Using Examinations and Testing to Improve Educational Quality

STEPHEN P. HEYNEMAN and ANGELA W. RANSOM

ACADEMIC achievement affects the eventual economic benefits of education. In industrialized countries, family socioeconomic background and other factors external to the school heavily influence the level of academic achievement. However, in low-income countries, the quality of the school influences student achievement as much as these factors, especially in science and mathematics.\footnote{1} School quality includes the amount of instructional resources available per student (including teacher training) and how these resources are used and managed. Rising population growth rates, slowed economic growth, and mounting foreign debt in developing countries mean that there are more children to educate and less money with which to educate them. The effect on the quality of schooling varies from region to region and from country to country. The poorest countries in the Third World have less money to invest in items that boost learning such as textbooks, instructional materials, good teachers, and teacher training. Although some middle-income countries in Asia, Latin America, and the Middle East are experiencing slight increases in school quality, the gap between them and countries belonging to the Organization for Economic Cooperation and Development (OECD) continues to widen.\footnote{2} As a result, educators and economists alike are

\textit{Correspondence}: Dr. Stephen P. Heyneman, Economic Development Institute, The World Bank, 1818 H St. N.W., Washington, D.C. 20433.

intensifying the search for an inexpensive means of improving the quality of education in developing countries.

Examinations can be a powerful, low-cost means of influencing the quality of what teachers teach and what students learn in school. Examination agencies have an important role to play in increasing the effectiveness of schools. They can act not only as ex post evaluators of educational achievement but can also take responsibility for making good teaching and learning happen. This paper suggests two criteria for grading the performance of examination agencies. One is the quality of the tests they design. Among other things, a good examination agency should design and efficiently administer excellent tests—that is, tests with an appropriate distribution of questions involving recall, application, synthesis, and evaluation.

The second criterion is the quality of the feedback on student performance an examination agency provides students, teachers, school administrators, and the community. The quality of feedback is determined by the level of analysis of the pattern of errors made in past student performance—that is, why wrong answers were chosen. Analysis of errors can be broken down, on the one hand, by looking at subject, type of skill or cognitive process, and specific test question, and, on the other, by looking at the country as whole, each region, each district, each school, each classroom, or in some cases each individual student. The lower the unit of analysis the higher the grade for the examination agency. The article also examines some special policy problems that face testing experts, including the language of testing, test and curriculum coverage versus fairness, and the professional and political demands on testing.

BACKGROUND AND THEORY

During the last two decades the demand for education has increased worldwide. High population growth rates and an unfavorable economic climate, in particular, have limited the ability of developing countries to respond. In many countries, enrollments can barely keep pace with the school-age population. The expansion that has occurred in some countries has been achieved with increasingly scarce resources and at the expense of overall quality. In other countries education has had to compete, more often than not unfavorably, with other sectors such as national defense and debt service.

As a result of serious resource constraints, many countries have untrained or poorly trained teachers and inadequate teaching materials. The situation varies widely across geographical regions according to the relative wealth of countries, how much of the national budget goes to education, and how education expenditures are allocated—for example, how much goes to teacher salaries as opposed to textbooks, equipment, or teacher training. For example, in 1980 Middle Eastern and North African countries spent on the average US$221 per pupil, East Asian
and Pacific US$210, Latin America and the Caribbean US$209, Sub-Saharan Africa US$92, and South Asian countries US$17. Within South Asia, primary school expenditures ranged from US$7 per pupil in Bangladesh and Burma to US$24 per pupil in India. OECD countries spent on average US$2,297 per pupil in 1980. For the same period, Latin America spent US$8.99 per pupil on instructional materials versus US$3.28 in the Middle East and North Africa, US$2.49 in Sub-Saharan Africa, US$2.06 in East Asia and the Pacific, and US$1.26 in South Asia. Industrialized countries spent on average US$92.32 per pupil.\(^4\) Low investments in education coupled with poorly qualified teachers and inadequate teaching practices are indirect indicators of declining school quality. Information on the effectiveness of school outcomes is more scarce but offers more direct evidence of declining educational performance. The evidence available reveals low achievement levels in many developing countries and an increasing gap in the average quality of educational services compared to the situation in industrialized countries.

Mean scores on the second International Association for the Evaluation of Educational Achievement science study of fifth graders were much lower for the Philippines (9.5 out of 24) compared to Hong Kong and Singapore (11.2 out of 24), or Japan and Korea (15.4 out of 24).\(^5\) The second IEA study of achievement in mathematics conducted in 1981–82 included Nigeria, Swaziland, and Thailand along with 14 industrialized countries and 3 upper-middle-income countries. The 3 upper-middle-income countries generally performed on the same level as the industrialized countries, while Nigeria, Swaziland, and Thailand performed at a much lower level. Students in Nigeria and Swaziland answered just over half as many items correctly as students in Japan, the highest-scoring country. When given the same test, students in Malawi performed lower in reading than students in all the IEA-surveyed countries and lower than students in all but three countries in science, even though the Malawian students were on average six years older.\(^6\) In Ethiopia the highest failure rates on the 1987 secondary school leaving examination were in math and physics, 48.1 and 44.8 percent respectively.\(^7\) In Bangladesh failure rates on the secondary and higher-secondary certificate examinations are 50 percent. Only 5 percent of those who enter primary education pass the higher-secondary certificate examination.

The examinations rely heavily upon recall of factual—sometimes esoteric—information. The emphasis in the classroom is on rote learning to pass classroom tests and the national examination. There is very little incentive to pay attention to developing reasoning, imagination, and independent inquiry. Graduates thus are not usually prepared for the demands of reasoning ability and initiative in the workplace. The majority of students may not even achieve the basic literacy and numeracy skills needed to improve farming, family nutrition, or health practices.\(^8\) Achievement also varies significantly from region to region within the same
country depending on the type of community and school. Sometimes private schools are of higher quality and have higher achievement scores than public ones. Urban schools often score higher than rural ones, largely because scores are also higher in higher per capita income areas where classroom resources are more abundant.

When school performance is inadequate it weakens the future skills, adaptability, and competitiveness of a country’s labor force. New pressures on already restricted education budgets have made making better use of scarce resources and maintaining acceptable standards of quality a high priority for educational officials in developing countries. Reforms in the area of examinations have the potential to contain unit costs and enhance quality. Examinations are valued as uniform mechanisms for identifying talent and measuring achievement. Especially in environments where education resources are limited or unequally distributed among schools, examinations can help to ensure that society is investing in those who will in turn make the most useful contributions to society. A well-designed examination system can monitor and measure achievement and, occasionally, aptitude; provide performance feedback to individual districts, schools, and students; and inform education officials about the overall strengths and weaknesses of their education systems and suggest directions for change and improvement. However, the most important aspect of examinations is the degree to which teachers, administrators, students, and parents think of them as important. Selection examinations are one of the most powerful motivational levers in the education sector. The question is whether their influence is positive or negative.

*Educational Reform and Examination Reform*

Deteriorating standards, high dropout and repetition rates, labor shortages in certain fields, high youth unemployment, and the demands of ambitious development plans are the main factors motivating educational and examination reform. Such reforms are usually aimed at making the curriculum "more relevant" to development needs or at redressing regional imbalances in the quality and opportunity for education. If reforms are contrary to public or parental aspirations, officials use the power of examinations to make reforms acceptable. However, if examinations themselves do not fulfill parental aspirations for social and economic advancement, reforms are likely to fail. For example, Sri Lanka abolished 0- and A-level examinations in 1974 and 1975. These are achievement examinations at the end of junior and senior secondary school, respectively, that certify competency in subjects of the curriculum. However, education officials no longer wished to use them as instruments of selection for further education but rather as measures of educational attainment. In an attempt to align education with the economy, it was argued that new national examinations should be designed to correspond to a new curriculum emphasizing prevocational subjects. Public
pressure forced a reversal of this decision. The new national examinations were not recognized by many universities for admissions purposes. These reforms failed in part because public confidence in the value of the certification of the new examination was undermined. The power of examinations lies in their ability to allocate life chances. If they lose or are perceived to lose that ability, they can also lose the power to influence educational policy.

*Examinations Are Necessary*

Attitudes toward testing have changed. In the 1960s, many educators viewed tests as antithetical to creative thinking, as unreliable predictors of future academic success, and as unfair to disadvantaged groups. In the United States, many thought that teaching the basic skills was only a part, and perhaps a lesser part, of schooling’s mission. Schools, it was said, should respond to pupil interests and needs to make learning more attractive and productive. Those who held such views tended to deny the educational value of tests. Newly independent countries sometimes considered testing an unwanted colonial legacy, irrelevant for building new national identities. In some countries such as Tanzania and the People’s Republic of China, education officials replaced standardized tests with school-based assessments and experimented with other criteria for selection and certification such as political attitudes or quotas based on socioeconomic, geographic, and ethnic origins or on gender. Unforeseen problems arose. With school-based assessments it is difficult to ensure that all teachers are judging on the same criteria. Since parents judge teachers on their ability to get their children ahead, teachers tend to inflate grades. Judging a child’s potential on political attitudes is not reliable since attitudes can be faked. Though quotas are sometimes necessary to ensure equal utilization of facilities, making selection decisions solely on ethnic or geographic criteria can lead to abuse and can also create resentment among groups not selected. However, whether criteria are academic or political, the need for some fair and efficient mechanism to recognize and reward ability is indisputable. Examinations are thus reemerging as instruments of educational and occupational selection. The challenge is how to make such a mechanism reinforce the teaching and learning of skills that are useful to a society.

*The Backwash Effects of Examinations*

Especially in the context of scarce resources, with declining educational standards and the ever-increasing demand for better-qualified manpower, education officials are looking for low-cost ways to improve their education systems. What makes testing powerful is its use to allocate life’s chances. This is especially the case in developing countries where rates of returns from education are higher and where in some instances high income-earning opportunities are available almost exclusively to those with access to postsecondary education.
When examinations determine a child’s advancement through school and his or her later life’s opportunities, parents understandably put pressure on teachers to ensure that their child succeeds. They hold the school system and particularly teachers accountable for their child’s results on examinations. The consequence or backwash effect of this public expectation is that teachers adjust their teaching to what the examination will cover to ensure that their students score the highest marks. Teachers are less concerned with whether the test measures the full range of competencies set out in the official curriculum or whether the knowledge, concepts, and skills tested are of as much use to students who will enter the workforce as to those who will continue to the next level of education. If examinations fail to test useful skills, teachers will have little incentive to teach them. The backwash effect of the examination thus restricts what is taught and learned in school. How can officials tell the extent to which their examinations measure the mastery of relevant skills and therefore have a positive impact on teaching and learning?

Making Backwash Effects Positive

One means of making the backwash effects of tests positive is to improve the content of examinations. The content of examinations should correspond to the many functions they perform. In developing countries one examination is often used to serve several functions. Examinations are used to select students to secondary and higher education and to certify mastery of the primary or secondary school curriculum. Rarely are they used to diagnose learning problems, place students at different levels of ability, or to help teachers plan what to teach. Usually the content of the examination is determined much more by its function as a selection instrument. This enhances the importance of the knowledge tested in the examination and consequently of what is taught and learned in the classroom. Ideally one should test separately for selection, certification, and monitoring functions; however, resources do not permit this in many developing countries. Another means of making backwash effects positive is to set up a good feedback mechanism to analyze and interpret student errors. Explaining the thought processes behind wrong answers and making performance results public give school and testing officials incentive to do their jobs better. A third means is to make sure that the examination body is financed and managed in such a way that it can set good test questions and provide the necessary feedback on student performance.
What Makes a Good Test Item?

Good examination questions should test more than the ability of the candidate to recall isolated facts. They should test the ability to observe, experiment, and interpret, to understand concepts and draw reasoned conclusions, to use knowledge and skills to solve problems, and make decisions in new situations and contexts both in and out of school. Good test questions require the mastery of knowledge and skills that when applied in everyday life improve the quality of life and help families use their limited resources better. Anthony Somerset, who worked on improving the quality and relevance of test questions for the Kenya primary school-leaving examination, suggests eight criteria for enabling an examination agency to design the right balance of test questions:

1. Most questions should require students to restructure information rather than simply reproduce it.

2. Many questions should require candidates to understand and use information they have not seen before, in new situations and contexts.


4. The knowledge elicited should require integrating existing and new knowledge explicitly related to the competent performance of some target behavior.

5. A high proportion of items, especially in science, health, agriculture, and geography, should draw upon both in- and out-of-school experience.

6. When an examination is given in languages other than the mother tongue, test designers should carefully monitor questions for language loading. Test questions should not contain idiomatic expressions or the use of "registers" not accessible to all. Good questions should measure mastery of knowledge and skills rather than mastery of the language of the test.

7. Though they are more expensive, some proportion of the questions should be open ended (short answer, long answer questions). Candidates should generate answers as well as select them. In the world outside of school, one is rarely presented with possible solutions to a problem from which to choose the correct response.

8. Some questions should test the creative, imaginative skills of learners—that is, be based upon giving an unusual response, despite the problems associated with marking such responses.
What Constitutes a Good Feedback Mechanism?

If test questions measure the acquisition of knowledge useful to school leavers as well as to those continuing to the next level of education, the resulting positive backwash effects can benefit the entire school system. To tap these benefits, examination agencies must first inform teachers and students, well ahead of time, of changes in the content of examination questions and of the nature of the skills being tested. Second, they have to analyze performance on individual questions to determine what errors students are making and why. Third, the examination body has to suggest ways to improve the teaching of skills that require mastery of difficult concepts.

In addition to teachers and students, the audience for such feedback on test results includes teacher training institutes—so that future teachers will learn to teach difficult cognitive skills; curriculum development institutes and textbook designers—so that they can improve the way concepts and principles are identified and explained; and in-service training programs for inspectors, education officers, and head teachers—so that they can better focus their attention on managing the most crucial functions of the education system.

Examination bodies can provide different levels and depths of feedback. They can report only individual performance as a general mean and standard deviation, or they can break mean scores and standard deviations down by subject (e.g. geography), type of skill (e.g. map reading), and individual test item. They can report these, in turn, for the country as a whole, each region, each district, each school, and, ideally, each classroom or each individual student. Any testing agency that can do all this has a strong feedback system. Unfortunately, few agencies even in OECD countries provide this level and depth of feedback. Most agencies see their task of explaining past performance as supplementary to their main function of grading individuals. Thus they rarely report more than general means and standard deviations for the country as a whole or for each region and seldom go into detail more specific than subject-by-subject results.

What are the aims of feedback systems? They help reduce performance gaps between different schools or districts by showing where differences in performance occur and why. They encourage better educational management by identifying specific cognitive skills that need to be strengthened. They also suggest what type of investments in schools—for example, more and better teachers, better textbooks, and better administrators—influences achievement.

What Makes a Good Examination Body?

A good examination body acts not only as an evaluator of educational achievement but also takes responsibility for making good teaching and learning happen. It does this by constructing tests that have a positive backwash effect on teaching and learning and by showing school systems the sources of children’s
learning problems. To provide high-quality feedback, an examination agency should analyze the pattern of errors made in past performance and break those patterns down to the lowest-possible unit of analysis—ideally that may even be down to the level of specific questions and to the level of the individual school or classroom. However, a testing agency cannot design good test items and set up a good feedback system if it is not efficient, credible, and autonomous. To be efficient it must produce tests on time, within its budget, and within its competence. To be credible it must maintain professional standards when setting and marking tests and not succumb to political pressures to test inappropriately. To be autonomous it must be managed as an entity separate from the Ministry of Education and be able to hire the technical expertise necessary at competitive salary levels and to purchase the technical equipment necessary independent of public-sector regulations. It therefore requires a budget separate from the normal rises and falls typical of most public-sector enterprises—that is, it will most probably have to be self-financed. It must create a demand for its services, for which clients are willing to pay. Once self-financed, a testing agency can pursue its professional responsibilities without political interference.

DILEMMAS AND TRADE-OFFS

Some issues in the area of examinations and testing have no hard and fast answers. They include dilemmas such as: test coverage of the curriculum versus fairness of the examination, mother tongue versus metropolitan language as the language of testing, professionalism of examination bodies versus political influence, and well-constructed test questions versus socioeconomic advantages.

Coverage versus Fairness

The extent to which the national examination covers the official curriculum varies from country to country. Tests and the curriculum influence what teachers teach in the classroom and what students actually learn. Discrepancies between test content and the taught curriculum pose a dilemma since tests are supposed to choose and certify individuals on the basis of their knowledge of the common curriculum. If testing experts intend to draw conclusions from test results about the quality of teaching and learning, then the match between what is actually taught and test content must be a close one. In Africa, for example, where schools are poorly equipped and learning materials scarce, teachers are typically able to cover only a small portion of the curriculum. The implications of this are particularly serious in countries where outside influences such as previous learning and home and social environment are less a factor and where achievement may be more directly traceable to school quality.

The amount of the curriculum covered varies enormously—rich schools will cover more, poor schools less. What should the test designers do in cases where
the quality of schools varies? Should they set questions on the intended cur-
riculum even though some parts may have been taught in only 10 to 15 percent
of the schools? If that is done, the test will not reflect what has been learned but
rather who had opportunity to learn. Alternatively, should the test emphasize those
parts of the curriculum for which it can be shown that a wide proportion of
students had an opportunity to learn—nonlaboratory sciences instead of labora-
tory techniques, for instance, since typically only wealthy schools can afford the
laboratory materials to teach that portion of the science curriculum. If this is done,
the test will be a truer reflection of what has been learned by the student
population, but it will not accurately reflect the intended curricular objectives.

National tests are designed to measure commonly taught knowledge, and
since the ability to teach that knowledge is shrinking, the portion that can be fairly
tested is also shrinking. Schools that wish to do well must teach the knowledge
and skills identified as important for pupils to learn. Better equipped schools, with
well-trained teachers and active head masters are more likely to cover the
curriculum than are poorer, ill-equipped schools. If test content is based on the
official curriculum rather than on what is actually covered in the classroom, it
will favor students from schools that are better able to cover the curriculum and
thus will risk measuring students’ opportunity to learn rather than what they have
actually been taught. If tests measure more than what is covered in the classroom
they can be unfair and unjust.

In situations where demand for schooling is high and school places scarce,
test scores are important for the future careers of students and teachers alike.
Students will decide on what areas to concentrate their studies based on what
areas are tested. Teachers will make decisions on what they will cover in the
curriculum and on how they organize and structure teaching on a similar basis.
Usual methods of policing teachers (better teaching materials, in-service training,
more frequent inspection) to make sure they are covering the curriculum will not
work if there are pressures on teachers to make sure their students do not fail. Test
content is a powerful influence on teachers to change their decisions on what part
of the curriculum to teach. Examination authorities can ensure a reasonable
correspondence between the content of tests, the intended curriculum, and what
is actually taught in the classroom. Success will depend on their willingness to
challenge current practice. In Kenya, for example, subjects such as health,
nutrition, Kiswahili, agriculture, and crafts were not tested and therefore not
emphasized in the curriculum. Schooling is the only way people in rural areas can
acquire knowledge of modern science and its applications for better health,
nutrition, or agricultural practices. To encourage the coverage of scientific sub-
jects in the curriculum the examinations council set new test papers in these
subjects. The tests were made compulsory. Carrying out the changes required
additional facili-ties, new textbooks, new equipment, and in-service teacher
training. Initial public and professional resistance to the changes added to the difficulty of the reform. However, the changes had the desired result—the coverage of the curriculum increased in both urban and rural schools. In this case, the testing body took the initiative and researched how much of the curriculum was being taught and where. It then set compulsory examinations to cover subjects of importance to society and advertised new coverage to encourage teachers to teach it.

The Language of Testing

The language of assessment poses many questions for testing and instruction. In multilingual societies, the rationale for using international languages such as English, French, or Spanish for teaching and testing is that they do not favor any particular linguistic group. How adequately is achievement measured in the second language? How does the tester of mathematics know that the scores of students are a reflection of what students know in mathematics rather than what they understand of the English-based mathematics questions? Does the language of the test determine how exam items are constructed—that is, do test questions measure understanding of concepts and underlying principles or rather the language-based recognition of the correct answers? Tests in languages other than the mother tongue may be measuring the ability to express concepts in the language of the test rather than the ability to understand the concepts themselves. There is more evidence of problem solving and other higher-order abilities when students are tested in their mother tongue. Particularly in science subjects, testing in the second language may discourage the use of concepts and principles of modern science in everyday life. The existing literature suggests a larger role for indigenous languages in testing on the grounds that testing in an international language may confuse subject knowledge with comprehension of the language of the test.

The primary goal of learning should be understanding concepts and principles in order to make inferences from that knowledge and to apply it in daily life. Tests should accurately measure students’ achievement of these competencies. Do tests in a second language elicit meaningful evidence of student achievement of competencies in a subject area? A good test should measure knowledge in a particular subject area, not understanding of the language of the test. If a language other than the mother tongue is the medium of the test, children with more access to the language of the test will do better regardless of whether they know the subject matter better. A testing agency that claims to certify that one student knows more than another in a particular subject must ensure that the language bias of test questions is minimized. In multilingual environments testing agencies can handle the language question in two ways. They can make sure that tests are free of language experience and that test questions are accurately measuring subject-
matter competence. In other words, they can pretest the language of the test item thoroughly and satisfy themselves ahead of time that there are no differences between performance of mother-tongue and non-mother-tongue students. Alternatively, they can use the mother tongue in testing to ensure that those who perform poorly in the second or international language, but well in subject matter, are not discriminated against. A testing agency that chooses the second option will probably run into political problems with those who have a vested interest in testing in one or the other language. They will also face the high costs and logistical complexity of setting questions, administering, marking, and analyzing results of subject matter tests in several languages.

**Professionalism versus Political Influence**

If most people agree that examinations are necessary, they also agree that examinations should be fair. In particular, parents insist that tests be fair in the interest of their children’s welfare. The quality and fairness of tests will depend on the testing agency. If testing agencies are solely dependent upon government funds to perform their functions, they will be more vulnerable to pressures to test inappropriately without regard for standards and resources. They may be pressured to test in a particular language for political or cultural reasons rather than for pedagogical ones. Such was the case with Algeria and Tunisia, which adopted classical rather than spoken Arabic as the medium of instruction in primary schools. Testing agencies may also be pressured to design tests to assess skills that are difficult to measure—such as prevocational subjects. In these instances examination content is often not only intended to measure competence but also to change public attitudes toward prevocational subjects and manual work. Usually high passes in one of the prevocational subjects become a requirement for entry into the next level of schooling.

**Assessing Practical Skills**

Making the curriculum relevant to local needs and to children who do not continue schooling to higher levels is a concern of education officials everywhere. The question is whether subjects that prepare students for progress up the educational ladder are relevant for those who do not continue their education; whether general academic subjects such as math, science, and languages are more useful to school leavers in the long run than are skill-specific courses such as carpentry, metal work, and auto mechanics. Competence in math, science, and languages is associated with the industrial and technological development of OECD countries. Similarly, competence in these areas will likely determine how easily developing-country labor forces keep pace with rapid technological change. Development theory in the 1960s and 1970s advocated prevocational subjects at lower levels of education for those who would end their schooling at
this level. Although these policies did not yield the quantity and quality of manpower desired, curricula of most developing countries still contain a practical-work component. The proper mix of theory-based subjects and practical work remains controversial.

In many developing countries, it is very common for politicians to fear that the schools and universities are not responding to the needs of society. Education is blamed for not having created job opportunities or for not having produced the quantity and quality of technical manpower needed for economic development. The content of the curriculum is seen as having little to do with the world of work but catering instead to the small number of students who attain the highest levels of education leading to professional occupations.

Neither parents nor teachers are convinced of the importance of practical subjects vis-à-vis academic ones. Parental and societal perceptions of the manual skills associated with practical work are usually negative. Parents aspire to ensure better lives for their children, usually equated with jobs in the modern sector that are better paid and less damaging to one's health than is working with one's hands. If the curriculum does not parallel those aspirations, parents react. In Kenya, for example, parents hired local carpenters to make the stools required by the schools for practical assessment in woodwork. This freed their children to concentrate on academic subjects and on preparing for the examinations that would determine whether they moved up the educational ladder. In Sri Lanka, the popular press criticized the field workbook portion of internal assessment for the junior-secondary social studies program because it took time away from other subjects and because it favored the children of wealthier, better-educated parents. Some parents did the field workbook for their children and purchased the commercially prepared scrapbooks so they could get a good grade. Children from poorer homes could not get suitable information from their parents, who also could not afford the cost of materials for field workbooks.

How to assess practical work in subjects such as woodwork, metalwork crafts, home economics, chemistry, physics, and biology also remains problematic. A large part of the problem of testing practical skills stems from political pressures to test them in the same manner as other academic subjects—in other words, to rank the performance of students against each other rather than against the mastery of agreed upon objectives (norm-referenced instead of criterion-referenced tests). In most African countries, testing agencies lack the money and the logistical and technical requirements to effectively assess practical skills using norm-referenced tests. Often the type of skills that are to be developed from the practical work is not clearly specified. Equipment and materials for practical work and assessment are costly and often not available. This keeps teachers from doing practical work throughout the school year. Equipment lists and guidelines sent to students prior to the practical examination give teachers an idea of the examina-
tion content and coverage and increase the risk of security leaks. Uniformity of standards of measure and marking is difficult to achieve especially in areas where there are many schools and moderators involved in the assessment process. As a result, grades tend to be inflated.

Even so, political officials feel that practical work is necessary, especially in the area of agriculture, where it is assumed that practice will foster the desire to become a farmer. They recognize the problems of high costs, added time, poor sampling, absence of reliable assessment criteria, and noncomparability of performance associated with assessing practical work. Education officials in India introduced changes to mitigate these limitations of practical assessment for science subjects. The first was to test only practical skills in practical work, assessing objectives such as knowledge, understanding, and application in written tests whenever possible. To improve the sampling the number of exercises assessed was increased. The skills associated with mastery of practical work in science subjects were carefully defined to improve their validity. Finally, examiners received intensive training in the new pattern of practical examination including detailed instructions on marking and interpreting procedures to minimize the variability in scoring.

However, costs remained high. It was difficult to equip all schools fully. Lack of a standard list of tools, equipment, and materials needed to do practical work led some schools to water down practical work to fit the facilities available. Sri Lanka ran into similar problems when implementing its prevocational studies program. Attempts to get the community to provide some materials were not entirely successful. Schemes to derive income from marketable goods risked competing with community suppliers and thereby defeating goals of breaking down traditional barriers between the school and community. Teachers found it easier to concentrate on teaching manual skills rather than on teaching applications of mathematical and scientific principles to a particular vocation or the appreciation of new values and attitudes toward manual work. Moreover, some vocations were too simple to have wider applications. The reform measures did result in more active participation of teachers and members of the community in developing curriculum and examination syllabuses.

Practitioners suggest possible ways to minimize the limitations of assessing practical skills. It may be possible to test practical work on the theory portion of tests for subjects like biology, chemistry, and physics, which do not lead to an occupation. Since the correlation between theory and practical skills is low, marks awarded may not give an accurate picture of mastery of practical skills. Continuous assessments involving the private sector and industry could be used for subjects with occupational applications. Student results from these would then be made available to school inspectors. Teachers and external examiners could be more adequately trained in assessment techniques for practical work, especially
when practical work is taken into account in the awarding of a certificate. However, given the enormous requirements and costs of adequately teaching and testing practical work in subjects that lead to an occupation, the question remains whether school is the right place to learn this type of knowledge.

CONCLUSION

The influence of assessment on instruction is well documented, particularly in many developing countries where incentive to perform well on examinations is very high due to the scarcity of educational opportunity. This article has suggested ways in which examination agencies can make effective use of this influence in order to improve the quality and usefulness of what students learn in school. It has argued for an expanded role of examination agencies—from that of evaluators of student achievement to promoters of good teaching and learning. Examination agencies can perform two essential functions that would have a positive impact on teaching and learning. First, they can design good test questions that could draw upon both in-and out-of-school experience and measure higher-order skills of evaluation, application, and synthesis useful to students who terminate their studies as well as to those who continue. Shifting the balance of test questions from measuring factual recall to reasoning and problem-solving skills will ensure that teachers teach these skills in the classroom. Second, examination agencies should provide students, teachers, school administrators, and the public with regular and systematic information on student performance on examinations. This requires a careful analysis of the reasons for and results from individual test questions, of why wrong answers occur, and of the distribution of error occurrence in the school system. With this information teachers and school administrators will know where weaknesses and gaps in performance are and can make appropriate decisions to redress them (i.e. more and better teacher training, instructional materials, etc.). Finally, an effective testing agency must be able to make professional decisions on politically and professionally controversial and sensitive issues such as the language of assessment, test coverage and fairness, and appropriate testing for practical subjects. It is unlikely that a testing agency can take on these responsibilities without some measure of financial and professional autonomy. This new and expanded role in education requires testing agencies to provide a high quality, competitive product for which their clients are willing to pay. An agency that has an independent source of revenue can make decisions on testing policy and practice. Free of political pressure to test inappropriately, it can achieve excellence in examination practice and thereby have a positive impact on educational quality.
NOTES

The views and interpretations in this article are those of the authors and should not be attributed to the World Bank.


3. Ibid.


