World Bank Lending for Educational Testing

A General Operational Review

By

Linda P. Larach
(Consultant)
and
Marlaine E. Lockheed

Education and Employment Division
Population and Human Resources Department
The World Bank

August 1992

This publication series serves as an outlet for background products from the ongoing work program of policy research and analysis of the Education and Employment Division in the Population and Human Resources Department of the World Bank. This document has a restricted distribution and may be used by recipients only in the performance of their official duties. Its contents may not otherwise be disclosed without World Bank authorization.
Acknowledgements

The research assistance of Maria Helena Maldonado-Villar and the helpful comments of Linda Dove, Himelda Martinez, Angela Ransom and Adriaan Verspoor are gratefully acknowledged.

The International Bank for Reconstruction and Development/
The World Bank, 1992
Executive Summary

The new emphasis on educational quality in Bank-financed projects and the search for valid and reliable indicators of education inputs are focusing increased attention on educational tests. The present review is designed to describe and analyze the Bank's involvement in supporting the development of educational testing capacity in member borrower nations. The major findings show that:

- Over the past 17 years, the World Bank has supported educational testing in 90 projects in 55 countries.
- World Bank lending for educational testing has risen since FY90.
- There is no discernable trend in the level of financing of testing subcomponents because specific budget figures for testing subcomponents are found in only 23 of the 90 projects.
- A higher proportion of education projects in LAC include testing subcomponents than do any other region, although the total number of education projects with testing subcomponents is higher in Africa and Asia.
- More projects support tests for selection and/or certification than for any other purpose; yet support for tests to monitor achievement trends is increasing, while support for tests for selection and/or certification is declining.
- Support for subcomponents related to changing an existing testing system is more frequent than support for creating a new testing system, and both types of subcomponents are more frequently found in projects than subcomponents designed to prepare for the establishment of a new testing system.
- The quality of testing subcomponents is improving but remains low; and technical quality is improving more rapidly than institutional or dissemination quality.

The recency of testing subcomponents in Bank lending operations makes drawing conclusions about the conditions under which these subcomponents succeed rather difficult. Nevertheless, five lessons emerge. First, projects that focus on improving institutional quality before addressing either technical quality issues or dissemination issues are more likely to succeed than projects that first try to improve technical or dissemination quality. Second, for projects that focus on improving technical quality, success is higher when both planning and implementation issues are addressed. Third, in order to give test scores meaning, Bank testing subcomponents need to pay more attention to standard setting, the process of developing an inelastic performance standard. Fourth, more support for dissemination of test results is essential. Fifth, testing subcomponents that are designated as identifiable project components or subcomponents in SARs, with specific budget figures attached, are more likely to succeed than those without such level of detail.
# Table of Contents

I. *Introduction: A New Focus on Education Outputs* ................................................. 1  
   Methodology ........................................................................................................ 4  
      Test purpose .................................................................................................. 5  
      Subcomponent purpose .............................................................................. 5  
      Subcomponent quality ................................................................................. 5  
      Project success ........................................................................................... 7  
   Organization .................................................................................................... 7  

II. *Lending for Testing* .......................................................................................... 10  
   Overview ........................................................................................................ 10  
      Lending Operation ...................................................................................... 10  
      Volume of Projects .................................................................................... 10  
      Lending Commitments ............................................................................ 13  
      Coverage of Testing Issues .................................................................. 13  
      Project Success ......................................................................................... 14  
   Support for Specific Test Purposes ............................................................... 16  
      General ....................................................................................................... 16  
      Trends ......................................................................................................... 17  
      Regional Differences ............................................................................... 19  
   Support for Specific Subcomponent Purposes ............................................ 19  
      General ....................................................................................................... 19  
      Trends ......................................................................................................... 22  
      Regional Differences ............................................................................... 22  
   Summary .......................................................................................................... 23  

III. *The Quality of Testing Subcomponents* ......................................................... 25  
   Institutional Quality ....................................................................................... 26  
      Institutional stability ............................................................................... 28  
      Staff quality ............................................................................................... 28  
      Training capacity ...................................................................................... 29  
      Peer communication ............................................................................... 30  
   Summary ......................................................................................................... 30  
   Technical Quality ........................................................................................... 31  
      Test Development ...................................................................................... 32  
      Survey Development ................................................................................. 35  
      Preparation, Administration and Data Processing .................................. 35  
      Analysis ...................................................................................................... 36  
      Summary .................................................................................................... 37  
   Quality of Dissemination .............................................................................. 39  
      Clear Objectives ......................................................................................... 40  
      Timeliness ................................................................................................... 40  
      Interpretability ........................................................................................... 41  
      Availability ................................................................................................. 41
World Bank Lending for Educational Testing

I. Introduction: A New Focus on Education Outputs

1. The World Conference on Education for All (WCEFA) catalyzed international commitment to ensure that all children be given an equal opportunity to achieve and maintain an acceptable level of learning. This explicit emphasis on the outputs of education was new. To assess progress toward this goal and to facilitate the development of appropriate policies to achieve it, the WCEFA World Declaration on Education for All identifies two essential activities: (a) defining "acceptable levels of learning acquisition for education programs," and (b) improving and applying "systems of assessing learning achievement" (WCEFA 1990, p.5). The WCEFA Framework for Action calls for increased international funding and long-term partnerships to assist countries reach their national goals.

2. Currently, the importance of monitoring both student learning, the key education output, and education inputs, is well recognized by the World Bank. The Bank's FY90 policy paper on primary education recommends national programs for assessing student learning as essential to improve the quality of education in developing countries. Nearly half of all Bank education projects approved in FY91 included activities related to the measurement of student learning. In FY92, 23% of education projects included an assessment subcomponent specifically designed to monitor learning.

3. Until recently, however, most Bank-supported monitoring activities focused on monitoring educational inputs, rather than educational outputs. The focus on educational inputs can be explained by two factors. First, educational projects during early years of Bank lending concentrated on facilitating educational inputs in terms of personnel, buildings and equipment; outputs were not expected during the
project's life because large amounts of time and energy were required to get educational building blocks in place. Thus only educational inputs were used to monitor a project's progress.

4. Second, educational inputs were assumed to be adequate proxies for educational outputs, which were considered too difficult to measure. This assumption is no longer accepted. International evidence has demonstrated that inputs differ so widely in their effectiveness across countries that they are very poor proxies for outputs (Fuller 1987). Furthermore, recent concerns with improving educational efficiency (the ratio of inputs to outputs) and with selecting wisely the most cost-effective educational inputs require separate measures of each (Hanushek 1986; Lockheed and Hanushek 1991). The new emphasis on educational quality in Bank-financed projects and the search for good output indicators are bringing tests into focus.

5. What are tests? Tests are the most common measure of student learning and are typically used to monitor learning. Tests are defined as "any series of questions or exercises or other means of measuring the skill, knowledge, intelligence, capacities or aptitudes of an individual or group" (Anderson, Ball and Murphy 1975). Educational achievement tests focus on skills, knowledge and capacities that are acquired through education (schooling); they do not seek to measure "intelligence" or "aptitude".

6. Educational achievement tests are constructed for many different purposes; among them, the most common six are:
   a. selecting students for further education
   b. certifying student achievement
   c. monitoring achievement trends over time
   d. evaluating specific educational programs or policies
e. holding schools, regions, etc. accountable for student achievement

f. diagnosing individual learning needs.

While in some cases educational achievement tests designed for one of these purposes can be used for another one, such is more often the exception than the rule. In most instances, the use of tests designed specifically for one purpose (such as selection) for another purpose (such as monitoring achievement trends over time) is inappropriate. The purpose of the test determines its design; different test designs are needed for different purposes. For example, selection tests are designed to discriminate between individuals, and hence include many items of a difficulty known to discriminate between those who should be selected and those who should not; as a consequence, the range of competencies tested will be highly constrained. By comparison, a test designed to monitor national achievement is designed to assess a large number of competency areas but to provide little information about specific individuals. Similarly, tests designed to monitor trends over time are designed to repeat at least some items; tests designed only to select once need not be equated. Test misuse is a serious problem and confusion about multiple uses for tests is found in Bank lending for testing.

7. Using tests for more than one purpose can be not only inappropriate but also more costly. For example, using tests for evaluative purposes (monitoring achievement or evaluating specific educational programs or policies) as diagnostic instruments is not only unproductive but it will also increase costs (LeMahieu and Wallace 1986). It is unproductive because evaluative testing instruments are not relevant, timely or brief enough to make for good diagnostic instruments. It is more costly because using evaluative tests for diagnostic purposes implies testing every student in the population; thus, increasing costs of a test that only requires a good sample in order to obtain an accurate estimate of student's performance.
8. Since 1963, the World Bank has supported educational testing in 90 projects in 55 countries (Annex 1). The purpose of this General Operational Review is to take stock of World Bank lending for testing and to examine trends in lending for testing.

Methodology

9. The data for this paper come from project documents for all education and social sector development projects initiated since 1963 that have supported educational testing; a total of 450 projects were reviewed. Very few projects (11%) have free-standing educational testing components; most projects (58%) support educational testing through a subcomponent of an educational quality, institutional strengthening or management component; but many projects (30%) support testing only as a part of a subcomponent that includes elements other than educational testing. In these projects, testing is seen as providing information and feedback to monitor and improve internal efficiency and educational quality and not as a free-standing element of the education system requiring scarce foreign exchange.

10. Since the majority of Bank projects support educational testing through project subcomponents (rather than components), in this paper all project elements related to testing will be referred to as subcomponents. These subcomponents are classified along three dimensions: (a) the purpose of the test, (b) the purpose of the subcomponent, and (c) the quality of the subcomponent. Examples of testing subcomponents classified according to the first two dimensions are provided in Table 1; quality issues

---

1 Staff Appraisal Reports (SARs), President Reports, background documents in project files, project completion reports, project performance audits and draft project documents.

2 Testing subcomponents for evaluation purposes are not included in the table because the subcomponent purpose classifications, which are related to systems are not relevant to these subcomponents. They are not relevant because evaluation subcomponents often only aimed at completing a one-shot study (i.e. they did not consider changing, creating or strengthening systems). The following are examples of testing subcomponents for evaluation purposes: (A) The project finances a study on "whether increased training of teachers results in improved pupil achievement." "Benchmark tests for pupil achievement and attitude, and tests of knowledge and competence" will be used in the study. (Pakistan: Primary Education Project, SAR #2305-PAK, p.26-27.; (b) The evaluation of the project impact and results will be executed in two
are discussed at length in Chapter III.

11. **Test purpose.** The paper classifies the purposes of educational tests according to the six most common purposes listed above; however, two purposes — selection and certification — are combined into a single category to reflect their usage in actual practice, and another category — multiple test purposes - is added. Thus, the six categories are: (a) selecting and/or certifying students, (b) monitoring achievement trends, (c) evaluating effectiveness of specific programs or policies, (d) holding schools/regions accountable for the performance of their students, (e) diagnosing individual student learning needs, and (f) serving more than one purpose.

12. **Subcomponent purpose.** The paper classifies the purposes of the project’s testing subcomponents into three categories: (a) strengthening an existing testing system, (b) creating a new testing system, and (c) preparing the ground for a new system. Subcomponent purposes relate to testing systems and thus rarely apply to subcomponents for evaluating the effectiveness of specific policies. Evaluation subcomponents typically have the single objective of completing a study and were not intended to strengthen, create or prepare the ground for a system for evaluating the effects of specific policies.

13. **Subcomponent quality.** Three aspects of a project’s testing subcomponent were judged: (a) institutional quality, (b) technical quality, and (c) dissemination quality. Because testing systems are complex, and the purposes of tests varied, it is difficult to develop a list of quality indicators common

---

[project states...][8] will concentrate on the following four categories of objectives: (i) *educational achievements of the target population*; (ii) availability and use of physical facilities, textbooks, teachers' guides and teaching/learning materials in project municipalities; (iii) the improvement in the progression rate of the school population in project schools; (iv) the quantity and quality of administrative records kept by the OMEs and the frequency/duration of each supervision paid to project schools, as a means for measuring the impact of the institution building element of the project at municipal level.* (Brazil: Northeast Basic Education Project, SAR#2815-BR, p.33).
<table>
<thead>
<tr>
<th>Test Purpose</th>
<th>Subcomponent Purpose</th>
<th>Preparing for a new system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selecting and/or Certifying Students</td>
<td>&quot;The Credit would support the consolidation of the two existing examination systems by providing the consolidated system with adequate office and computing facilities. This would involve (a) minor capital improvement in the Zambia office or Interim Consulate of Education and Testing Board to accommodate the transfer of professional staff from Lilongwe; and (b) computing and printing equipment to accommodate increased work load and (c) technical assistance in the form of a consultant to develop better test measurement. The legal measures necessary to merge the two systems will be taken prior to credit effectiveness.&quot; (Malawi: Education Sector Credit, SAR #0428-M-1, p.36)</td>
<td>No projects</td>
</tr>
<tr>
<td>Monitoring Achievement Trends Over Time</td>
<td>&quot;SIA/CIS, the Achievement Assessment System, will be strengthened to ensure that: (a) all tests used to measure student learning achievement meet stringent psychometric standards of reliability and validity; (b) information gathering activities are carried out in a timely and efficient manner; (c) results reporting will be easily interpretable and will facilitate improving instructional practice at the school level and policy decision concerning the efficient use of school resources; and (d) results analysis will be done by competent research specialists to provide feedback to school authorities on educational and social significance of the assessment findings.&quot; (Chile: Primary Education Improvement Project, SAR #9769-CH, p.25)</td>
<td>Planning and management will be enhanced through the introduction of an annual assessment program of elementary student performance at the National Educational Testing and Research Center, to routinely provide achievement data for effective monitoring and evaluation.&quot; (Philippines: Second Elementary Education Project, SAR #5995-PH, p.14)</td>
</tr>
<tr>
<td>Holding schools, regions, etc. accountable for student achievement</td>
<td>No projects</td>
<td>No projects</td>
</tr>
<tr>
<td>Diagnosing individual learning needs</td>
<td>&quot;The project will support the further development of the Essential Learning Curriculum and Continuous Pupil Assessment and their introduction into Classes I through V in all schools.&quot; (Bangladesh: General Education Project, SAR #8015-BD, p.15).</td>
<td>To improve the capacity of teachers to diagnose pupil learning difficulties, the project will carry through the following actions: &quot;development of pilot test instruments and training programs for selected teaching staff in using diagnostic tools and defining remedial actions.&quot; (Thailand: Education Sector Reform Program, SAR #5923-MOR, Annex C page 1 of 4)</td>
</tr>
</tbody>
</table>
to all testing subcomponents. Instead, for each test purpose, an appropriate list of "quality indicators" was compiled and the extent to which these indicators were discussed in project documents was recorded. The quality indicators for the institutional requirements for research and analysis comprised those specific to: (a) institutional stability, (b) staff quality, (c) institution-based training, and (d) communications. The technical quality indicators for tests and related surveys comprised those specific to: (a) test development, (b) survey development, (c) preparation, administration and data processing, and (d) analysis. The quality indicators for dissemination systems comprised those specific to: (a) clear objectives, (b) timeliness, (c) interpretability, (d) availability, (e) stimulating demand for test results, and (f) evaluation of information use. Within each group of indicators, up to eight specific indicators were coded; Box 1 describes these in detail.

14. Project success. For completed projects, project completion reports were reviewed for evidence of the success or failure of the educational testing subcomponent. Since only 26 (29%) of the projects are closed, it is difficult to draw detailed conclusions on best practice.

Organization

15. This paper begins by surveying testing in Bank educational projects (extent, purposes, trends, regional patterns). The next section provides greater detail on the quality (institutional quality, technical quality and quality of dissemination) of testing subcomponents for four major test purposes (selection/certification, monitoring trends, evaluating programs and individual diagnosis). The final section discusses lessons learned.
Box 1: Definitions of Quality Issues

Institutional Quality

3. Stable Staff: Encouraging methods to retain qualified staff, such as revising salary scales.

Staff Quality

1. University-based training: Supporting high-level advanced training in relevant fields, such as psychometrics, statistics and survey research (encouraging especially scholarship programs with institutional linkages).

Institution-based training

1. Post doctoral: Supporting institution-based practical training programs that provide opportunities for postdoctoral research.
2. "Extension workers": Providing applied, professional training opportunities for "extension agents" and implementation specialists.

Communication

1. Professional: Providing means for professional staff to communicate with peers (e.g. through journals, library development, travel to professional meetings, networks, and others.)

Technical Quality

Test Development

1. Conceptual Planning: Entails gathering teachers, administrators, specialists and/or consultants to define uses and objectives of tests, design procedures, select test development team, document and disseminate plans.
2. Setting Test Parameters: Determining levels, test format, types of skills, and sampling methods.
3. Writing Test: Selecting grades and subjects; identifying time and item frameworks for each objective; writing and reviewing preliminary items.
5. Pretesting Items: This is an important stage since it not only includes procedures for pretesting items (selecting pre-test groups, distributing tests, training testers, securing tests, and administering tests; collecting, inputting and cleaning data) but also procedures of item analysis and reliability and validity check.
6. Selecting Final Items: Test development team group meets to estimate and document final test items.
7. Final lay-out: Laying out, editing and proofing final items.
8. Standards/Normaling: Tests for monitoring progress toward national educational goals and for selection/certification purposes require setting standards of performance. Norms, basically a standard based on average performance, are useful for national assessments that provide indicators of the system as a whole and of types of students, classrooms, schools, and areas.

Survey Development

1. Conceptual Planning: Determining, documenting and disseminating survey’s uses, objectives and designing procedures. A development team should be identified at this stage.
2. Writing the Survey: Selecting areas of interest, writing and reviewing preliminary questions.
3. Graphic and Printing: Preparing preliminary questions for pre-testing (i.e. typing, lay-out, editing, proofing and printing).
4. **Pretesting questions:** Selecting pre-test groups, training surveyors, administrating and coding of survey, inputing and cleaning data, question reliability and validity check.

5. **Selecting Final Questions:** Survey development team identifies and documents final questions.

6. **Survey Production:** Preparing final survey questions for printing (lay-out, editing, proofing).

**Preparation, Administration and Data Processing**

If surveys are to be used simultaneously with tests, then these may also apply to surveys.

1. **Test Production:** Printing, securing, packaging and distributing tests.


3. **Manual Revision:** Professionally editing and printing manual.

4. **Preparation:** Selecting sample population, registration of test-takers and training of test administrators. If needed, test centers need to be established.

5. **Informing regions and schools of tests:** Writing and disseminating an informative text.

6. **Administration:** Scheduling test administrations; administering and collecting tests in a secure manner.

7. **Processing:** Counting, marking and batching tests.

8. **Scoring**
   - a) **Machine Scoring:** If tests are to be machine scored, then data needs to be inputted, cleaned and processed.
   - b) **Hand Scoring:** If tests are to be hand scored, then tests must be scored, aggregated and recorded.

**Analysis**

1. **Aggregation/Trends:** Aggregating results will be necessary for many tests, especially if results are needed at classroom, school or national levels. Trend analysis can then be carried out.

2. **Policy or Pedagogical analysis:** If tests for assessment of individual performance are to provide guidance for improving pedagogy (e.g. remedial instruction), then teachers must be provided with appropriate tools to analyze test results for this purpose. And if tests are to provide indicators of the effectiveness of specific policies on student achievement, then research must integrate background surveys and test results.

**Quality of Dissemination System**

1. **Clarity of Objectives of Dissemination System:** Clearly defining when dissemination is to be carried out, for what purposes and for what audiences.

2. **Timeliness:** Supporting approaches that increase the speed with which data collection and analysis is carried out.

3. **Interpretability:** Tailoring the format of reports for different audiences, assisting the correct interpretation of scores. The importance of selecting appropriate tests to meet intended purposes should be reinforced.

4. **Availability-Methods of Dissemination:** Using techniques (e.g. seminars, briefings, public media) to make information easily accessible to test-users, such as policy makers.

5. **Stimulate Demand for Information:** Providing means for stimulating test-users demand for test results.

6. **Evaluation of information use:** Assessing dissemination performance to ensure responsiveness to test-users needs and usage patterns.

**Sources:** Ilon (Draft 1992); Lockheed (1991).
II. Lending for Testing

Overview

16. The World Bank invests in educational testing through two major approaches: (a) education projects, such as free-standing primary or secondary education projects, and (b) education components in non-education projects, such as social sector development projects. In both cases, testing subcomponents are often included in educational quality, institutional strengthening or management components rather than as free-standing components. This review examines educational testing in 85 Bank education projects or education components of non-education projects\(^3\) from FY75 (when the first testing subcomponent was supported) through FY92 and 5 projects in FY93\(^4\). The majority of the 90 projects containing testing subcomponents that are reviewed in this paper are currently active (62%); a few are in preparation (7%) and about one-third (31%) are closed.

17. *Lending Operations.* Bank projects support a variety of educational testing purposes, although the majority (83%) of projects provide support for a single testing subcomponent related to a unique test purpose, for example, a project subcomponent to support the development of a new system (the subcomponent purpose) for monitoring progress toward national goals (the test purpose). This chapter reviews overall trends in lending operations, support for specific test purposes and support for specific subcomponent purposes.

18. *Volume of Projects.* The number of projects with educational testing subcomponents, as well as

---

\(^3\) Hereafter referred to as "education projects."

\(^4\) Five projects planned for FY93 were sufficiently well-developed for analysis; two had yellow cover SARs and background documents, one had a green cover SAR and background documents, and two had sufficient background information specific to the testing subcomponent. In order to fairly evaluate these projects which are still in preparation, testing subcomponents were coded wherever possible, but indicators about which no information was available were coded as "missing" rather than as valid cases.
the share of these projects in total education lending, has increased dramatically since the Bank started lending for education. From FY63 to FY74, no projects supported testing subcomponents. In 1975, one project (5% of all education projects) supported these subcomponents. This rose to 12 projects in FY91 and 10 projects in FY92, 46% and 38% of education projects, respectively, approximately doubling every five years (Figure 1). This reflects the Bank’s increasing attention to monitoring the outcomes as well as the inputs of educational systems in developing countries.

19. Projects with testing subcomponents occur in all geographical regions, with 30 projects in Africa, 25 in Asia, 19 in LAC and 16 in EMENA (Figure 2). Between regions, however, the percentage of projects with testing subcomponents differs sharply. Since FY75, a higher proportion of projects in LAC have included testing (33%) than have projects in other regions (25% of projects in Africa, 25% of projects in Asia and 23% of projects in EMENA). In part, this is due to the recent renewed lending for education in Latin America, which coincided with the growth in lending for testing.

20. Despite their widespread inclusion in Bank projects, educational testing subcomponents rarely are prepared in detail; this affects both the level of lending commitments for educational testing and the coverage of testing issues in SARs. For earlier projects, it also seems to be related to the success of the testing subcomponent.
Countries with World Bank Supported Projects with Testing Subcomponents
1975-1992
21. **Lending Commitments.** Exact lending commitments for educational testing are difficult to determine because few projects provide the necessary budget details. Only 22 (24%) of projects reviewed provided sufficiently detailed costing information to isolate lending commitments for educational testing. Among those projects that provide cost figures for testing subcomponents, few provide disaggregated information. Only 18% of projects provide costs for any individual testing subcomponent (Figure 3).

In general, testing subcomponents are costed as part of a project component (47%) or as part of a subcomponent (16%) that includes other elements than testing.

22. For the projects for which costs could be identified, the level of financial support for educational testing was modest, ranging from $70,000 to $11 million, and from .10% to 8% of total project costs, with 14 (64%) of these projects allocating fewer than 2% of total project costs to educational testing (Figure 4).

23. **Coverage of Testing Issues.** Educational testing issues are treated very meagerly in both SARs and their related background documents. Only 8% of SARs allocate as much as one page to discussion of testing issues, while 61% of SARs treat educational testing issues in fewer than 200 words, or less than a half page. The lack of detail in
SARs may be explained by their summary nature; however, background documents, which is where such detail should be found, also lacked detail on testing issues. Background documents for only 24% of projects specifically dealt with testing issues. Twenty-seven percent of projects were informed by background documents that treated testing issues in no more than a single page of text and 41% did not refer to testing at all (Figure 5).

24. Trends in coverage of testing issues in SARs and background documents are discernable, however, with projects from recent years having substantially greater detail. The percentage of projects with extensive coverage of testing issues in the SAR or a background document increased from 4% in FY75-79 to 16% in FY90-92. Even in the most recent periods, more than three quarters of all projects with testing subcomponents do not report key design features. (Figure 6).

25. Project Success. For most projects, it is too early to judge success. However, for some of the earliest projects, Project Completion Reports (PCRs) provide some evidence. Project Completion Reports were available for 23 of the 26 closed projects which discussed testing subcomponents in their SARs. Testing was addressed in 20 of these PCRs. Of these 20 projects, 10 dealt with selection/certification tests only, 8 dealt with testing for evaluating the effectiveness of specific programs or policies, and two dealt with both test purposes.

---

5Extensive coverage in an SAR is defined as more than a page of text devoted to testing; extensive coverage in a background document is defined as a background document dealing specifically with testing issues.
In general, if a project had been implemented, the PCR indicated that the test subcomponent was successful.

26. The PCRs for only four implemented projects indicated that the testing subcomponent had been unsuccessful (Haiti 83, Kenya 78, Pakistan 79 and Pakistan 81). The reasons for lack of success included political problems, institutional weakness, technical faults, and procurement delays. Political upheavals and instability in government limited the progress and continuity of Haiti's FY83 project's testing subcomponent. Kenya's FY78 testing subcomponent did not achieve the anticipated institutional strengthening because technical assistance was delayed or not provided, computer equipment was not procured until the very end of the project, and the government and the technical assistance agency lacked coordination. Technical and research design problems restricted the validity of the evaluations carried out in Pakistan's FY79 testing subcomponent; the expected output of Pakistan's FY81 testing subcomponent was never achieved because the testing agency's capacity had been overestimated during project appraisal.

27. Most PCRs provided little information about strategies that contributed to the success of the testing subcomponent. However, a few projects indicated areas of success around institutional quality, technical quality and quality of dissemination. For example, institutional quality was enhanced in Liberia by providing technical training for members of the staff of the West African Examinations Council and,
in Kenya, by supporting long-term technical consultants in the areas of research, statistics and systems analysis. Technical quality of tests developed for the evaluation of a project in Bangladesh was assured by contracting the University of Dhaka to develop the tests according to acceptable procedures for item preparation and selection and by monitoring test reliability. Test security in Zambia was enhanced by the provision of a special strong room for storage examination papers. Efficiency of testing in Liberia was ensured by providing printing equipment which reduced the cost of printing tests, improved their security and resulted in on-time delivery of tests. Finally, quality of dissemination was enhanced through providing support for routinely collecting school-based data and using these data to inform education investment decisions.

28. PCRs provide more information on reasons for failure than on reasons for success, much of which information deals with problems of implementation. With respect to poor institutional quality, PCRs noted that technical assistance was under-utilized or inappropriately utilized, training fellowships were not used, and appropriate institutions were not given responsibility for the work. Failure of testing subcomponents is generally attributable to failure of the project to develop an appropriate institution to carry out the technical and dissemination subcomponents. With respect to poor technical quality, PCRs noted that research designs were overly complex or failed to adjust for local conditions, tests were not printed or administered on time, data were not collected, and analyses were not conducted as a result of failing to maintain computers.

Support for Specific Test Purposes.

29. General. Thirty-three percent of projects with testing subcomponents support tests for student certification and selection, 21% support tests for monitoring student progress toward national educational goals, 19% support tests for evaluating the effectiveness of specific policies, 7% support tests for teacher
diagnosis of student learning, 2% support tests for mixed or multiple purposes, while 17% of projects include more than one testing subcomponents (Figure 7).6

30. Tests are supported at all levels of education (basic, secondary, tertiary) and a single subcomponent frequently supports tests at more than one level. Most monitoring (96%), evaluation (86%) and diagnostic tests (93%) are prepared for basic education. Selection/certification tests, in comparison, are more evenly supported across educational cycles, with 43% prepared for basic education, 40% for secondary, 38% for vocational, and 13% for higher education (Figure 8).

31. *Trends, FY75-FY92.* Trends in the type of tests supported in Bank projects are apparent (Figure 9). The percentage of projects that support tests for selection and certification has been steadily growing over the past 15 years, while the percentage of projects that support tests for evaluating the effectiveness of specific policies has declined. Support for tests both for monitoring progress toward national education goals and for teacher diagnosis of student learning has increased sharply over the past decade. Monitoring has shown the greatest increase, growing from 0% of projects before FY88 to 27% of projects in FY91. The increase in monitoring reflects greater attention to strengthening the capacity of governments, particularly Ministries of Education, to gather education output information for purposes of policy analysis and improved management.

---

6 Although certification and selection are conceptually distinct, most examination systems in developing countries use a single test for both purposes, and we have therefore combined these purposes in this discussion.
Educational Level of Testing Subcomponents
32. Regional Differences. Regional differences in test purposes are also apparent (Figure 10). In Africa and Asia, nearly half (46% and 44%, respectively) of all projects support tests for selection and certification, while in EMENA and LAC the largest share of projects support tests for monitoring achievement trends (29% and 40%, respectively).

33. These regional differences reflect historical patterns of education in the region. African and Asian education systems have included national selection and certification examinations at all levels (Figure 11), and Bank lending has been designed to strengthen existing examinations. Where national examinations have played a lesser role in educational systems, as in LAC, Bank lending has been designed to support new systems of monitoring and assessment.

Support for Specific Subcomponent Purposes.

34. General. Three purposes of testing subcomponents in Bank-supported projects are to change an existing system of testing, create a new system of testing or prepare for a new system of testing. Most projects (52%) include testing subcomponents designed to change an existing system; the projects related to creating a new system account for 34% of projects with testing subcomponents; very few projects (3%) included testing subcomponents related to preparing for a new system. Eleven percent of projects include more than one subcomponent.
Test Purposes of Testing Subcomponents in World Bank Supported Projects in AFRICA, ASIA, EMENA and LAC
Figure 11.

Percentage of AFRICAN, ASIAN, EMENA, LAC countries that have Selection/Certification Examinations
35. Subcomponent purposes and test purposes are related. Bank projects typically support strengthening testing systems for selection/certification tests and creating new testing systems for monitoring tests. Of the 40 projects with selection/certification testing purposes, 31 (78%) envisioned changing the existing system by strengthening examination institutions and/or improving instruction by improving the scope and content of selection examinations; some evidence from Kenya supports this expectation (Somerset 1987). Of the 26 projects with testing purposes related to monitoring learning achievement, 15 (58%) intended to create a new system (Figure 12).

36. Trends, FY75-FY92. An examination of trends, FY75-FY92, shows a steady increase in all types of testing subcomponents. Projects with testing subcomponents aiming at changing a new system increased from 5% during FY75-FY79 to 21% during FY90-92, and those with subcomponents for creating a new system increased from 1% to 15% during the same periods. Subcomponents related to preparation for a new system are evident only in the latest period, however (Figure 13).

37. Regional Differences. Regional differences in subcomponent purposes are substantial and reflect the relationship between subcomponent purposes and test purposes. Projects in LAC and EMENA have the highest percentage of subcomponents related to creating a new testing system because they have the largest share of projects supporting tests for monitoring trends, which are new endeavors. In contrast, Africa and Asia have the highest percentage of projects supporting selection/certification tests and thus,
have the largest share of subcomponents related to changing a testing system. The two subcomponents associated with preparing for a new system are found in Africa and LAC (Figure 14).

Summary

38. World Bank support for educational testing has increased significantly in the recent past, catalyzed by growing evidence regarding the poor performance of educational systems in developing countries, the desire by countries to manage their education systems better and more efficiently, the 1990 World Conference on Education for All’s call for national systems for monitoring learning, and the need to monitor the impact of Bank education lending operations on student learning. Projects that include testing subcomponents are found in all regions and address a wide variety of purposes. The major type of test that is supported is tests for student selection or certification, although testing subcomponents related to tests for monitoring achievement trends have become a focus of attention in projects in the most recent years. The principal purpose of testing subcomponents is to change an existing system of testing. The designs of testing subcomponents rarely are elaborated fully, providing little detail about the cost or quality of the testing subcomponent.
Regional Differences in Subcomponent Purpose

AFRICA

Change System 57
Prepare New System 3
Create New System 40

ASIA

Change System 72
Create New System 28

EMENA

Change System 36
Create New System 64

LAC

Change System 47
Prepare New System 6
Create New System 47
III: The Quality of Testing Subcomponents

39. Testing subcomponents were judged according to three dimensions of quality: institutional quality, technical quality, and quality of dissemination. As the number of projects supporting testing subcomponents has increased, so has the general quality of the subcomponents, with quality improving dramatically in the past two years. The average percentage of quality issues addressed in Bank projects grew from less than 10% in FY75-FY79 to nearly 35% in FY90-FY92. Moreover, the quality of projects has improved more sharply in the past few years than the percentage of projects with testing subcomponents, suggesting that projects with testing subcomponents are better designed since FY90. Aside from two projects with well-developed testing subcomponents related to evaluation7, the average percentage of quality issues addressed in Bank projects remained below 20% until FY90. However, even with 40% of quality issues addressed, the percentage of quality issues not addressed remains disturbingly high (Figure 15).

40. The improvement in quality across time can be explained by greater (a) ministerial commitment to quality and willingness to change institutions; (b) international awareness of the value of educational testing; (c) national experience with testing and evaluation culture; (d) availability of local expertise; (g) capacity and quality of technical assistance agencies; and (h) Bank task managers’ understanding of testing issues and commitment to change. In recent years, educational testing has been given more importance in education projects and task managers’ understanding of testing issues has been reinforced through training seminars.

7Brazil: Northeast Basic Education I (FY80) and Bangladesh: Fourth Education Project (FY80)
Significant variation across projects remains; however, the present report was not able to explain this variation.

41. Technical quality issues are addressed more often than either institutional quality or quality of dissemination, but in FY90-92 these differences diminished greatly in comparison with earlier years (Figure 16). Projects providing good coverage of these issues are listed in Table 2.

**Table 2: Projects with High Coverage of Quality Issues**

<table>
<thead>
<tr>
<th>Project</th>
<th>Test purpose</th>
<th>Average percentage of issues covered</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Institutional Quality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lesotho FY92: Education Sector Development</td>
<td>Select/Certify</td>
<td>83</td>
</tr>
<tr>
<td><strong>Technical Quality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lesotho FY92: Education Sector Development</td>
<td>Diagnose</td>
<td>100</td>
</tr>
<tr>
<td>Mexico FY92: Primary Education Project</td>
<td>Monitor</td>
<td>97</td>
</tr>
<tr>
<td>Philippines FY91: Second Elementary Education</td>
<td>Monitor</td>
<td>93</td>
</tr>
<tr>
<td>Indonesia FY85: Secondary Education and Management</td>
<td>Select/Certify</td>
<td>92</td>
</tr>
<tr>
<td>Brazil FY80: Northeast Basic Education I</td>
<td>Evaluate</td>
<td>91</td>
</tr>
<tr>
<td>Mozambique FY91: Second Education Project</td>
<td>Monitor</td>
<td>91</td>
</tr>
<tr>
<td>Lesotho FY92: Education Sector Development</td>
<td>Select/Certify</td>
<td>88</td>
</tr>
<tr>
<td>Madagascar FY90: Education Sector Reinforcement Project</td>
<td>Monitor</td>
<td>84</td>
</tr>
<tr>
<td>Belize FY92: Education and Training</td>
<td>Select/Certify</td>
<td>83</td>
</tr>
<tr>
<td>Chile FY92: Education IV</td>
<td>Monitor</td>
<td>80</td>
</tr>
<tr>
<td><strong>Dissemination Quality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chile FY92: Education IV</td>
<td>Monitor</td>
<td>100</td>
</tr>
<tr>
<td>Ecuador FY92: First Social Development Project</td>
<td>Monitor</td>
<td>83</td>
</tr>
</tbody>
</table>

**Institutional Quality**

42. A major focus of educational testing subcomponents in Bank lending is *improving institutional quality*; the target institution is often one that is responsible for the development and administration of national tests and examinations. Because testing for teacher diagnosis in developing countries rarely
involves standardized, centrally developed and administered tests, testing subcomponents related to teacher testing for student diagnosis rarely include institutional strengthening as an objective. Three types of institutional strengthening objectives characterize lending for testing: (a) improving existing testing institutions, (b) building new testing institutions where none currently operate, (c) preparing for the development of a new system. Bank projects support testing subcomponents related to all three objectives.

43. Effective educational testing institutions have characteristics common to other applied science institutions. They are stable, are staffed by professionals who have received high-level doctoral, postgraduate training, provide master's level training opportunities for junior staff and "extension workers" such as test administrators, and enable their professional staff to communicate with their peers internationally. Examples of such testing institutions from North America include Educational Testing Service and the American Council on Testing; examples from developing countries include the Caribbean Examinations Council and the National Educational Testing and Research Center in the Philippines. Strengthening or establishing national testing institutions involves creating the research capacity necessary for these institutions to carry out their responsibilities for developing a variety of tests that are valid, reliable and can be used for a range of purposes, including individual student certification and selection and in policy analysis.
44. Thus, support for institutional strengthening typically seeks to improve: (a) institutional stability, (b) staff quality, (c) institutional capacity for in-house training, and (d) opportunities for professional peer communication (Table 3). On average, monitoring subcomponents have addressed a larger percentage of these institutional strengthening issues than selection/certification and evaluation subcomponents. Monitoring subcomponents address approximately 34% of institutional quality issues; whereas selection/certification and evaluation subcomponents address fewer than 15% of these issues (Figure 17).

45. Institutional stability. Project subcomponents provide little support for strengthening the stability of testing institutions. Three key indicators of institutional stability are: (a) the clarity of the institution's objectives, (b) the intensity and duration of its financial support and (c) the tenure of its staff. Overall, 38% of testing subcomponents in Bank projects had clearly specified objectives; clarity of objectives differed across testing uses, with greatest clarity of institutional objectives in projects for monitoring (70%), followed by those for selection/certification (38%) and least for evaluation (9%). A maximum of six percent of subcomponents provided long-term financial support for institutions, and a maximum of 10% of subcomponents provided support for staff stability. Thus, while projects typically identified a testing institution's raison d'être, they rarely provided the necessary financial support for a stable institution.

46. Staff quality. A second requirement for an institution to carry out the type of technical work
required to ensure the technical quality of its product (in this case tests and related surveys) is a "critical mass" of professionals with high levels of advanced training. In developing countries, where the total number of professionally trained individuals in any given field is relatively small, highly trained individuals are often drawn away from educational institutions, including ministries of education and related agencies, to the more favorable salaries and working conditions in the private sector. To retain staff, special efforts must be made to make government employment more competitive with employment in the private sector. However, no testing subcomponents supported this aspect of assuring staff quality. Instead, projects supported the development of staff quality, largely through the provision of university-based training for existing staff. Again, projects with monitoring subcomponents showed the greatest attention to this aspect of institutional development. 48% of monitoring projects provided for university training, 25% of subcomponents related to selection/certification testing provided for university-based training, and no project subcomponents related to evaluation did so.

47. Training capacity. A third aspect of a stable institution is its ability to replicate itself through its training programs. Two types of training are typically carried out in applied research institutions, such
as testing agencies: practical training for recent doctoral recipients whose formal training has been university based and largely theoretical, and the direct training of master’s level students and "extension workers." Much of this training is carried out in the context of actual work, such as that involved in developing and analyzing tests. Fewer than 25% of testing subcomponents provided support for training for newly trained researchers, thus limiting the possibility that support for scholarships at institutions outside the country would have a direct impact on increasing the capacity of the country to ultimately provide its own professional training in the area of testing. Only training for "extension workers" such as teachers received any significant support, with over 40% of testing subcomponents related to monitoring providing training for teachers.

48. **Peer communication.** A fourth aspect of a stable institution is that it facilitates communication among its professional staff and similar professionals working elsewhere; mechanisms include travel, libraries, journals, research abstracting services and institutional "twinning" arrangements in which an institution in a developing country has a continuing relation with one in a developed country. Support for peer communication is seldom included in Bank projects. Only 5% of testing subcomponents supporting tests for selection/certification or for evaluation have supported professional peer communication. Testing subcomponents supporting monitoring are more likely to include peer communications, with 45% of such subcomponents providing support.

49. **Summary.** Most testing subcomponents in Bank projects do not provide support for sustained institutional strengthening. Lesotho’s Education Sector Development project is one of the few projects that show significant support for institutional development (Box 2). Testing subcomponents in support of monitoring, on average, show the most support for institutional strengthening, but these subcomponents are typically too new to assess their long-term effects.
Box 2. Lesotho Education Sector Development Project: Institutional Quality

As part of its Five-Year Education Development Plan and with the support of the World Bank, Lesotho will strengthen the capacity of the Examination Council of Lesotho (ECOL). Approximately 2.6% ($1.9 million) of Lesotho’s FY92 Education Sector Development project costs are directed to this effort. Strengthening ECOL’s capacity involves addressing issues related to institutional stability and staff quality, training and communication.

Institutional Stability. Stable research institutions have clear objectives, low staff turnover rates and are financially stable. In order to use resources efficiently, ensure that exams reflect the curricula adequately, and increase the general quality of exams, Lesotho plans to establish a single institutional focus for examinations. Therefore ECOL’s institutional objectives have been expanded to encompass not only the management of the Junior Certificate Examination but also the administration of the Primary School Leaving Certificate. ECOL will be responsible for preparing the exams, registering students, printing and distributing tests, appointing test supervisors and markers and preparing and publishing results. Although ECOL’s institutional objectives and responsibilities are clearly defined, the institution is not being provided with complementary means for ensuring institutional stability. Methods to assure long-term financial support and avoid high staff turnover rates are not promoted. Thirty-three percent of institutional stability issues are addressed.

Staff Quality. Successful research institutions are also characterized by the quality of their staff. To improve ECOL’s capacity, the project will finance foreign fellowships for four ECOL staff. These staff members will develops skills in disciplines related to test development. One-hundred percent of staff quality issues are addressed.

Institution-based Training. Creating the ability to replicate itself through training programs is essential to stable research institutions. The project will foster on-the-job training by supporting two long-term (36-staff months each) foreign consultants, one examinations specialist and one measurement expert. Experiences in other countries have shown the benefits of ensuring long-term commitments of foreign consultants whose principal task is on the job training. Post-doctoral practical training will be stimulated through ECOL’s research and development program, which will receive approximately 44% of funds allocated to various research activities. One-hundred percent of institution-based training issues are addressed.

Communication. Another important aspect of well-established research institutions is that they facilitate good communication among researchers working elsewhere. ECOL will attempt to do the same by establishing a reference library on testing, estimated to cost $US 10,700. One-hundred percent of communication issues are addressed.

Source: Levine, Victor (Draft 1990); Lesotho: Education Sector Development Project, SAR#9529-LSO.

Technical Quality

50. A second major emphasis of educational testing subcomponents is improving the technical quality of tests. The process of creating high quality tests involves improving: (a) test development, (b) preparation, administration, and data processing and (c) data analysis. In addition, the development of
background surveys can be a vital complement to testing subcomponents for monitoring or evaluating purposes.

51. Issues related to the four aspects of technical quality differ for each test purpose. All twenty-four identified issues (eight for test development, six for background survey development, eight for preparation, administration and data processing, and two for data analysis) are relevant to tests for monitoring purposes. Twenty-two are pertinent to tests for evaluation, 17 to selection/certification tests, and six to tests for individual diagnosis (Table 4).

52. In general, technical quality is only partially addressed in Bank supported testing subcomponents. The coverage of technical quality issues is highest for monitoring subcomponents with 38% of the issues covered. Fewer than a third of the relevant technical quality issues are covered for selection/certification, evaluation and diagnostic testing subcomponents (Figure 18). This section examines coverage of technical issues in these Bank testing subcomponents, highlighting issues that are particularly important, most emphasized, or greatly disregarded.

53. Test Development. Test development issues receive little attention in project SARs and background documents. On average, 35% of test development issues are addressed in subcomponents for monitoring, evaluating, selecting/certifying, and diagnosing individual learning achievements. Approximately three out of eight relevant issues (38%) are covered for
<table>
<thead>
<tr>
<th>Issues</th>
<th>Select/Certify</th>
<th>Monitor</th>
<th>Evaluate</th>
<th>Diagnose</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Development</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conceptual Planning</td>
<td>54%</td>
<td>78%</td>
<td>41%</td>
<td>15%</td>
</tr>
<tr>
<td>Setting Test Parameters</td>
<td>50%</td>
<td>86%</td>
<td>32%</td>
<td>23%</td>
</tr>
<tr>
<td>Writing the test</td>
<td>43%</td>
<td>73%</td>
<td>29%</td>
<td>8%</td>
</tr>
<tr>
<td>Graphic &amp; Printing</td>
<td>23%</td>
<td>27%</td>
<td>9%</td>
<td>NR</td>
</tr>
<tr>
<td>Pretesting Items</td>
<td>39%</td>
<td>59%</td>
<td>27%</td>
<td>NR</td>
</tr>
<tr>
<td>Selecting Final Items</td>
<td>28%</td>
<td>32%</td>
<td>18%</td>
<td>NR</td>
</tr>
<tr>
<td>Final lay-out</td>
<td>20%</td>
<td>27%</td>
<td>5%</td>
<td>NR</td>
</tr>
<tr>
<td>Standards/Norming</td>
<td>5%</td>
<td>14%</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td><strong>Survey Development</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conceptual Planning</td>
<td>NR</td>
<td>50%</td>
<td>18%</td>
<td>NR</td>
</tr>
<tr>
<td>Writing the Survey</td>
<td>NR</td>
<td>41%</td>
<td>18%</td>
<td>NR</td>
</tr>
<tr>
<td>Graphic &amp; Printing</td>
<td>NR</td>
<td>27%</td>
<td>5%</td>
<td>NR</td>
</tr>
<tr>
<td>Pretesting Questions</td>
<td>NR</td>
<td>41%</td>
<td>14%</td>
<td>NR</td>
</tr>
<tr>
<td>Selecting final Questions</td>
<td>NR</td>
<td>23%</td>
<td>9%</td>
<td>NR</td>
</tr>
<tr>
<td>Survey Production</td>
<td>NR</td>
<td>23%</td>
<td>5%</td>
<td>NR</td>
</tr>
<tr>
<td><strong>Preparation, Administration And Data Processing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Production</td>
<td>48%</td>
<td>64%</td>
<td>27%</td>
<td>NR</td>
</tr>
<tr>
<td>Test Manual Development</td>
<td>18%</td>
<td>50%</td>
<td>14%</td>
<td>NR</td>
</tr>
<tr>
<td>Test Manual Revision</td>
<td>3%</td>
<td>19%</td>
<td>0%</td>
<td>NR</td>
</tr>
<tr>
<td>Preparation</td>
<td>38%</td>
<td>41%</td>
<td>23%</td>
<td>27%</td>
</tr>
<tr>
<td>Informing regions and schools of Test</td>
<td>18%</td>
<td>35%</td>
<td>9%</td>
<td>NR</td>
</tr>
<tr>
<td>Administration</td>
<td>40%</td>
<td>55%</td>
<td>32%</td>
<td>NR</td>
</tr>
<tr>
<td>Processing</td>
<td>35%</td>
<td>41%</td>
<td>32%</td>
<td>NR</td>
</tr>
<tr>
<td>Scoring</td>
<td>a) Machine Scoring</td>
<td>13%</td>
<td>32%</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>b) Hand Scoring</td>
<td>23%</td>
<td>9%</td>
<td>9%</td>
</tr>
<tr>
<td><strong>Analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggregations/Trends</td>
<td>21%</td>
<td>41%</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Policy or Pedagogical Analysis</td>
<td>NR</td>
<td>64%</td>
<td>41%</td>
<td>36%</td>
</tr>
</tbody>
</table>

NR = NOT RELEVANT
monitoring and selection/certification subcomponents, about two out of seven (29%) for evaluation subcomponents, and about one out of three (33%) for diagnosing individual students.

54. Conceptual planning, setting test parameters, and writing the test are pertinent issues for all four test purposes. All types of subcomponents give the greatest attention to issues of conceptual planning and setting test parameters; yet only for monitoring subcomponents do more than 50% of subcomponents address these issues. Writing the test is also given comparatively better coverage than other issues except in subcomponents for diagnosing individual student learning needs. Only 8% of diagnostic subcomponents consider writing the test an important issue.

55. Test reliability and validity are essential for evaluating the technical quality of tests for monitoring, selection/certification, and evaluation purposes. A high degree of reliability (test data consistency and generalizability) and an acceptable level of validity (whether or not the test measures what it is supposed to and whether appropriate inferences can be made from its results) are needed before test results can be used to make decisions. High levels of reliability are especially important for high stakes tests such as those for selection purposes. However, in spite of the importance of reliability and validity checks, only 39% of selection/certification and less than a third of evaluation subcomponents express concern over these crucial issues. A higher percentage of monitoring subcomponents propose to carry out reliability and validity checks (59%).

56. Monitoring and selection/certification tests require setting standards of performance (i.e. the specific levels of proficiency that test takers are expected to attain). Tests for monitoring purposes require an inelastic performance standard, such as is provided by norming (which is basically setting a standard based on the performance of an average sample of students, the "norming sample") or by establishing criterion-referenced standards. Inelastic performance standards are also useful for comparing students,
classrooms, schools and regions. Bank testing subcomponents, nevertheless, give the least attention to issues of standard setting. Less than 15% of monitoring and 5% of selection/certification subcomponents mention these issues.

57. **Survey Development.** Assessment systems for monitoring trends and tests for evaluation purposes benefit from surveys that collect supplementary measures of student background, and classroom and school inputs and processes. In fact, supplementary measures are indispensable for evaluation subcomponents, because measures of the policies or programs in question must be obtained. Yet, only 11% of evaluation subcomponents supported by the Bank address issues concerning survey development. Paradoxically, survey development is addressed in a higher percentage of subcomponents related to monitoring (29%) than to evaluation, even though they are somewhat less important to this subcomponent.

58. **Preparation, Administration and Data Processing.** After tests and necessary surveys are developed, they must be administered. High quality test administration requires preparation for test administration (e.g. test production, test manual development and revision, registration of test-takers, training of test administrators, and informing regions and schools of tests). It also requires that tests be collected in a secure manner, processed and scored. Scoring can be done by hand or by machine.

59. All eight of these issues are important for monitoring, selection/certification, and evaluation subcomponents, but only two (preparing teachers for test administration and training teachers on test scoring) are important for diagnostic testing subcomponents. Coverage of these issues is slightly better for monitoring subcomponents where, on average, three out of eight identified issues are covered. Evaluation and selection/certification subcomponents cover only two out of the eight. Training teachers to administer and to hand score tests is fundamental for diagnostic purposes, but just 27% of diagnostic
subcomponents propose to prepare teachers for test administration and 15% to train teachers on test scoring.

60. Selection/certification, monitoring, or evaluation subcomponents require that test scores be comparable. Thus, test administration and scoring (as well as the test itself) must be standardized. This requires training test administrators, scheduling and administering the tests, informing regions and schools of tests, and scoring tests by machine or by uniform hand scoring procedures. Subcomponents for these three test purposes place greater emphasis on training test administrators, scheduling and administering tests, and processing and scoring answer sheets than on informing schools and regions of tests. Only 35% of monitoring, 18% of selection/certification, and 9% of evaluation subcomponents aim at writing and disseminating valuable information about tests and their administration.

61. Furthermore, selection and evaluation subcomponents give little attention to test manual development especially in comparison to monitoring subcomponents. If all eight issues were ranked from the highest percentage coverage to the lowest, then manual development would rank third for monitoring subcomponents but sixth for selection/certification and evaluation subcomponents. Less than a quarter of any of the three subcomponents, however, contemplate revising the manual.

62. Analysis. After tests have been administered and scored, test results need to be appropriately processed and analyzed to serve specific purposes. This may involve providing score results for individuals, analyzing individual score results for classroom instructional guidance, aggregating scores for groups of test-takers, or analyzing aggregate scores for policy analysis.

63. The purpose of the test determines the type of appropriate analysis. For monitoring trends over time or providing information at classroom, school or national levels, test scores must be aggregated.
By comparison, for selecting students or certifying their accomplishment, test scores must be reported individually. Trend analysis with selection/certification tests is rarely feasible because the tests often are not comparable over time, even though policy makers, school or university administrators and other users of selection tests find aggregated results of interest. Forty-one percent of Bank supported monitoring subcomponents provide specific reference to compiling all test results for trend analysis. Considering that aggregation is less important for selection/certification tests, it is not surprising that less than a quarter of these subcomponents refer to score aggregation.

64. Policy analysis stemming from score results is the essence of evaluation and monitoring subcomponents. Analyzing the effectiveness of specific policies on student achievement requires integrating score analysis with measures of the policies or programs in question. Yet fewer than half of Bank evaluation subcomponents provide information on how policy analysis is to be carried out. In comparison, 64% percent of monitoring subcomponents allude to analyzing scores for informing policy-making.

65. Pedagogical analysis and not policy analysis is important for diagnosing individual learning achievements. Teachers need to know how to use test results for pedagogical purposes (i.e. identifying individual students' strengths, weaknesses and educational progress and informing decisions on teaching strategies). A little over a third (36%) of Bank diagnostic testing subcomponents aim at providing teachers with appropriate tools for analyzing test scores.

66. Summary. Certain issues, such as test reliability and validity, setting standards, uniform administration and scoring, and appropriate analysis, are critical to the development of high quality tests. Bank supported testing subcomponents have not given adequate attention to these issues and even monitoring subcomponents, which provide the best coverage overall, overlook crucial issues such as
Box 3. Mexico Primary Education Project: Technical Quality

Assessments of the quality of education have been conducted in Mexico for the past twenty years. Yet it was not until recently, as a result of the Education Modernization Program (1989-1994), that the Secretariat of Public Education's Directorate for Evaluation (DGEIR) has made special efforts to systematize Mexico's student assessment system in all 31 states. The system will help monitor progress toward national goals and the World Bank will help finance its expansion to the states of Chiapas, Guerrero, Hidalgo and Oaxaca. One percent ($3.5 million) of project costs for Mexico's FY92 Primary Education project are being allocated to this testing subcomponent.

The 67-page plan for the testing subcomponent provides sound coverage of issues related to the technical quality of assessment systems. Twenty-three of 24 important issues concerning test development, survey development, preparation, administration and data processing, and analysis are addressed.

Test Development. A system of tests in all primary education grades will be instituted, encompassing a total of forty-eight different tests measuring cognitive achievement for four subjects (math, Spanish, social and natural sciences). Half of these tests will correspond to the present program of studies and half to the new programs that will be soon be implemented. The steps for the development of these tests are clearly delineated in the plan of action. Issues of conceptual planning, setting test parameters, writing the test, printing preliminary tests, pretesting items, carrying out item analysis and reliability and validity check, selecting final items, preparing tests for printing and the creation of standards are all addressed. One hundred percent of test development issues are addressed.

Survey Development. One of the testing system objectives is to compare student achievement by geographical area, gender, type of school and school inputs, and to investigate what factors are associated with student achievement. Thus, additional data relative to school inputs, family and community characteristics and social and economic environment will be gathered. Surveys for teachers, parents and classroom processes will be developed for each grade; that is, a total of 18 survey guides will be created. One school director interview guide will be developed and administered to each school. Steps for planning, writing, printing and pretesting questions, selecting final questions and preparing the final version are specified. One hundred percent of survey development issues are addressed.

Preparation, Administration and Data Processing. Plans for test production, test manual development, preparation for test administration, informing regions and schools of tests, test administration, test processing, and computerized scoring of tests are proposed; and to support these processes, the project will finance data collection and data processing equipment. One issue, the revision of test manuals, is left unaddressed. Eighty-eight percent of preparation, administration and data processing issues are addressed.

Analysis. The plan also defines processes for test score aggregation and analysis. In fact, four specific types of inferential analysis are specified and research on student achievement will integrate background data to control for school and non-school factors. The project will finance these analyses and the distribution of information. One hundred percent of data analysis issues are addressed.


standard setting. Coverage of technical issues, however, has improved since 1975, increasing from 0% coverage in FY75 to 56% in FY92. Mexico's Primary Education project is an example of a FY92 project that provides good coverage of technical quality issues (Box 3).
Quality of Dissemination

67. A third major emphasis of educational testing subcomponents is building **efficient dissemination systems**, which are more important for monitoring, evaluation and selection/certification tests than for classroom-level diagnostic testing. As teachers carry out the diagnosis of individual learning needs they typically obtain immediate feedback on student performance. By comparison, policy-makers, teachers, parents and other users of tests for monitoring, evaluation, and selection/certification are distanced from processes of test development, administration and scoring. This distance means that special efforts must be made to communicate testing results to them. A continuous dialogue between test producers and users regarding appropriate tests and score interpretation can be supported through high quality dissemination.

68. A high quality dissemination system is characterized by: (a) having clear objectives, (b) providing timely information, (c) supplying easily interpretable results, (d) using different communication forms, (e) stimulating demand for test results and (f) having mechanisms for ensuring responsiveness to test-users' needs and usage patterns (Table 5). Bank testing subcomponents address, on average, fewer than 30% of these issues, with monitoring subcomponents covering the highest percentage (28%) of issues. (Figure 19). Only monitoring subcomponents have shown considerable improvement in coverage of dissemination issues, from 0% coverage in FY88 to 58% in FY92. The following section discusses each of these six crucial elements and their coverage in Bank testing subcomponents for the three test purposes that are relevant to these concerns: selection/certification, monitoring and evaluation.

![Figure 19. Coverage of Dissemination Quality Issues](image)
Table 5: Percentage of Testing Subcomponents Covering Dissemination Quality

<table>
<thead>
<tr>
<th>Issues</th>
<th>Test Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Select/Certify</td>
</tr>
<tr>
<td>Clear Dissemination Objectives</td>
<td>23%</td>
</tr>
<tr>
<td>Timeliness</td>
<td>10%</td>
</tr>
<tr>
<td>Interpretability</td>
<td>5%</td>
</tr>
<tr>
<td>Availability-Methods of Dissemination</td>
<td>NR</td>
</tr>
<tr>
<td>Stimulate Demand for Information</td>
<td>NR</td>
</tr>
<tr>
<td>Evaluation of Information Use</td>
<td>5%</td>
</tr>
</tbody>
</table>

NR = Not Relevant

69. **Clear Objectives.** Clear dissemination objectives that match test purposes are fundamental. Agreement must be reached on three issues: when dissemination is carried out, for what purposes, and for what audiences. More than half of testing subcomponents for monitoring address these questions. Subcomponents for selection/certification and evaluation, in contrast, cover this issue only in 23% and 32% of the cases, respectively.

70. The most common target audience is national officials. Eighty-nine percent of monitoring subcomponents that provide details on dissemination view national officials as the main users of test results; 71% of selection/certification subcomponents and 45% of evaluation programs also target national officials for dissemination. Other frequently mentioned targets are regional officials for monitoring subcomponents, teachers and school directors for selection/certification tests, and teachers and local researchers for studies evaluating specific programs. Parents and students are rarely mentioned.

71. **Timeliness.** A common reason for not using test results for policy analysis is that they are not
available on time. Techniques to avoid overloading national computer centers, such as providing better computing equipment or subcontracting analysis to a non-governmental institution, can help increase the speed with which data collection and analysis is carried out. Furthermore, greater efficiency can be reached by using optical scanning equipment. Less than 40% of monitoring, selection/certification, and evaluation subcomponents, however, allude to methods for ensuring timely dissemination of results.

72. **Interpretability.** Another obstacle to appropriate test usage is that results are not presented in easily interpretable ways. To avoid this problem, report formats can be tailored for different audiences. Test-users, such as policy-makers, curriculum developers and school directors, can benefit from clear, non-technical, journalistic and graphically illustrated reports. Furthermore, it is critical that reports aim at preventing test misuse (i.e. warning test-users of applying results for purposes for which test was not intended). In spite of the fact that misinterpretation and ill-use of test results are serious problems that test developers and users confront, less than one-fourth of testing subcomponents address this issue. Only 24% of monitoring, 14% of evaluation, and 5% of selection/certification subcomponents give careful consideration to how to report results for appropriate purposes.

73. **Availability.** Written reports are not always the most effective form of communication. Other means of dissemination (e.g. seminars, briefings, workshops, public media) can help make information more accessible to test-users. Making information easily accessible is particularly important for tests intended to guide national or regional policy deliberations. Nevertheless, only 29% percent of monitoring and 9% of evaluation subcomponents attempt to make information more accessible to policy-makers and other test-users.

74. **Stimulating Demand for Information.** Intended test-users may not be aware of the benefits and uses of test results. Therefore techniques for stimulating demand (e.g. creation of test-user networks,
ministry commissions, and training in policy analysis) should be encouraged. Encouraging test-use, similar to issues of availability, is particularly important for tests for monitoring and evaluation. Test-takers and test-users of selection/certification tests are most likely to perceive the value of test scores; whereas, the value of results from tests for monitoring or evaluation is not immediately apparent to policy-makers. Yet less than 40% of monitoring and evaluation subcomponents attempt to inform policy-makers or other test-users of the importance of test results.

75. **Evaluation of information use.** Responding to test-users needs and usage patterns is key and, in order to ensure responsiveness, mechanisms for evaluating test usage are needed. Evaluations of test usage can inform judgements on what dissemination system improvements are required and, in the view of some experts, reviewing the consequences of decisions or policies made from test-based inferences can complement judgements on test validity. Of all six dissemination issues, evaluating test use is given the least attention in Bank testing subcomponents. Nineteen percent of testing subcomponents for monitoring and 5% of those for selection/certification tests address this issue. None of the testing subcomponents for evaluation mentioned surveying the use of study results.

76. **Summary.** Even if high quality tests are developed, the production and analysis of test results remain worthless if the tests do not achieve their intended purpose. In order to achieve their goal, selection/certification, evaluation and monitoring test scores need to be effectively and efficiently disseminated. Chile’s Primary Education Improvement project recognizes the importance of high quality dissemination systems (Box 4). However, of all three quality issues that have been reviewed, dissemination issues are given the least attention in the majority of Bank supported testing subcomponents.
Box 4. Chile Primary Education Improvement Project: Dissemination Quality

SIMCE is the Chilean Ministry of Education’s system for monitoring trends in the quality of education. The system presently monitors the quality of primary education but, in spite of SIMCE’s sophisticated tests, procedures, and organization, test results are currently underutilized. Therefore, in addition to improving the technical quality of tests and adding two subsystems (one for multigrade, rural schools and one for secondary education), Chile is also strengthening SIMCE’s dissemination system. These efforts are supported by the World Bank through Chile’s FY92 Primary Education Improvement Project. Three percent ($6 million) of project costs are allocated to this testing subcomponent.

All six dissemination quality issues (100%) are addressed in SIMCE’s dissemination plans. Importance is given to having clear objectives, providing timely information, supplying easily interpretable results, using different communication forms, stimulating demand for test results, and having mechanisms for ensuring responsiveness to test-users’ needs and usage patterns.

Clear Dissemination Objectives. SIMCE’s disseminates test results to educational authorities at national (MINEDUC), regional and provincial levels, school directors, teachers and parents. It is expected that SIMCE results will facilitate education, national, regional, and school decision-making processes.

Timeliness. Timely availability of information is an important characteristic of efficient dissemination systems. SIMCE is equipped with optical scanning equipment that helps accelerate data processing which, in turn, increases the possibilities of delivering information on time. The project addresses this issue by asserting that SIMCE will be strengthened to ensure that “information gathering activities are carried out in a timely and efficient manner.” (file: Primary Education Improvement Project, SAR#9769, p. 25).

Interpretability. In previous years, SIMCE has attempted to facilitate test score interpretation by providing guidelines on how to interpret results. The project will continue and strengthen these efforts.

Availability. SIMCE has also attempted to make information more accessible by conducting meetings which have been attended by supervisors, directors and teachers, used audio-visual materials to discuss the meaning of test results. In addition to continuing these meetings, the Ministry of Education, with support from the World Bank, is making SIMCE results more accessible by establishing an Educational Information System (EIS). This EIS will link SIMCE with other ongoing information systems (educational statistics, programming and budgeting, budget disbursements, MINEDUC’s administration and accounting).

Stimulating Demand for Information. The demand for SIMCE results is stimulated by disseminating information about SIMCE to all possible test-users. The reports define SIMCE’s objectives and activities, describe the value of SIMCE results and provide information about test-use.

Evaluation of Information Use. SIMCE recognizes the importance of establishing a system through which test-users can define their needs and usage patterns. By receiving feedback on the type of information and form in which it is distributed, SIMCE can be more responsive to test-users demands.

Sources: Morales Frias (1990); Chile: Primary Education Improvement Project, SAR#9769.
IV: Lessons Learned

77. The present review is designed to describe and analyze the Bank’s involvement in supporting the development of educational testing capacity in member borrower nations. A comprehensive review of education projects and education components in non-education projects, FY63-FY93, found 90 projects with these objectives. This chapter highlights findings regarding Bank lending practices in this area and draws some conclusions about the conditions for success of testing subcomponents.

78. The major findings of this report are as follows:

1. Bank lending for educational testing has risen rapidly since FY90. Both the number of projects and the detail of project description reflect an increased focus on the outputs of education systems; at the same time, design of educational testing components remains often vague in both SARs and their related background documents.

2. Specific budget figures for testing subcomponents are found in only 23 of the 90 projects; there is no discernable trend in the level of financing of testing subcomponents in these projects, and the amount budgeted for testing subcomponents represents less than 2% of overall project costs.

3. A higher proportion of education projects in LAC include testing subcomponents than do any other region, although the total number of education projects with testing subcomponents in LAC is lower than that of Africa or Asia, and only slightly higher than that of EMENA.

4. Support for tests to monitor achievement trends is increasing, while support for tests for selection/certification is declining (Figure 9). Although there is an evolution towards supporting tests for monitoring purposes, more projects support tests for selection and/or certification than
for any other purpose.

5. Support for subcomponents related to changing an existing testing system is more frequent than support for creating a new system, and both types of subcomponents are more frequently found in projects than subcomponents designed to prepare for the establishment of a new system.

6. The quality of testing subcomponents is improving, but remains low. Technical quality is improving more rapidly than institutional or dissemination quality.

79. The recency of testing subcomponents in Bank lending operations makes drawing conclusions about the conditions under which these subcomponents succeed rather difficult. Nevertheless, five lessons emerge:

1. First, projects that focus on improving institutional quality before addressing either technical quality issues or dissemination issues are more likely to succeed than projects that initially try to improve technical or dissemination quality. The principal reason for this is that without a strong institutional base, technical expertise supported under the project will be lost at its conclusion. Strategies in Bank-supported projects for improving technical or dissemination quality often rely on foreign and local technical assistance and training for local staff on technical matters. When foreign or local specialists complete their tour of duty, their expertise leaves the country or the sector unless a strong institution has been established to maintain a critical mass of technical expertise. Similarly, when local staff are trained but local testing institutions are not available to employ their expertise, they migrate to positions outside the sector that can better employ their services. If, on the other hand, the project concentrates on establishing a strong institutional base for technical specialists, then these skills will not be lost to the sector. Nevertheless, institutional
strengthening is not an easy task and its success does not necessarily ensure high technical and dissemination quality. Institution building, technical quality and dissemination are all embedded in political and evaluation cultures that condition the interactions among them.

2. Second, for projects that focus on improving technical quality, success will be higher when both planning and implementation issues are addressed. In many such projects, only planning receives significant attention, and implementation issues, such as certain aspects of test development, test administration, data processing and analysis are rarely elaborated in detail. Without successful implementation of these tasks, a testing system will not be operable. The failure of testing subcomponents in early projects has been attributed to failure to attend to these details.

3. Third, for test scores to have meaning over time, student performance must be measured against some inelastic yardstick of achievement, such as a norm reference group or criteria references (Anastasi 1988; Berk 1984). Yet standard setting — the process of developing an inelastic performance standard — is overlooked in virtually all projects. Since standard setting is typically overlooked, most tests that are developed in Bank-supported projects are highly elastic; their meaning changes from year to year. This elasticity constitutes a factor that contributes to the overall inutility — and hence lack of success — of the testing subcomponent, especially of those supporting tests for monitoring purposes.

4. Fourth, for testing to help improve instruction and to provide information for accountability, dissemination of results is important. Yet few testing subcomponents support the development of efficient dissemination systems, including the dissemination of information from the system
to policy makers. This is likely to lead to declining support for the testing system and its ultimate failure.

5. Fifth, testing subcomponents that are designated as identifiable project components or subcomponents in SARs, with specific budget figures attached, are more likely to succeed than projects that pay little attention to the testing subcomponent. These testing subcomponents are more likely to be fully elaborated, with separate background documents, than testing subcomponents lacking specific budgets, and are therefore more likely to receive supervision. They also are more fully thought through, with attention paid to implementation of specific "quality" issues.
<table>
<thead>
<tr>
<th>COUNTRY</th>
<th>PROJECT</th>
<th>FY</th>
<th>LOAN</th>
<th>M</th>
<th>E</th>
<th>A</th>
<th>S/C</th>
<th>D</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALGERIA</td>
<td>Basic Education Institutional Development</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARGENTINA</td>
<td>Social Sector Management Technical Assistance Project</td>
<td>89</td>
<td>L2984</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BANGLADESH</td>
<td>Fourth Education Project</td>
<td>80</td>
<td>C1054</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BANGLADESH</td>
<td>Technical Education Project</td>
<td>84</td>
<td>C14900</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BANGLADESH</td>
<td>Second Primary Education Project</td>
<td>85</td>
<td>C15740</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BANGLADESH</td>
<td>General Education Project</td>
<td>90</td>
<td>C2118</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BELIZE</td>
<td>Education and Training</td>
<td>92</td>
<td>L3422</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BHUTAN</td>
<td>Primary Education Project</td>
<td>88</td>
<td>C18990</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOLIVIA</td>
<td>First Education and Vocational Training Project</td>
<td>77</td>
<td>LN1404</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOTSWANA</td>
<td>Fourth Education Project</td>
<td>86</td>
<td>L26440</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRAZIL</td>
<td>Northeast Basic Education I</td>
<td>80</td>
<td>L1867</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRAZIL</td>
<td>Innovations in Basic Education</td>
<td>91</td>
<td>L3375</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRAZIL</td>
<td>Second Northeast Basic Education</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BURKINA FASO</td>
<td>Fourth Education Project</td>
<td>91</td>
<td>C2244</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAPE VERDE</td>
<td>Primary Education Upgrading</td>
<td>88</td>
<td>C18530</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CENTRAL AFRICAN REP.</td>
<td>Second Education Project</td>
<td>83</td>
<td>C13590</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHILE</td>
<td>Education IV</td>
<td>92</td>
<td>L3410</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 M=MONITOR, E=EVALUATE, A=ACCOUNTABLE, S/C=SELECT/CERTIFY, D=DIAGNOSE, C=COMBINATION

2 All FY 93 projects are in proposal stage.
<table>
<thead>
<tr>
<th>COUNTRY</th>
<th>PROJECT</th>
<th>FY</th>
<th>LOAN</th>
<th>M</th>
<th>E</th>
<th>A</th>
<th>S/C</th>
<th>D</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHINA</td>
<td>Provincial Education</td>
<td>92</td>
<td>C2339</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>COLOMBIA</td>
<td>Subsector Project for Rural Basic Education</td>
<td>83</td>
<td>L21920</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>COLOMBIA</td>
<td>Second Subsector for Primary Education Project</td>
<td>89</td>
<td>L30100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>COSTA RICA</td>
<td>Basic Education Rehabilitation Program</td>
<td>92</td>
<td>L3414</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DJIBOUTI</td>
<td>A First Education Project</td>
<td>85</td>
<td>C15430</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DOMINICAN REPUBLIC</td>
<td>Vocational Training Project</td>
<td>85</td>
<td>L2558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DOMINICAN REPUBLIC</td>
<td>Primary Education Development</td>
<td>91</td>
<td>L3351</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>ECUADOR</td>
<td>First Social Development Project</td>
<td>92</td>
<td>L3425</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETHIOPIA</td>
<td>Fifth Education Project</td>
<td>81</td>
<td>C1141</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAMBIA</td>
<td>Education Sector Project</td>
<td>90</td>
<td>C2142</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHANA</td>
<td>II Education Sector Adjustment</td>
<td>90</td>
<td>C2140</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAITI</td>
<td>Education III - Basic Education Project</td>
<td>83</td>
<td>C13050</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAITI</td>
<td>Fifth Education Project</td>
<td>91</td>
<td>C2264</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INDIA</td>
<td>Vocational Training Project</td>
<td>89</td>
<td>C20080</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INDIA</td>
<td>Second Technician Education Project</td>
<td>91</td>
<td>C2223</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INDONESIA</td>
<td>Secondary Education and Management</td>
<td>85</td>
<td>L24720</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INDONESIA</td>
<td>II Secondary Education and Management</td>
<td>90</td>
<td>L3158</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INDONESIA</td>
<td>Primary Education Quality Improvement</td>
<td>92</td>
<td>L3448</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JAMAICA</td>
<td>Social Sectors Development Project</td>
<td>89</td>
<td>L3111</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JAMAICA</td>
<td>Educational Social Sector II</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>JORDAN</td>
<td>Sixth Education Project</td>
<td>84</td>
<td>L23780</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JORDAN</td>
<td>Seventh Education Project</td>
<td>88</td>
<td>L28900</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COUNTRY</td>
<td>PROJECT</td>
<td>FY</td>
<td>LOAN</td>
<td>M</td>
<td>E</td>
<td>A</td>
<td>S/C</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------------------------------------------------</td>
<td>----</td>
<td>--------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-----</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>JORDAN</td>
<td>Human Resources Development Sector Investment Loan</td>
<td>89</td>
<td>L3106</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KENYA</td>
<td>Education III</td>
<td>76</td>
<td>L11840</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KENYA</td>
<td>Education IV</td>
<td>78</td>
<td>C07970</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KOREA</td>
<td>Program for Science and Technology Education</td>
<td>84</td>
<td>L2427</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESOTHO</td>
<td>Education Sector Development</td>
<td>92</td>
<td>C2287</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIBERIA</td>
<td>Education I</td>
<td>76</td>
<td>L12660</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>LIBERIA</td>
<td>Fourth Education Project</td>
<td>83</td>
<td>C13060</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MADAGASCAR</td>
<td>Education Sector Reinforcement Project</td>
<td>90</td>
<td>C2094</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALAWI</td>
<td>Education Sector Credit</td>
<td>87</td>
<td>C17670</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALAWI</td>
<td>Second Education Sector Credit</td>
<td>90</td>
<td>C2083</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALAYSIA</td>
<td>Primary and Secondary Education</td>
<td>86</td>
<td>L26850</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALAYSIA</td>
<td>III Primary &amp; Secondary Education Sector</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALI</td>
<td>Education Sector Consolidation</td>
<td>89</td>
<td>C20540</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEXICO</td>
<td>Primary Education Project</td>
<td>92</td>
<td>L3407</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOROCCO</td>
<td>Education Sector Reform Program</td>
<td>86</td>
<td>L26640</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOROCCO</td>
<td>Second Vocational Training Project</td>
<td>87</td>
<td>L2779</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOROCCO</td>
<td>Rural Primary Education Project</td>
<td>89</td>
<td>L30260</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOROCCO</td>
<td>Basic Education Development Project</td>
<td>91</td>
<td>L3295</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOZAMBIQUE</td>
<td>Second Education Project</td>
<td>91</td>
<td>C22000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEPAL</td>
<td>Second Education Project</td>
<td>82</td>
<td>L1198</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEPAL</td>
<td>Basic and Primary Education Project</td>
<td>92</td>
<td>C2357</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIGERIA</td>
<td>Primary Education Project</td>
<td>91</td>
<td>C2191</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COUNTRY</td>
<td>PROJECT</td>
<td>FY</td>
<td>LOAN</td>
<td>M</td>
<td>E</td>
<td>A</td>
<td>S/C</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------------------------------------------------------</td>
<td>----</td>
<td>-------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-----</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>OMAN</td>
<td>Third Education Project</td>
<td>87</td>
<td>L27990</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAKISTAN</td>
<td>Primary Education Project</td>
<td>79</td>
<td>C892</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAKISTAN</td>
<td>Fifth Education (Vocational Training) Project</td>
<td>81</td>
<td>C1109</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAKISTAN</td>
<td>Sindh Primary Education Development Program</td>
<td>90</td>
<td>C2102</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAKISTAN</td>
<td>Punjab Middle Schooling Project</td>
<td>92</td>
<td>C2354</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAPUA NEW GUINEA</td>
<td>Primary Education Project</td>
<td>81</td>
<td>C1087</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAPUA NEW GUINEA</td>
<td>Secondary Education Project</td>
<td>84</td>
<td>L23950</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHILIPPINES</td>
<td>Third Education Project</td>
<td>76</td>
<td>L1224</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHILIPPINES</td>
<td>Fourth Education Project</td>
<td>77</td>
<td>L1374</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHILIPPINES</td>
<td>Educational Radio Technical Assistance Project</td>
<td>78</td>
<td>S008</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHILIPPINES</td>
<td>Sector Program for Elementary Education</td>
<td>81</td>
<td>L2030</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHILIPPINES</td>
<td>Vocational Training Project</td>
<td>83</td>
<td>L2200</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHILIPPINES</td>
<td>Second Elementary Education</td>
<td>91</td>
<td>L3244</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RWANDA</td>
<td>Education I</td>
<td>75</td>
<td>C05670</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RWANDA</td>
<td>First Education Sector Project</td>
<td>91</td>
<td>C2227</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SENEGAL</td>
<td>Third Education Project</td>
<td>79</td>
<td>C908</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SENEGAL</td>
<td>Primary Education Development Project</td>
<td>87</td>
<td>C17350</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOLOMON ISLANDS</td>
<td>Secondary Education Project</td>
<td>86</td>
<td>C16860</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOMALIA</td>
<td>Third Education Project</td>
<td>78</td>
<td>CR738</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOMALIA</td>
<td>Education Rehabilitation Program</td>
<td>90</td>
<td>C21110</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TANZANIA</td>
<td>Education VII</td>
<td>81</td>
<td>C10560</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRINIDAD and TOBAGO</td>
<td>Education and Training for Youth Employment</td>
<td>91</td>
<td>L3328</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COUNTRY</td>
<td>PROJECT</td>
<td>FY</td>
<td>LOAN</td>
<td>M</td>
<td>E</td>
<td>A</td>
<td>S/C</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------------------------------</td>
<td>----</td>
<td>---------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-----</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>TUNISIA</td>
<td>Education and Training Sector Loan</td>
<td>89</td>
<td>L30540</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TURKEY</td>
<td>Non-formal Vocational Training Project</td>
<td>87</td>
<td>L2776</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TURKEY</td>
<td>National Education Development</td>
<td>90</td>
<td>L3192</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UGANDA</td>
<td>Fifth Education Project</td>
<td>93</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>YEMEN ARAB REPUBLIC</td>
<td>Technical Training Project</td>
<td>86</td>
<td>L16450</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZAIRE</td>
<td>Education Sector Rehabilitation</td>
<td>91</td>
<td>C2213</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZAMBIA</td>
<td>Fourth Education Project</td>
<td>77</td>
<td>L1356</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ANNEX 2: Coding Form

SAR Report No.:  
Project Name:  
Component:  
Subcomponent:  
Paragraphs, pages, Annex in SAR:

GENERAL INFORMATION

Form: Is this the first or only form for this project?  
Yes or No

Country:  
Region:  
1. AFRICA  
2. ASIA  
3. EMENA  
4. LAC

Fiscal Year:  
Stage. The project is:  
1. a project in preparation stage (not yet approved)  
2. an ongoing project  
3. a closed project  
4. other ____________

Comp. How does the support for testing fit into the project?  
1. A component of it’s own  
2. A subcomponent of it’s own  
3. Part of a subcomponent.  
4. Other ____________
WHAT MEASURES OF STUDENT LEARNING ARE SUPPORTED

Purposes of Measures of Student Learning Supported in the Project:

Y or N (M) Monitoring Progress Toward National Educational Goals
Y or N (E) Evaluating Effectiveness of Specific Policies
Y or N (A) Holding Schools Accountable for Performance
Y or N (S) Selection and/or Certifying Students
Y or N (T) Teacher Assessment of Individual Student Performance
Y or N (X) Mixed Purposes

If the test is mixed, then which purposes are combined?

Y or N x1. Monitoring Progress Toward National Educational Goals
Y or N x2. Evaluating Effectiveness of Specific Policies
Y or N x3. Holding Schools Accountable for Performance
Y or N x4. Selection and/or Certifying Students
Y or N x5. Teacher Assessment of Individual Student Performance

Purpose ID #: Remaining questions specific to testing activity will apply to projects support of tests with the following purpose (only one):

(1) Monitoring Progress Toward National Educational Goals
(2) Evaluating Effectiveness of Specific Policies
(3) Holding Schools Accountable for Performance
(4) Selection and/or Certifying Students
(5) Teacher Assessment of Individual Student Performance
(6) Mixed Purposes

PREVIOUS EXPERIENCE WITH MEASURES OF STUDENT LEARNING

Q11. With what type of testing does the country seem to have had experience?

Y or N (a) Monitoring Progress Toward National Educational Goals
Y or N (b) Evaluating Effectiveness of Specific Policies
Y or N (c) Holding Schools Accountable for Performance
Y or N (d) Selection and/or Certifying Students
Y or N (e) Teacher Assessment of Individual Student Performance
Y or N (f) None of the above
Y or N (g) Hard to know from provided information
PROJECT COST INFORMATION: (in US$ MILLIONS)

Q1. Total Project Cost: US$ ____________

Q2. Disaggregation of Project Costs by Sources of $

<table>
<thead>
<tr>
<th>Source</th>
<th>1. US$ millions</th>
<th>2. % of total project costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. IBRD/IDA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Government</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Other 1.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total Project Cost 100%</td>
</tr>
</tbody>
</table>

Q3. What is the project's total base cost and total contingencies?

<table>
<thead>
<tr>
<th>Category</th>
<th>1. US$ millions</th>
<th>2. % of total project costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Total Base Cost</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Contingencies</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total Project Cost 100%</td>
</tr>
</tbody>
</table>

Q4. What lending instrument is used for the education project?

1. Specific Investment Loans (SILs)
2. Sector Investment Loans (SECILs)
3. Sector Adjustment Loans (SECALs)
4. Hybrids

Comments:
LEVEL OF SUPPORT FOR MEASURES OF STUDENT LEARNING (in US$ millions)

Q5. What is the most disaggregated amount provided for the level of $ support?

0. Total project costs
1. Total costs for Component including other elements than testing activities.
2. Total costs for the Subcomponent including other elements than testing activities.
3. Total costs for more than one testing activities.
4. Total costs for one testing activities.

***** If (3) or (4) was answered in Q5, then please answer Q6, and Q7 *****

Q6. How much $ (in millions) is provided for measures of student learning?

   1. US$ millions   2. % of total project costs
   a. Measures of Student Learning
   b. Other

   Total Project Cost 100%

Q7. For how many measures of student learning does the amount provided in Q6 refer?
   1  2  3  4  5  6

*****   *****   *****

Q8. Is another agency, bank or country (other than the local government and IDA/IBRD) providing part of this $ for supporting measures of student learning?

Yes or No

Comments:
**TYPE OF SUPPORT**

Q9. What type of support is the Bank providing?

Y or N a. Recurrent Costs

Y or N 1. Not specified
Y or N 2. Technical Assistance (Human Resources Investment)
    Y or N a. Not specified
    Y or N b. Consultants and/or Specialists
    1. If Yes, then does the project specify whether the type of consultants (i.e. short-term or long-term and local or foreign)?
       Y or N If yes, consultants and/or specialists will be:

Y or N c. Other

Y or N 3. Staff Development
    Y or N a. Local Training
    Y or N b. Training Abroad
    Y or N c. Fellowships
    Y or N d. Study tours
    Y or N e. Other

Y or N 4. Research and Studies
Y or N 5. Piloting program/piloting of tests
Y or N 6. Printing Costs
Y or N 7. Administration Costs
Y or N 8. Other

Y or N b. Capital Costs

Y or N 1. Not specified
Y or N 2. Computer Equipment
Y or N 3. Other Office Equipment
Y or N 4. Vehicles
Y or N 5. Furniture
Y or N 6. Civil Works
Y or N 7. Other

Comments:
**AVAILABLE DESIGN/PLAN INFORMATION:**

Q10. How much information is provided in SAR?
   0. SAR not yet available (project under preparation)
   1. a little (less than a paragraph)
   2. some (a paragraph)
   3. substantial (more than one paragraph but less than a page)
   4. exceptional (a page or more!)

Q11. Is there a background document, consultant work or other document which addresses the testing subcomponent? If yes (1-3), how much information does it provide?
   0. None available
   1. Document with little information (less than a page)
   2. Document with some information (more than a page, but document not specific to testing subcomponent)
   3. Document with exceptional amount of information (document is specific to testing subcomponent)
   9. Document not ready yet, but will be prepared.

Comments:

**SPECIAL AGREEMENTS DURING NEGOTIATIONS**

Q12. Were any special agreements concerning the measures of student learning subcomponents made during negotiations?

   YES  or  NO

If yes, what was the agreement(s)?
Q13. For what educational cycle(s) are the test(s)?:

Y or N a. Basic Education (Primary, Elementary, or mixed—e.g. grades 2, 6, 9)
Y or N b. Secondary Education
Y or N c. Vocational Education
Y or N d. Higher Education
Y or N e. Other ________________________

Q14. Are the objectives for the testing clearly stated?

Yes or No

Q15. The project attempts to:

Y or N a. Strengthen/Change an existing testing system.
   What kind of changes are being made?
   Y or N 1. Test quality improvements:
            e.g. A periodic update: adding a new section or removing dated items.
            Ongoing development: testing new items by imbedding them in the current test or by rolling norming.
            Major overhaul: An existing test is totally revamped, reviewing at least the most fundamental framework and parameters.
   Y or N 2. Test expansion and/or addition of tests
            e.g. Expanding coverage and/or modifying test to cover specific areas (urban/rural) or grade levels.
            e.g. Adding a new test for different educational level
   Y or N 3. Institutional improvements
   Y or N 4. Dissemination improvement
            e.g. Improving test usage.
   Y or N b. Create a new system.
   Y or N c. Carry out a feasibility study for creating a new system.
   Y or N d. Other ________________________
            e.g. Study; impact evaluation.

Comments:
ORGANIZATIONAL CONTROL (institutional arrangements)

Q16. Where are the testing systems going to be housed?

0. Not available
1. Private or semi-autonomous Institution
2. Government (MOE)
3. Government will subcontract out to a semi-autonomous or private institution.
4. Other

Q17. Is the same institution managing more than one testing systems (e.g. selection/certification and national assessment)?

0. No
1. Yes
9. Don’t Know

Q18. For those projects planning to house the systems in either (2) or (3) (of Q16) —that is, some control in Government —, what strategy for institutionalizing the assessment system will be used?

0. Unclear
1. Single unit: A single unit is responsible for the development and/or management of the system.
2. Collaboration: One government unit coordinates collaborative efforts of various units and/or divisions within ministries of education or other government units.
3. Other

Q19. Are any of the participating units, divisions and/or institutions new ones created in the project?

YES or NO

If yes, what is the new unit, division and/or institution?

Comments:

Project ID:
ARE THE RIGHT ISSUES ADDRESSED? (Just numbered headings need to be circled. Non-numbered sub-elements are to give an idea of headings)

Q20. Technical Quality of Tests and Background Surveys

Y or N A. Test Development

Y or N 1. Conceptual Planning
- Defining Uses and Objectives
- Designing Procedures
- Selecting Development Team
- Documenting
- Disseminating

Y or N 2. Setting Test Parameters
- Deciding on Levels, forms, item type
- Deciding on types of skills
- Deciding on samples, uses of reports

Y or N 3. Writing the Test
- Selecting objectives
- Selecting grades and subjects
- Identifying time and item frameworks for each objective
- Writing preliminary items
- Reviewing preliminary items

Y or N 4. Graphic & Printing
- Artist work
- Typing
- Graphic lay-out
- Editing/proofing

Y or N 5. Pretesting Items
- Printing versions
- Selecting pre-test groups
- Distributing tests
- Training testers
- Securing tests
- Administering the test
- Collecting completed tests
- Inputting tests into the computer
- Data editing and clean-up
- Item Analysis
- Reliability and Validity Check
Y or N 6. Selecting Final Items
   - Group meetings
   - Estimating final test times
   - Documenting

Y or N 7. Final lay-out
   - Graphics layout
   - Editing/proofing

Y or N 8. Norming
   - Selecting norming groups
   - Norming
   - Reliability and validity estimation
   - Documenting

Y or N B. Survey Development-- (When relevant, e.g. Monitoring or Evaluating Purposes.)

Y or N 1. Conceptual Planning
   - Defining Uses and Objectives
   - Designing Procedures
   - Selecting Development Team
   - Documenting
   - Disseminating

Y or N 2. Writing the Survey
   - Selecting objectives
   - Selecting areas of interest
   - Writing preliminary questions
   - Reviewing preliminary questions

Y or N 3. Graphic & Printing
   - Artist work
   - Typing
   - Graphic lay-out
   - Editing/proofing

Y or N 4. Pretesting Questions
   - Printing versions
   - Distributing surveys
   - Training surveyors
   - Securing surveys
   - Administering survey
   - Collecting completed surveys
   - Inputting answers into the computer
   - Data editing and clean-up
Y or N 5. Selecting Final Questions
   - Group meetings
   - Deciding on final items
   - Document

Y or N 6. Survey Production
   - Graphics layout
   - Editing/proofing

Y or N C. Preparation, Administration and Data Processing

Y or N 1. Test Production
   - Printing
   - Securing
   - Numbering and Counting
   - Packaging
   - Distributing

Y or N 2. Test Manual Development
   - Selecting team
   - Meetings with test developers
   - Reviewing similar test booklets
   - Writing
   - Editing
   - Copying

Y or N 3. Test Manual Revision
   - Editing
   - Printing

Y or N 4. Preparation
   - Selecting sample population
   - Editing and proofreading
   - Registration
   - Training

Y or N 5. Informing regions and schools of tests
   - Writing text
   - Dissemination Information

Y or N 6. Administration
   - Scheduling
   - Administering
   - Collecting
Y or N  7. **Processing**
   - Counting
   - Marking
   - Batching

Y or N  8. **Machine Scoring**
   - Computer input
   - Cleaning data
   - Scoring

Y or N  9. **Hand Scoring** (especially for Teacher Assessment)
   - Scoring
   - Aggregating
   - Recording

Y or N  D. **Analysis**

Y or N  1. **Aggregation/Trends**
   - Aggregation by Hand
   - Aggregation by Machine
   - Trend Analysis

Y or N  2. **Policy Analysis or Pedagogical Analysis**
   - Research integrating background surveys
Q21. Institutional Requirements for Research and Analysis (Remember to mark numbered subheadings also!)

Y or N A. Institutional Stability
   Y or N  1. Clear objectives
   Y or N  2. Long-term financial support
   Y or N  3. Stable staff

Y or N B. Staff quality
   Y or N  1. University-based training

Y or N C. Institution-based training
   Y or N  1. Post doctoral
   Y or N  2. "Extension workers"

Y or N D. Communication
   Y or N  1. Professional (peer)

Comments:
Q22. Dissemination Quality

Y or N  A. Clarity of Objectives of Dissemination System
- what audience?
- when?
- for what?

Y or N  B. Timeliness
- computers/optical scanners
- contract out (thus, not overloading government units, computers)

Y or N  C. Interpretability
- form of report: how to make stats easy to understand.
- attention to different forms for different audiences.

Y or N  D. Availability - Methods of Dissemination
- seminars
- briefing
- news

Y or N  E. Stimulate demand for information
- networks
- ministry commissions
- training in policy analysis

Y or N  F. Evaluation of information use
- research, surveys on information use for improving dissemination system.

Comments:
SUBCOMPONENT "SUCCESS" INFORMATION:

If this is a closed project (e.g. stage = 3), then answer following section.

Q23. Is a PCR or PPAR available for the project and to what extent does it address subcomponents including measures of student learning?
   0. No PCR or PPAR is available
   1. PCR or PPAR has no information on testing
   2. a little (less than a paragraph)
   3. some (a paragraph)
   4. substantial (more than one paragraph but less than a page)
   5. exceptional (a page or more!)

Q24. If information was available in the PCR and/or PPAR (i.e. answers 2-5 on Q23), then was the subcomponent carried out and was it successful?
   0. No
   1. Yes, but not "successful"
   2. Yes and "successful"

What exactly does the PCR or PPAR tell us? (Describe)
Q25. DETAILS ON DESIGN

Circle one answer for A-C:

A. To whom administered:
   0. Not available
   1. Sample of students
   2. Grade census
   3. Universe of students

Comments: (e.g. initially sample but moving towards grade census?)

B. When administered:
   0. Not available
   1. Periodically
   2. One time

C. Content objective:
   0. Not available
   1. Selected domains of curriculum
   2. All domains of curriculum

D. Behavioral objective
   Y or N 1. Not available
   Y or N 2. Knowledge
   Y or N 3. Higher order thinking skills
   Y or N 4. Practical skills

E. Format
   Y or N 1. Not available
   Y or N 2. Objective
   Y or N 3. Performance

F. Standardized
   Y or N 1. Not available
   Y or N 2. Not specified but mentions to be "standardized"
   Y or N 3. Content
   Y or N 4. Administration
   Y or N 5. Scoring

68
G. Reference
Y or N  1. Not available
Y or N  2. Norming Group
Y or N  3. Performance Criteria

H. Supplementary measures
Y or N  1. Not available.
Y or N  2. Not specified (but reference to background surveys)
Y or N  3. Student background
Y or N  4. Classroom/school inputs
Y or N  5. Classroom/school processes
Y or N  6. Student attitudes
Y or N  7. Teacher attitudes
Y or N  8. Other __________________________

Comments: (e.g. Background information is collected but not simultaneously with test administration.)

Other interesting details on design? What are they?
Q26. DETAILS ON DISSEMINATION

A. Is Dissemination information available?

0. No
1. Somewhat (vague, implied)
2. Yes, clear.

B. If yes or somewhat, to whom will info be disseminated:

Y or N 1. Students
Y or N 2. Teachers
Y or N 3. Parents
Y or N 4. Principals (directors)
Y or N 5. Regional officials
Y or N 6. National officials (MOE central)
Y or N 7. Researchers (local)
Y or N 8. Researchers (international)
Y or N 9. Public media
Y or N 10. Curriculum developers
Y or N 11. Textbook writers
Y or N 12. Teacher Training institutions
Y or N 13. Others
ANNEX 3: Glossary of Basic Testing Terminology for World Bank Projects

A test score is an estimate based on sampling what the test taker knows or can do. Both achievement tests and aptitude tests estimate aspects of an individual’s developed abilities. Aptitude tests, however, are intended to capture knowledge acquired both in and out of school, whereas achievement tests are more linked to what a student knows and can do in a specific subject only as a result of schooling.

Standardized tests are tests that are administered and scored under conditions uniform to all students. Standardization is needed to make test scores comparable and to assure as much as possible that test takers have equal chances to demonstrate what they know. Most people associate standardized tests with the multiple-choice format, yet standardization is a generic concept that can apply to any testing format—from written essays to oral examinations to producing a portfolio.

The World Bank has supported standardized achievement tests through education projects. Since the majority of Bank education projects support educational testing through project subcomponents (rather than free-standing components), all project elements related to testing are referred to as testing subcomponents. The purpose of the subcomponents are to strengthen an existing testing system, build a new testing system, or prepare for a new system. Some subcomponents supporting tests for evaluation purposes only aimed at completing a one-shot study (i.e. they were not concerned with changing, creating or strengthening testing systems.) The subcomponents supported tests for six purposes: (a) selecting students for further education; (b) certifying student achievement; (c) monitoring achievement trends over time; (d) evaluating specific educational programs or policies; (e) holding schools, regions, etc. accountable for student achievement; and/or (f) diagnosing individual learning needs.

Tests are imprecise because they are based on samples of behavior and scores can vary for

---

reasons unrelated to the individual’s actual achievement. Test scores can only evaluate what skills have been mastered, but they cannot alone, explain why learning has occurred, or prescribe ways to improve it. Achievement is affected by schools, parents, home background, and other factors which need to be taken into consideration when drawing inferences about schools, programs and policies. In short, test scores must be interpreted carefully. To interpret tests cautiously, it is important to understand important test characteristics such as standards, reliability, validity, equating, and methods of interpretation.

When applied to tests, the word *standards* has at least two meanings. First, in the more general sense, it denotes goals, desirable behaviors, or models to which students, teachers, or schools should aspire. That is, standards describe optimal performance and what is desirable for students to know. The word, in its more technical definition, denotes the specific levels of proficiency that students are expected to attain. Thus, setting a standard of performance on a test is setting a passing score for that test.

*Reliability* refers to the consistency and generalizability of test data. It describes the extent to which a test is consistent in measuring whatever it does measure or to the degree to which repeated measurement of the same individual would tend to produce the same result. A test needs to demonstrate a high degree of reliability before it is used to make decisions especially those made from high-stakes tests such as selection exams.

*Validity* refers to whether or not a test measures what it is supposed to measure, and whether appropriate inferences can be drawn from test results. There are different kinds of validity depending on the type of test and its purposes and different types of evidence are appropriate for each type of validity. In the views of some experts, even the consequences of translating test-based inferences into decisions or policies is suitable evidence for judging validity. An acceptable level of validity must be demonstrated before a test is used to make decisions.

*Equating* tests allows test results from different tests (or forms of tests) to be comparable. It is the process of adjusting the raw statistics obtained from a particular sample to corresponding statistics.
estimated for a base group or reference population. There are different methods for equating tests but, in general, equating involves comparing examinees' performance on old and new test questions. Equating is necessary if different tests are used to monitor trends in achievement.

There are two ways of interpreting test results. The first method is to *describe a student's test performance as it compares to that of other students* (e.g. she typed better than 90 percent of her classmates). *Norm-referenced tests* are designed to make these comparisons. The second method is to describe the skills and performance that the student demonstrates (e.g. she typed 75 words per minute without mistakes). *Criterion-referenced tests* are designed for these interpretations.
Annex 4: Sample Flowchart of Test Development Process

The flowchart presented in the following pages is a simplified abstraction of a highly complex and diverse process (Hunter 1979). Hunter's flowchart models the process generally followed when developing tests at ETS. Most test development efforts follow a significant portion of the process detailed in this annex; however, few follow it precisely.
KEY TO FLOWCHART SYMBOLS

**Process Step** - Upper portion identifies person, function, or area responsible for process step. Lower portion contains a brief description of the activity being performed.

**Document** - Source document or product such as planning memos, schedules, test books, answer sheets, etc.

**Data Card** - Used to indicate standard ETS item cards.

**Off Page Connector** - Provide linkage for processing flow between two or more pages. Code consists of a page number (12 in the example) and a unique letter that will allow one to differentiate between several connectors on one page. The number in parentheses to the right ([6] in our example), indicates the page from which the flow originated (where appropriate).

**On Page Connector** - Similar to the off page connector, this symbol indicates flow continuing elsewhere on the same page. A unique letter ("A" in our example) is used to differentiate between multiple connectors.

**Document File** - Office files of cards, memos, etc.

**Terminal Step** - Used to initiate or terminate process flow.

**Decision Diamond** - Indicates process flow branching to two or three paths.

**Broken Line** - Indicates optional path, process step, or document.
TEST DEVELOPMENT PROCESS
TEST COMMITTEE SELECTION

PROG. ADM.  
TD STAFF  
DETERMINE NEED FOR COMMITTEE  

COMM. NEEDED?  

NO  
END COMMITTEE SELECTION  

YES  
CLIENT  
PROG. ADM.  
TD STAFF  
APPOINT AND CONFIRM TEST COMMITTEE  

PROF. ASSNS.  
PRODUCE LISTS OF PROSPECTIVE COMMITTEE MEMBERS  

PROSPECTIVE COMMITTEE MEMBERS  

LIST OF COMMITTEE MEMBERS  

TD STAFF  
DISTRIBUTE BACKGROUND MATERIALS  

BACKGROUND MATERIALS  

TO COMMITTEE
TEST DEVELOPMENT PROCESS
TEST SPECIFICATIONS

EXISTING SPECS

YES
COURSE TEST SPECS

NO

COMMITTEE

TD STAFF
REVIEW
CURRENT
TEST SPECS

RESEARCH

TD STAFF
ANALYZE JOB,
CURRICULUM,
ETC.

COMMITTEE

TD STAFF
DEVELOP AND
REFINE TEST
SPECS

TD STAFF
DOCUMENT
TEST
SPECS

TEST SPECS

PLANO REVIEWER

TEST SPECIALIST

STAT. ANAL.

CLIENT

PROG. ADM.

COMMITTEE

TD STAFF
REFERENCE
THROUGHOUT TD
PROCESS
TEST DEVELOPMENT PROCESS

TEST PRODUCTION
TEST DEVELOPMENT PROCESS
TEST ADMINISTRATION PREPARATION

PRETEST OFFICE
COORDINATE
PRETEST AS
APPROPRIATE

STAT. ANAL.
PREPARE AND
DIST. STAT.
PROCEDURE

TO
STAFF

DATA ENTRY SYS.
PREPARE
ANSWER SHEET
DESIGN

PROG. ADM.
PREPARE
SUPERVISORS
MANUAL DRAFT

DRAFT
SUPERVISORS
MANUAL

TEST
BOOKS

STAT. PROCEDURE

TD COORD.
REVIEW AND
CONSULT

PUBLICATIONS
COORDINATE
ANSWER SHEET
PRODUCTION

PUBLICATIONS
COORDINATE
SUPERVISORS
MANUAL PROD.

SUPERVISORS
MANUAL

TEST
SHIPMENTS

TD COORD.
REVIEW AND
CONSULT AS
NECESSARY

OPERATIONS
SHIP TEST
MATERIAL
TO CENTERS

13A (12)

TO TEST
CENTERS

13A (12)
TEST DEVELOPMENT PROCESS
TEST CENTER RECEIPT PROCESSING

TEST CENTERS

IRREG. REPORTS
SUPERVISORS REPORT FORMS

SUPERVISOR REL.
DISTRIBUTE TO TD AS APPROPRIATE

IRREGULARITY REPORTS
SUPERVISORS REPORT FORMS

TD STAFF REVIEW AND EVALUATE

PROBLEMS?
YES

TD STAFF RESOLVE PROBLEMS
SUPERVISOR AND IRREG REPORTS

TD STAFF ADVISE STAFF AND FILE REPORT

ANSWER SHEETS

15A 16A

TEST BOOKS

OPERATIONS RETAIN IN STORAGE

PROBLEM RESOLUTION

PRETEST?
YES

PROG. ADM.
STAT. ANAL.

TD STAFF RESOLVE PROBLEMS

NO

TD STAFF TRANSMIT PROBLEM SOLUTIONS

VARIOUS MEMOS

PROBLEM IDENTIFICATION NOTICE

TD STAFF
CORRECT ITEM CARD(S)

89
TEST DEVELOPMENT PROCESS

RETIREFG OLD FORMS

ITEM FILES
PREPARE LIST OF ACTIVE TEST FORMS

LIST OF ACTIVE TEST FORMS

STAT. ANAL.

TD STAFF CONSULT AND ADVISE

PROG. ADM.
REVIEW TEST FORMS FOR USEFULNESS

USEFUL FORM?

NO

PROG. ADM.
NOTIFY STAFF TO DISCONTINUE ACTIVE USE

YES

PROG. ADM.
CONTINUE TO USE AS ACTIVE FORM
References


