the results in a way that focuses on teaching and learning.
7. Use the results to argue for budget allocations for staff development, curricular or textbook revisions, and development of additional instructional and support materials.
8. Meet with teacher educators to urge and discuss improvements in teacher training based on test results.
9. Ensure that test results are reported fairly.

EXCERPT FROM KENYA'S REPORT TO TEACHERS"

One of the questions in the science section of the test deserves special mention because so many students answered it incorrectly:

62. Saifa goes to the market to buy a fish. She notices that the pointer of the balance is set as shown in Diagram P. When the fish is placed in the pan, the pointer is as shown in Diagram Q. Which one of the following is the most likely weight of the fish?

- A. 500 g.
- B. 600 g.
- C. 700 g.
- D. 800 g.



Only 18.5% of rural candidates and 20.1% of Nairobi candidates (excluding those in high-cost and private schools) correctly worked out that the most likely weight of the fish was 600 g. More than 4% of candidates in both samples chose 700 g. This, of course, is the reading shown in the second diagram (Q), with the fish in the pan. These candidates ignored the information given in the first diagram (P), which shows that before the fish was placed in the pan the pointer was set to 100 grams. Thus the fish probably weighs 100 grams less than the reading shown in Q. If Saifa made the same mistake as these candidates, she probably paid too much for this fish!

When a customer buys sugar, meat, or fish from a shop, or when a farmer sells his pyrethrum, coffee, or maize, he needs to be able to check that the weighing is done accurately. During geography and science field trips, pupils should become familiar with as many different types of balances and scales as possible. Pupils can observe weighing being carried out in places such as coffee factories, post offices, and tea centres (places where tea farmers sell their green tea leaves) as well as shops and markets.

A CALL TO ACTION: FOURTEEN STEPS TO A BETTER TESTING SYSTEM

Educational policymakers are the initiators and overseers of testing reform. To do this requires understanding, time and commitment. The following is a list of steps for designing and implementing a high-quality testing system to enhance teaching and learning.

1. Form an oversight, advisory committee of representatives of various stakeholders including teachers, parents, curriculum specialists, teacher educators and policymakers to review and advise on aspects of redesigning and implementing the testing system.
2. Arrange for testing staff and some teachers to receive training in strategies for developing criterion-referenced tests and in using test formats that promote good teaching and learning, such as performance tasks.
3. Review the curriculum and textbooks to make sure they are consistent with the principles of good teaching and learning. Make revisions where needed.
4. Revise examinations and develop tests for use in national assessment.
5. Develop booklets for teachers which describe what will be tested and how it will be tested, and provide suggestions for teaching.
6. Develop booklets for parents which explain proposed changes in the examinations and contain suggestions for how parents can help their children learn what will be measured on the examinations.
7. Develop booklets for students to use in preparing for selection examinations. All students should have equal access to these booklets. If there is a charge, it should be affordable by all parents.
8. Provide teachers with training and assistance in classroom assessment, instruction and use of test results.
9. Report test results on a timely basis and provide teachers with non-technical, easy-to-read reports which analyze students' problems and offer concrete suggestions for how to improve instruction.
10. Report assessment results in a way that acknowledges that results are a function of a variety of factors other than schooling, such as parent income and education, geographic location and opportunity.
11. Make clear how the Ministry of Education is using test results to improve parts of the educational system, including
 - the curriculum,
 - textbooks and teacher guides,
 - teacher training and staff development,
 - technical assistance to schools and teachers, and
 - conducting follow-up research, including using teachers as researchers.
12. Provide advance notice on changes in examinations so that teachers, students and parents will feel they have had sufficient time to prepare for the examinations. Explain changes concretely.

13. Involve teachers in scoring essay tests, performance tasks and open-ended questions.

14. Try out new tests once or twice in small-scale pilot activities and use information gathered in the pilot activities to make improvements in the tests and to give teachers, students and parents feedback on how students are performing. At this stage, feedback should not identify specific schools, teachers or students but should reflect general strengths and weaknesses and provide suggestions for improvement.

Conclusion

Tests are critical to how students view their educational experience; they are critical to schools as guides for improvement; and they are critical to a nation because they measure educational accomplishment. Quality tests can help build a quality nation—and for that reason alone, they are worthy of the most careful attention from concerned educators.

Communicate with parents about proposed changes in the examinations and about how they can help their children learn.



GLOSSARY

Assessment	A process for obtaining information that is used for making decisions about students, curricula and programs, and educational policy. (See Figure 1, page 12, for a classification of some of the types of decisions.)
Assessment Techniques	Include formal and informal observations of students, paper-and-pencil tests, student homework and research papers, project performances, oral questioning, analyses of students' records, lab work, etc.
Criterion-referenced Tests	Tests developed to determine whether an individual has learned specific skills or knowledge.
Evaluation	The process of making a value judgment about the worth of someone or something. May or may not be based on measurements or test results.
Measurement	A procedure for assigning numbers (usually called scores) to a specified attribute or characteristic of persons in such a way that the numbers describe the degree to which the persons possess the attribute.
Norm-referenced Tests	Tests designed to compare student performance with a norm group of students, rather than to determine how proficient a student is in a particular subject, skill, or concept.
Reliability	The extent to which a test yields the same results on repeated administrations, or the extent to which different raters assign the same score to a student's performance on a test or task.
Tests	Instruments or systematic procedures for observing and describing one or more characteristics of a student using either a numerical scale or classification scheme. Also referred to as examinations in some countries. Examinations refer to tests used to 1) certify that a student has learned what was taught during that cycle, 2) determine if the student has learned enough to justify promotion to the next level, and/or 3) select students for a limited number of places at the next level of schooling. Can provide information about an individual or a system. Can be used to determine whether school systems have attained certain goals. When used to describe systems, the score is usually the average score of the pupils who took the test in a school, district, region, or state.
Validity	The extent to which a test measures what it purports to measure. Concerned with the interpretations and decisions made which are based on the test results.

ENDNOTES

¹ Uganda. Ministry of Education (1989) Report of Educational Policy Review Commission, Kampala.

² Rowell, P. and Prophet, R. (1990) Curriculum-in-action: The 'Practical' Dimension in Botswana Classrooms, *International Journal of Educational Development*, Vol. 10, No. 1, p. 20.

³ Puryear, J.M. (1993) Status and Problems of International Education Statistics and Research. Paper presented at a meeting of the Board on International Comparative Studies in Education, National Academy of Sciences, Washington, DC, p. 5.

⁴ Adapted from IOX Assessment Associates, Los Angeles, CA

⁵ Sample test items from California Field Tests, California Assessment Program (1991) California State Department of Education, Sacramento, CA.

⁶ Ibid.

⁷ Adapted from sample test items, California Assessment Program (1989). *A Question of Thinking: A First Look at Students' Performance on Open-ended Questions in Mathematics*. State Department of Education, Sacramento, CA.

⁸ Adapted from Baker, E.L., Aschbacher, P., Niemi, D., and Sato, E. (1992) CRESST Performance Assessment Models: Assessing Content Area Explanations. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, p. 37.

⁹ Ibid., pp. 69-72.

¹⁰ Adapted from *The Writing Assessment Handbook*, The Pennsylvania Department of Education, Division of Evaluation, 1992, p. 3.

¹¹ Adapted from Kenya National Examinations Council (1979) C.I.P.E. Newsletter: 1979, Nairobi, Kenya, p. 193.

ORDERING INFORMATION

Technical Papers

Project ABEL may be preparing additional technical advisory papers on topics related to improving examinations and assessment systems. Below is a partial list of possible topics. Please check those in which you would be most interested and feel free to add others.

- | | |
|---|---|
| 1. <input type="checkbox"/> Testing language-minority students | 9. <input type="checkbox"/> Testing approaches in other countries |
| 2. <input type="checkbox"/> Ensuring that students have equal opportunity to learn what is tested | 10. <input type="checkbox"/> Setting standards |
| 3. <input type="checkbox"/> Bias and fairness in testing | 11. <input type="checkbox"/> Publishing tests results |
| 4. <input type="checkbox"/> Generalizability of performance tasks | 12. <input type="checkbox"/> Matrix sampling |
| 5. <input type="checkbox"/> Setting passing scores | 13. <input type="checkbox"/> Item banks |
| 6. <input type="checkbox"/> How to decide what should be tested | 14. <input type="checkbox"/> School-based assessment |
| 7. <input type="checkbox"/> Teachers as researchers—explaining test results | 15. <input type="checkbox"/> Continuous assessment |
| 8. <input type="checkbox"/> Using examinations or national assessments to evaluate teachers and schools | 16. <input type="checkbox"/> Other topics _____ |
| | _____ |
| | _____ |

Check if you would like to receive samples of the following types of materials.

- | | |
|---|--|
| 1. <input type="checkbox"/> Performance tasks, open-ended questions, portfolio guidelines | 3. <input type="checkbox"/> Materials for training markers |
| 2. <input type="checkbox"/> Scoring guidelines | 4. <input type="checkbox"/> Specifications (specify subject and grade level) |
| | 5. <input type="checkbox"/> Other _____ |

Regional Professional Development Testing Networks

The Academy will support, subject to funding, several regional professional development networks in testing. The networks would be designed to strengthen the technical measurement capacity of network participants and to provide an opportunity to learn from, and build upon, the work of one another.

Each network could consist of 30 to 40 individuals from neighboring countries who would meet four to five times per year, rotating meeting sites among participating countries. Although the focus of the networks would be on examinations and national assessments, participating countries would be encouraged to send staff from other departments related to testing, such as curriculum, textbook, and teacher training. The reason for including members of these other groups is to increase the potential for having a positive impact on teaching and learning. Senior-level policymakers also would be included in selected network meetings, as would teacher representatives.

The aim of the networks would be to upgrade the technical testing expertise of the participants, as well as their understanding of how testing fits into the broader educational policy framework. Another goal would be to provide a mechanism for participants to communicate and share with each other and to open up opportunities for internships and exchanges with U.S.-based testing organizations. The networks would be designed so that participants would determine their own needs and set the ongoing agenda of training and other related activities based on their collective perceived needs and interests.

If you are interested in participating in a network, ordering additional copies of this executive summary, or ordering copies of the complete book, **Testing to Learn—Learning to Test: A Policymaker's Guide to Better Testing**, please complete the order form on the reverse side.

(See order form on next page)

ORDER FORM

Name: _____ Title: _____

Institution: _____

Address: _____

City: _____ State: _____ Country: _____ Zip or Postal Code: _____

Telephone: _____ Fax: _____

Specify Language Preference: _____

Number of Copies

Executive Summary _____

Complete Testing Book _____

Networks

Check if you are interested in finding out more about professional development networks in testing.

Yes ___ No ___

Others who should receive copies of this summary:

Name: _____ Title: _____

Institution: _____

Address: _____

City: _____ State: _____ Country: _____ Zip or Postal Code: _____

Training and Technical Assistance in Testing

Please indicate the appropriate contact person if you are interested in receiving training or technical assistance in testing.

Contact Name: _____ Title: _____

Telephone: _____ Fax: _____

Institution: _____ City/State/Country: _____

Send orders to: Testing Book

Project ABEL

Academy for Educational Development

1875 Connecticut Avenue, N.W.

Washington, D.C. 20009-1202

FAX: (202) 884-8400 Voice: (202) 884-8000

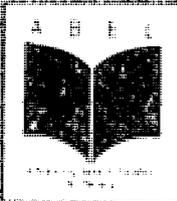
Internet: "ALMAZ+aBASIC+__ED%AEDWASH"@MCIMAIL.COM

2/10

This publication is a summary of a larger, more comprehensive book on testing for policymakers. The table of contents of this more complete book is listed below.

TABLE OF CONTENTS

- Chapter 1 Using Testing as a Policy Tool
The Policymaker's Role in Testing
- Chapter 2 Principles of Good Teaching and Learning: Implications for Testing
Principle 1: Coherence
Principle 2: Cover Topics in Depth
Principle 3: Connect New Knowledge to What Students Already Know
Principle 4: Active Learning
Principle 5: Use Real-world Tasks
Principle 6: Make Students Aware of Their Learning
Principle 7: Organization
Principle 8: Teach Students to Manage Their Own Learning
- Chapter 3 Testing for Better Teaching and Learning
Performance Tasks
Open-ended Questions
Objective-type Questions
Promoting Equity and Addressing Needs of Terminating Students
Sample Performance Tasks, Open-ended Questions, and Enhanced Multiple-choice Questions
- Chapter 4 Examinations, Assessments and Continuous Assessment: Differences and Uses
Examinations
Norm-referenced and Criterion-referenced Tests
National Assessments
How Assessments and Examinations Differ
Continuous Assessment
Integrating Examinations, Assessments, and Continuous Assessment
Can Examinations Be Used for Assessment Purposes?
Validity and Reliability
- Chapter 5 Preparing a Test Plan
- Chapter 6 Item Specifications
Why Are Specifications Important?
What Do Specifications Look Like?
Using Specifications to Communicate with Teachers
Sample Specifications
Examples of Communications with Teachers
- Chapter 7 Essay Tests
Testing to Improve the Teaching of Writing
Increasing the Validity and Reliability of Essay Tests
Specifications for Essay Tests
Writing Essay Prompts
Scoring Guides
Training for Markers and Marking in Groups
- Chapter 8 Reporting and Using Test Results:
A Dialogue with Educators
The Policymaker's Role
Allocating Accountability Fairly
What Goes into a Report for Teachers?
What Goes into a Report for Policymakers?
Using Attitude and Background Data to Help Explain Results
- Chapter 9 Pitfalls to Avoid



This publication is funded by USAID Research and Development/Education. It is published under the auspices of Project ABE, Advancing Basic Education and Literacy. Correspondence should be directed to:

Project ABE
Academy for Educational Development (AED)
1527 Connecticut Avenue, N.W.
Washington, D.C. 20004-2202

TEL: 202/462-6400
TELEX: 202462-6400

Project ABE is a USAID-funded world-wide education project to enhance literacy and provide basic education. It provides direct technical assistance, reports on project progress, research, and materials on program activities and techniques developed in the field. The project is coordinated by the Academy for Educational Development, in close cooperation with Creative Associates International (CAI), the American Council on Educational Development (ACED), and the Research Triangle Institute (RTI).