



Association for the Development of Education in Africa

**ADEA Biennial Meeting 2003
(Grand Baie, Mauritius, December 3-6, 2003)**

**Monitoring Performance:
Assessment and Examinations in Africa**

by Thomas Kellaghan and Vincent Greaney

**Working Document
DRAFT**

PLEASE DO NOT DISSEMINATE

Doc. 2.D

This document was commissioned by ADEA for its Biennial Meeting (Mauritius, December 3-6, 2003). The views and opinions expressed in this volume are those of the authors and should not be attributed to ADEA, to its members or affiliated organizations or to any individual acting on behalf of ADEA.

The document is a working document still in the stages of production. It has been prepared to serve as a basis for discussions at the ADEA Biennial Meeting and should not be disseminated for other purposes at this stage.

© Association for the Development of Education in Africa (ADEA) – 2003

Association for the Development of Education in Africa (ADEA)

International Institute for Educational Planning

7-9 rue Eugène Delacroix

75116 Paris, France

Tel.: +33(0)1 45 03 77 57

Fax: +33(0)1 45 03 39 65

adea@iiep.unesco.org

web site: www.ADEAnet.org

Contents

ABSTRACT	8
1. EXECUTIVE SUMMARY	9
1.1. PUBLIC (EXTERNAL) EXAMINATIONS	9
1.2. NATIONAL ASSESSMENT	10
1.3. INTERNATIONAL ASSESSMENTS.....	11
1.4. CLASSROOM ASSESSMENT	11
1.4.1. <i>Problems in the use of information derived from assessments in classrooms.....</i>	<i>12</i>
1.5. MECHANISMS TO SUPPORT THE USE OF ASSESSMENT.....	12
2. THE CONTEXT OF PROPOSALS TO USE ASSESSMENT TO IMPROVE THE QUALITY OF EDUCATION	13
2.1. EXTENT AND QUALITY OF EDUCATIONAL PROVISION IN AFRICA	13
2.2. OBSTACLES TO MEETING EFA GOALS	14
2.3. RECOMMITMENT IN DAKAR TO EFA	15
2.4. FOCUS OF THE PAPER.....	16
2.5. PUBLIC (EXTERNAL) EXAMINATIONS.....	17
2.6. THE QUALITY OF PUBLIC EXAMINATIONS IN AFRICA	18
2.7. IMPROVING THE QUALITY OF PUBLIC EXAMINATIONS.....	19
2.8. USING DATA FROM EXAMINATIONS TO DESCRIBE STUDENT ACHIEVEMENTS	19
2.8.1. <i>Chief examiners' reports</i>	<i>20</i>
2.8.2. <i>Other feedback mechanisms</i>	<i>21</i>
2.8.3. <i>Identification of categories of pupils who perform poorly</i>	<i>22</i>
2.8.4. <i>Limitations of the data provided by examinations.....</i>	<i>22</i>
2.9. USING EXAMINATIONS TO IMPROVE STUDENT ACHIEVEMENTS	23
2.10. EFFECTS OF CHANGES IN EXAMINATIONS	23
2.10.1. <i>Is a change in examinations likely to result in an improvement in the level of student achievements?</i>	<i>24</i>
2.10.2. <i>Will a change in examinations result in change in students' cognitive processing skills?</i>	<i>25</i>
2.10.3. <i>Mechanisms to bring about change in school practice and student achievement.....</i>	<i>26</i>
2.10.4. <i>Competition between schools</i>	<i>26</i>
2.10.5. <i>Provision of guidance to schools.....</i>	<i>27</i>
2.10.6. <i>Support for underperforming schools.....</i>	<i>27</i>
2.10.7. <i>Linkage with curriculum.....</i>	<i>28</i>
2.10.8. <i>Particular problems when high stakes are attached to assessment information.....</i>	<i>28</i>
2.10.9. <i>Guidelines to improve public examinations</i>	<i>31</i>
3. NATIONAL ASSESSMENTS	33
3.1. INFORMATION SOUGHT IN NATIONAL ASSESSMENTS.....	34
3.1.1. <i>Uses of national assessment information</i>	<i>35</i>
3.2. VARIATION IN NATIONAL ASSESSMENTS.....	35
3.3. NATIONAL ASSESSMENT ACTIVITY IN AFRICA	35
3.4. IMPACT OF NATIONAL ASSESSMENTS	38
3.5. ISSUES IN NATIONAL ASSESSMENTS	39
3.5.1. <i>Conditions and practices required to develop and institutionalize the capacity to carry out a national assessment.....</i>	<i>41</i>
4. INTERNATIONAL ASSESSMENTS	43
4.1. INTERNATIONAL ASSESSMENTS IN AFRICA	43

4.2. PROBLEMS ASSOCIATED WITH INTERNATIONAL ASSESSMENTS	44
5. CLASSROOM ASSESSMENT	46
5.1. THE QUALITY OF CLASSROOM ASSESSMENT.....	46
5.2. IMPROVING CLASSROOM ASSESSMENT.....	48
5.3. THE USE OF SCHOOL-BASED ASSESSMENT IN PUBLIC EXAMINATIONS.....	49
6. FACTORS THAT MAY INHIBIT THE USE OF INFORMATION DERIVED FROM ASSESSMENTS AND EXAMINATIONS IN CLASSROOMS.....	52
6.1. MODE OF INTERVENTION	52
6.1.1. <i>Relevance of information</i>	52
6.1.2. <i>Teacher competence</i>	53
6.1.3. <i>Teachers' understanding of implications of changes</i>	53
6.1.4. <i>Complexity of teaching</i>	54
6.1.5. <i>Classroom context</i>	55
6.1.6. <i>Opposition based on the perception that a change will involve a risk to pupils.</i>	55
7. CONCLUSION	56
7.1. RECOMMENDATIONS	60
8. REFERENCES	62

Acronyms and abbreviations

ADEA	Association for the Development of Education in Africa
CONFEMEN	Conférence des Ministres de l'Éducation des pays ayant le Français en partage
MLA	Monitoring Learning Achievement
NESIS	National Education Statistical Information Systems
PASEC	Programme d'Analyse des Systèmes Éducatifs des Pays de la CONFEMEN
PRSP	Poverty Reduction Strategy Papers
SACMEQ	Southern Africa Consortium for Monitoring Educational Quality
SAP	Structural Adjustment Programs
UNESCO	United Nations Educational, Scientific and Cultural Organization

Boxes

Box 1. Chief examiners' reports in Swaziland.....	20
Box 2. Language of instruction.....	21
Box 3. Problems with league tables	27
Box 4. How teachers are influenced by examinations	29
Box 5. Some forms of examination corruption	31
Box 6. Guidelines to improve public examinations.....	32
Box 7. Main elements of a national assessment.....	33
Box 8. Differences in national assessment procedures	36
Box 9. South Africa's experience with international assessments.....	45
Box 10. Classroom Assessment	46
Box 11. Suggestions to improve classroom assessment.....	49
Box 12. Teachers' difficulties in adjusting to changes in examinations	54
Box 13. The complexity of teaching	54

Acknowledgement

We wish to acknowledge the assistance of the following in the preparation of the paper: R. Agabu, Tunde Adekola, Jean-Marc Bernard, E.D. Bononewe, Jaap Bregman, Vinaygum Chinapah, Paul Coustère, Brigitte Duces, Mourad Ezzine, Birger Fredriksen, Jacques Hallak, Soumana Hamidou, Cathal Higgins, Anil Kanjee, John Khozi, John Lauglo, Lily Mulatu, Paud Murphy, Bob Prouty, Rachidi B. Radji, Ken Ross, Charlotte Sedel, Endale Shewangizaw, Mercy Tembon, Gary Theisen, Adriaan Verspoor, and David Weere. Secretarial assistance was provided by Hilary Walshe.

ABSTRACT

The rationale for, and experience with, the use of assessment to improve the quality of education in African education systems over the past two decades are described for four categories of assessment: (i) public examinations, (ii) national assessments; (iii) international assessments, and (iv) classroom assessment. The case for reform of *public examinations* is based, first, on a perception that their quality is poor in many countries and, secondly, on the belief that since teachers align their teaching to the demands of examinations, improving them will effect change in what teachers teach and, as a consequence, in the quality of students' learning. Examples of the use of examination data to improve student achievements are provided. Evidence is considered relating to the effects of changes in examinations on the content of teaching, level of student achievements, and students' cognitive processing skills. National assessment activity spread through the whole of Africa during the 1990s. Although information derived from *national assessments* is obtained from individual students (chosen to represent all students at a particular grade in the education system), data are aggregated to provide an assessment of the education system. The primary purpose of such an assessment is to describe how well students (and identifiable subgroups of students) are learning. To date, use of information derived from assessments seems limited for the most part to government commissions, reviews, and education reform programs in a number of countries. *International assessments*, which share many procedural features with national assessments, provide comparative data on achievement in more than one country. Few African countries have participated in them. However, some of the national assessments that have been carried out allow international comparisons. Although *classroom assessment* has attracted the least attention in proposals to use assessment to improve the quality of education, it is likely to have a greater impact on student learning than any other form of assessment. Advantages and disadvantages of school-based assessment in contributing to grades in public examinations (a feature of examinations in some countries) are considered. Proposals to improve public examinations, national assessments, and classroom assessment are presented. Factors that may inhibit the use of assessment to improve the quality of education are described.

1. EXECUTIVE SUMMARY

1. The World Bank policy paper of January 1988, *Education in Sub-Saharan Africa, Policies for Adjustment, Revitalization, and Expansion*, placed a major focus on improving academic standards through the reform of examination systems. In June of the same year, Donors to African Education (now the Association for the Development of Education in Africa) signalled their interest in examinations when they set up a Working Group on School Examinations to study and promote the role of examinations in quality improvement. Following this, the World Bank Education and Training Division of the Africa Technical Department prepared terms of reference to undertake studies of examinations in primary and secondary education in Sub-Saharan Africa. The studies were carried out, with funding from the World Bank and the Irish government in six anglophone countries, six francophone countries, and two other countries. A report of the studies, *Using Examinations to Improve Education. A Study in Fourteen African Countries*, was prepared by the authors of the present paper and was published by the World Bank in 1992.

2. This paper reviews what happened in the field of assessment since then. It deals with public examinations, but differs from the 1992 report in that new data were not specifically collected for it (except in regard to a few minor details). Rather, it revisits many of the issues that were raised in the earlier paper, particularly in relation to the role that assessment can play in improving the quality of students' learning. It also differs from the earlier paper in that its consideration of assessment is not limited to public examinations.

3. The World Declaration on Education For All in Jomtien in 1990, not only gave fresh impetus to issues relating to assessment, it necessitated the introduction of a new form of assessment – system assessment or national assessment – to determine if children were acquiring the useful knowledge, reasoning ability, skills, and values that schools promised to deliver. National assessment is the second major area of assessment addressed in this paper. International assessments, which share many procedural features with national assessment, are also considered. The fourth type of assessment addressed in the paper is classroom assessment.

1.1. Public (External) Examinations

4. Public (external) examinations have played a major role throughout the history of modern education in Africa. They serve a number of functions, the most important of which is to select students for successive levels in the education system. Despite their central role, many criticisms have been made of their quality. In particular, it has been observed that examinations are limited in the areas of knowledge and skills that they assess; they contain little reference to the knowledge and skills that students need in their everyday life outside the school; and they tend to measure achievement at a low taxonomic level.

5. This situation has two implications. First, issues arise about the validity of the examinations, in particular the extent to which they are biased towards the testing of competencies needed by students continuing their education into the next cycle, and so fail to adequately reflect the goals of curricula for students (a majority in most countries) who will not proceed to secondary education. And secondly, since teachers focus their teaching on what is assessed in an examination which has important consequences attached to performance, serious concerns arise about the character and quality of teaching and learning in schools which examinations foster.

6. Several education systems have taken steps to improve the quality of examinations over the past two decades by, for example, using more diverse modes of assessment, by including items that test higher-order thinking skills, and by assessing students' ability to apply their knowledge and skills in situations outside the school as well as in scholastic contexts. The expectation is that such improvements will impact on what is taught and what is learned. To reinforce the power of examinations to impact on practice in schools, a number of additional procedures are in place in some countries. These include the publication of examination results to increase competition between schools, the provision of information on student performance, the provision of support and guidance to underperforming schools, and establishing linkages between examinations authorities and those responsible for curriculum development and implementation.

7. While efforts to improve the quality of examinations are to be commended and encouraged, one should hesitate before placing too great a reliance on examinations to radically alter the quality of learning of all students. First, while a change in the content areas examined will result in a shift in the content to which students are exposed in class, there is only limited empirical evidence that changes in examinations will result in an improvement in the overall level of student achievements or in their cognitive processing skills. Secondly, the fact that high stakes are attached to examination performance can have a number of negative, if unintended, consequences for school practice, whatever the quality of the examinations. These include narrowing of the implemented curriculum, neglecting what is not examined, emphasizing learning strategies that are superficial or short-term (such as memorizing, rehearsing, and rote learning), devoting a significant amount of time to test-preparation activities, and a heavy reliance on extrinsic rather than intrinsic motivation in student learning. Of particular significance in the context of Education For All is the fact that teachers, because their reputations depend on how well their students perform in examinations, may focus their efforts on students who are most likely to succeed. This, in turn, may be associated with high rates of grade retention and early dropout.

1.2. National Assessment

8. While public examinations are a long-standing feature of education systems, national assessments (sometimes called system assessments, learning assessments, or less appropriately assessments of learning outcomes) are relatively new. A national assessment may be defined as an exercise designed to describe the level of achievements, not of individual students, but of a whole education system, or a clearly defined part of it (e.g., fourth grade pupils or 11-year olds). National assessments were introduced to address the fact that the educational data on inputs to education which had typically been collected in the past were often of little relevance or use to educational planners. National assessments would address this issue by providing information on the "products" or "outcomes" of schooling (e.g., student achievements, inequalities in the system) which, it was hoped, could be used in conjunction with input data to provide a sounder basis for policy and decision making in education.

9. The main issues addressed in national assessments are: How well are students learning? Is there evidence of particular strengths or weaknesses in their knowledge and skills? Do achievements of subgroups in the population differ? To what extent is achievement associated with characteristics of the learning environment? Do the achievements of students change over time?

10. A great deal of activity relating to national assessment was generated in Africa during the 1990s. Four major categories can be identified. Three involve similar activities in a number of countries: the "Monitoring Learning Achievement" (MLA)

project; the Southern Africa Consortium for Monitoring Educational Quality (SACMEQ) project; Programme d'Analyse des Systèmes Educatifs des Pays de la CONFEMEN (PASEC). In a fourth category, national assessments unrelated to any of these activities were carried out in individual countries. To date, MLA assessments have been carried out in 47 countries in Africa, SACMEQ assessments in 15, and PASEC in nine.

11. All the assessments share a number of features. All are policy-oriented, involving planners and managers, and are designed to provide information for policy making. All assess students' achievements in basic curriculum areas. All are sample-based. All view capacity building and the strengthening of the policy-maker/research nexus as major objectives. All adopt an input-output model of the education system and attempt to identify factors associated with achievement.

12. Impact has been reported for all the assessment programs in a number of areas: in policy debate and formulation, in reviews of educational policy, in national education sector studies, to reappraise the adequacy of resources, and to support policy decisions. Several countries have had as many as three national assessments (each sponsored or supported by a different agency). This would suggest that the assessments were not a response to locally perceived needs, and are not integrated into the normal structures and activities of ministries. Given the need for such integration, the cost of these activities, and problems in recruiting personnel with the required competence to carry them out, it is clear that there is an urgent need for rationalization.

1.3. International Assessments

13. International assessments differ from national assessments in that they can provide some indication of where the achievements of students in a country stand relative to the achievements of students in other countries. They also provide evidence on the extent to which the treatment of common curriculum areas differs across countries.

14. Only a few African countries have participated in international assessments. However, while MLA, PASEC, and SACMEQ were designed as national assessments, international comparisons have been made on the basis of published reports.

1.4. Classroom Assessment

15. The assessment of students' learning in the classroom (both by teachers and by students themselves) is an integral component of the teaching-learning process. Much of this kind of assessment is subjective, informal, immediate, on-going, and intuitive, as it interacts with learning as it occurs, monitoring student behavior, scholastic performance, and responsiveness to instruction. Its role is to determine students' current level of knowledge, skill, or understanding, to diagnose problems they may be encountering, to make decisions about the next instructional steps to take (to revise or to move on), and to evaluate the learning that has taken place in a lesson.

16. There is evidence that the quality of teachers' assessment practices may be deficient in many ways. Problems that have been identified include the use of poorly focused questions, a predominance of questions that require short answers involving factual knowledge, the evocation of responses that require repetition rather than reflection, and a lack of procedures designed to develop students' higher-order cognitive skills.

17. With some notable exceptions in francophone countries, the improvement of classroom assessment has received little attention in reforms. However, given its central

role in the teaching-learning process, such improvement should be accorded high priority in any reform that has as its objective the improvement of students' learning.

18. Several examination systems in Africa have introduced, or are planning to introduce, an element of school-based assessment in their public examinations. Indeed, some would hope ultimately, especially at the primary school level, to replace external examining completely with school-based assessment, perceiving that to be the only way in which the range of competencies specified in curricula can be validly assessed, and in which the negative effects of external examinations on teaching and learning can be removed. However, the implementation of school-based assessment as a component of public examinations has proved problematic. While the aspiration and motivation to introduce it have been high, practical difficulties have on more than one occasion resulted in failure, postponement, or the limitation of the school-based element to a minimal, almost token, amount.

1.4.1. Problems in the use of information derived from assessments in classrooms

19. Getting information to teachers, and effecting changes in their behavior, poses many challenges. Expectations that information from assessments and examinations (in publication of results, newsletters, workshops) will radically alter the culture of schools and substantially raise the achievements of all students need to be tempered by a consideration of factors that may serve to frustrate the intentions of reformers. Unless these factors are recognized and addressed, policies involving assessment may be very limited in their impact on the quality of student learning. The following factors are discussed: mode of intervention, relevance of information, teacher competence, teachers' understanding of implications of changes, complexity of teaching, classroom context, and opposition by teachers and parents based on the perception that a change would involve a risk to pupils.

1.5. Mechanisms to support the use of assessment

20. A conclusion from the review presented in this paper is that while assessment information can improve policy and the management of resources in education and can shape teachers' instructional practice, success is not assured. Success in the first place will be dependent on the political will of governments to support the effort. Effort is required at two levels. At one level, national assessment systems should be institutionalized and integrated into the structures and processes of government policy making and decision making. It will also be important to ensure that national assessments are aligned with other major instructional guidance systems in the education system -- other assessment systems (including the alignment of standards), curricula, teacher education, school capacity building, and measures to address inequities. Action is also required at the level of the school. This requires opening adequate channels of communication to inform teachers of the implications of assessment findings. This can be a difficult task when problems relate to instructional objectives, equity, and quality, in which case teachers may need considerable and continuing support in interpreting reforms and in devising appropriate teaching strategies based on their interpretation.

2. THE CONTEXT OF PROPOSALS TO USE ASSESSMENT TO IMPROVE THE QUALITY OF EDUCATION

21. The World Conference on Education For All (EFA) in Jomtien, Thailand in 1990 held out the promise of improving the equity and quality of education systems throughout the world as nations sought to secure their economic competitiveness in a global economy in preparation for the 21st century. Not only should educational opportunities be open to all children, but those who avail of them should acquire useful knowledge, reasoning ability, skills, and values (UNESCO, 1990). Since the Conference, student achievement has become a major point of reference in judging the quality of education. Because the Conference recommendations represented the convergence of views among national governments, organizations, and donor agencies, they have had enormous influence, and countries throughout the world have responded by giving priority to basic education; promoting gender equity in access to education; improving the educational opportunities of rural and poorest population groups; improving the quality of education at all levels; addressing the issue of the practical relevance for students' lives of school curricula; and capacity building.

22. The aspirations of the Conference were not very novel in many respects. From the 1960s to the 1980s, many developing countries had proposed programs to improve primary education which involved both quantitative expansion and qualitative improvement. Programs included the establishment of a better balance between academic and practical/technical subjects; the integration of school and community activities; the improvement of teacher preparation; the extension of inspection, supervision, and advisory services for teachers; improving the methods of assessing pupils; increasing the availability of textbooks; and increasing the efficiency of educational planning (Grisay & Mählck, 1991). However, the Jomtien declaration differed from these, not only in its international and inter-agency dimension, but in requiring definitions or targets for quality improvement, and in specifying a need to assess the effects of reforms on student achievements.

2.1. Extent and quality of educational provision in Africa

23. It is easy in retrospect to say that the target of the Declaration on Education For All to ensure that all nations would have universal basic education by 2000 was over-ambitious. The target was not met. Some countries forged ahead, others sustained their gains without making any breakthrough, and others slipped backward (S. Nalovu cited in ADEA, 1999). While net enrolment increased in Sub-Saharan Africa, the non-enrolled school-age population actually increased since 1980, following a decrease in the 1970s (UNESCO, 2000a). The magnitude of the task that remains to be accomplished becomes clear when one considers that 42.4 million school-aged children in Africa are not attending school (UNESCO, 2002). However, getting all children into school is only a first step in achieving the goal of education for all, since reform must focus on actual learning acquisition and outcomes, rather than exclusively on enrolment. Again, the magnitude of the task involved is obvious when one considers the quality of provision for those who are attending school. At a regional conference on Education for All for Sub-Saharan Africa, held in Johannesburg in December 1999, attended by ministers of education, representatives of civil society and international development agencies, it was concluded that for those who were in the education system, what was provided was "poor

and the curricula often irrelevant to the needs of the learners and of social, cultural and economic development” (UNESCO, 2000b, p. 26); that “education planning and management capacity ... remain largely underdeveloped” (p. 26); that “only a small proportion of children are reaching the minimum required competencies and our education systems are not performing to the standards we expect of them” (p. 28).

24. Commentary on education in individual countries throughout the continent does nothing to alleviate this gloomy picture. It has been noted, for example, that education in Ghana is confronted with the massive problem of poor quality as attested by “abysmal national criterion-referenced test scores” from 1992 through 1996 (N’tchougan-Sonou, 2001). In Ethiopia in 1990, over 20% of candidates at the grade 6 examination, 40% of the grade 8 examination, and 70% at the senior secondary leaving examination failed. And this was so despite the fact that about 20% of pupils in grades 1 and 9 were repeaters. In Mozambique, one-third of grade 5 pupils and one half of grade 7 pupils did not pass national examinations (Takala, 1998). In Senegal, between 1990 and 1995, average success rates on the end of primary school examination were lower than 30% (ADEA, 2002). In Guinea, only 6% of pupils achieved “mastery” in French and none in mathematics (5% achieved “partial mastery”) in tests that were constructed to reflect the objectives of the curriculum (Carron & Châu, 1996).

2.2. Obstacles to meeting EFA goals

25. It was unfortunate that efforts to implement policies designed to provide Education For All occurred at a time when many countries on the continent were suffering from civil strife, natural disasters, and the HIV/AIDS pandemic. They also occurred as Structural Adjustment Programs (SAP), which made expenditure on education vulnerable, were being implemented. In the case of education, SAP measures typically were aimed at increasing efficiency (reducing unit costs) and cost recovery in second-level education (involving raising fees and encouraging private provision), measures not likely to improve quality in conjunction with wider provision. However, “ultimately, whether or not education contributes to development depends less on how many children complete school than on what they do or do not learn” (Samoff, 1999, p. 199).

26. Given the lack of human and financial resources to carry out the EFA mandate, one would expect the enormous expansion of education that took place in Africa in recent decades to have been accompanied by deterioration in quality (see Bude, 1993). For example, how could Uganda be expected to maintain, much less improve, quality at a time when primary school enrolment was increasing from 2.5 million (in 1996) to 6.5 million (in 1999) and pupil-teacher ratio from 38:1 to 62:1 (ADEA, 1999)? The problem was recognized in Zambia where the National Task Force, *Focus on Learning*, in a follow-up to Jomtien, noted that the “price that has been paid for the rapid quantitative developments of earlier years is a serious deterioration in the quality of school education” (cited in Takala, 1998). However, in a number of countries, as in the original Education For All proposals, little attention seems to have been paid to possible trade-offs in policy between quantitative expansion and quality improvement. As a result, many parents have come to question the value of education for their children as quality dropped sharply, less qualified teachers were employed, classes became larger, facilities poorer, and textbooks in shorter supply (Chapman & Mähle, 1993).

2.3. Recommitment in Dakar to EFA

27. Progress during the 1990s can only be regarded as disappointing. In 2000, in the Dakar Framework for Action, national governments, organizations, and donor agencies recommitted themselves to a revised set of goals relating to Education For All (UNESCO, 2000a). These include:

- ensuring that by 2015 all children, particularly girls, children in difficult circumstances and those belonging to ethnic minorities, have access to and complete, primary education of good quality (goal 3);
- eliminating gender disparities in primary and secondary education by 2005, and achieving gender equality in education by 2015, with a focus on ensuring girls' full and equal access to and achievement in basic education of good quality (goal 5); and
- improving all aspects of the quality of education and ensuring excellence of all so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy and essential life skills (goal 6) (UNESCO, 2002, p. 13).

28. Unlike the Jomtien Declaration, the Dakar Framework specified a set of actions that were designed to instil a sense of urgency and create a climate of accountability (UNESCO, 2002). Policies and practices adopted in pursuit of EFA goals will obviously depend on local circumstances, and so will vary from country to country. However, since all countries are committed to ensuring that recognized and measurable learning outcomes are achieved by all, some form of assessment will be required to determine if this happens.

29. The need to assess student achievements will involve a departure from practice in which, though there have been exceptions, the main focus in assessing the quality of education was on inputs (e.g., student participation rates, physical facilities, curriculum materials, books, and teacher training) (Grisay & Mählck, 1991). Though inputs are not irrelevant in an assessment of quality, assumptions that high quality inputs are invariably associated with high levels of student achievement are not always tenable. Neither are what are often considered low quality inputs always associated with low levels of achievement.

30. Growth in interest in ascertaining what students learn as a result of their school experience is not attributable solely to the need to provide information to monitor the effect of EFA policies. Nor is it confined to developing countries. In recent years, in both the industrialized and developing world, interest in developing a corporatist approach to administration and an associated rise in "managerialism" is evident. The approach draws on ideas from the business world involving strategic and operational planning, continuous improvement, a focus on "deliverables" and results, quality assurance, incentive and accountability systems based on results, and the concept of the citizen as consumer. In this approach, performance is defined in terms of results, performance targets are set, the extent to which results are achieved are determined on the basis of performance indicators, and decisions regarding resource allocation are based on performance information. The purpose of the activities is to provide rapid and continuous feedback on a limited number of outcome measures considered to be of interest to politicians, policy makers, administrators, and stakeholders (Kellaghan, 2003; Kellaghan & Madaus, 2000).

2.4. Focus of the paper

31. The focus of this paper is on the use of assessment to monitor and evaluate learning. Acceptance that there is a need to monitor what students learn gives rise to four more specific needs: (i) the need to describe the knowledge and skills that constitute “quality education”; (ii) the need for technical means and organization capacity to measure student achievement; (iii) the need for technical means to evaluate progress towards the goal of “quality education”; and (iv) the need for expertise in translating assessment data into policy and instructional procedures that will improve the quality of learning (UNESCO, 2000b).

32. The term “assessment” will be used to refer to any procedure or activity that is designed to collect information about the knowledge, attitudes, or skills of a learner or group of learners (Kellaghan & Greaney, 2001b). Four major categories of assessment will be considered: public (external) examinations; national assessments; international assessments; and classroom assessment. It should be noted that each category comprises a variety of practices. However, the categorization is justified on the basis that administrative procedures vary for the categories and the major purposes of assessment in the four categories differ.

33. On the basis that government response to problems in the past was constrained by the fact that little objective information on which to make decisions was available to managers, we shall be interested as we consider each category in determining the extent to which information derived from assessments can aid decision-makers in improving management and, in turn, the quality of the education system (see Chapman & Mählck, 1993). Use of information derived from assessments, as we shall see, may not be limited to management decisions. When assessments have important consequences attached to performance (as is the case when performance on a public examination determines graduation, promotion, or selection for further education or a job), they are likely to impact directly on teaching and learning, and so merit consideration as a mechanism for improving student achievements.

34. The main interest in assessment in the context of Education For All has been in monitoring *the outcomes of the education system* rather than in appraising the achievements of individual students. This interest will be addressed in our description of national and international assessments. However, given the important role that the assessment of *individual students* in public (external) examinations plays in the education systems of Africa, we shall also consider how examinations might be used to shape teachers’ pedagogical practices in desirable ways, and thus improve the quality of students’ learning.

35. Classroom-based assessment has received little attention in proposals regarding the use of assessment to improve students’ learning. This may be because it is quite difficult for policy makers to influence the classroom practices of teachers over a short period of time or without getting involved in what is likely to be expensive inservice provision. It is easier and less expensive to use an existing centrally controlled system of assessment (public examinations) or to mandate an external assessment (in a national assessment). However, since teachers’ assessment of students’ learning in the classroom occupies a central role in the learning process of all students, it merits attention when proposals to improve learning through the use of assessment are being considered.

2.5. Public (external) examinations

36. Public (external) examinations have played a major role throughout the history of modern education in Africa. Most countries operate three major examinations which are administered by an agency outside the school (usually a national examinations authority in anglophone countries and a ministry of education in francophone countries). The first is administered at the end of primary schooling when students are examined in the major subjects of the curriculum (e.g., English or French, a local language, science, and social studies). Performance on the examination usually determines whether or not a student will proceed to secondary school. After two or three years in secondary school, students sit another examination (usually in a wider range of subjects), the results of which will determine who will move into the higher grades of secondary education. At the end of secondary school, a third public examination, the results of which will affect students' further educational and vocational prospects, is administered.

37. Examinations in Africa serve a number of important functions which reflect the social and educational contexts in which they are administered. First, they control the disparate elements of the education system, helping to ensure that all schools teach to the same standards, something that was especially necessary in colonial times when most schools were under private management. Second, they are used to select students in pyramidal education systems in which the number of places diminishes at each successive level. Third, the examinations have a certification function, though this is often lost sight of because of the emphasis on their use for selection. Formal certification of academic achievements, however, can be important for some students in gaining access to employment or training, though lower level certificates are losing their currency in the labor market as the numbers possessing them increase. Finally, public examinations may serve an accountability function for teachers and schools. This will especially be the case when the results of students' performance on examinations are published (Kellaghan, 1992; Kellaghan & Greaney, 1992; Omolewa & Kellaghan, 2003).

38. The continued existence and central importance of public examinations in Africa can be attributed to a number of factors. They are perceived to allocate scarce educational benefits in an objective and unbiased way, though concern has sometimes been expressed that they may discriminate against minorities, rural populations, girls, and students whose first language differs from that of the examination. They provide a specification of clear goals and standards for teachers and students. They can be used to underpin changes in curriculum and teaching methods, and to maintain national standards. Finally, especially at the end of secondary schooling, they legitimate membership in the international global society, and facilitate international mobility.

39. In this section we shall first consider the quality of public examinations in Africa. While there is some variation between systems, and efforts have been made in many countries over the past two decades to improve quality, many examinations are inadequate in a number of respects. Some of the inadequacies are inherent in any terminal pencil-and-paper test. In other cases, improvements could be made while maintaining the terminal pencil-and-paper form. Following a description of the quality of examinations, we shall address the use of data from examinations to describe student achievements. Analysis of student performance can provide insights into the nature of the achievements of students at the end of a course of study (e.g., curriculum areas in which they perform poorly), and how those achievements are distributed by gender and through schools and regions. We shall then consider the rationale behind proposals to use examinations as a lever of reform, as well as the limited empirical evidence relating to the effects of changes in examinations on what is taught in classrooms and on student achievements. We shall also describe a number of mechanisms which are designed to use assessment data to bring about change in school practice and pupil achievements. Following a

consideration of problems that arise when examinations to which high stakes are attached are used to influence teaching and learning, we propose a number of guidelines designed to improve the quality of examinations.

40. It should be noted that our commentary relates to the formal education system. This is not to imply that the non-formal sector is not an important component of educational endeavor in Africa. There are examples of students in this sector being prepared for public examinations. In many systems, there are also large numbers of “external” candidates who are not attending recognized schools. However, these are usually students who have performed less well than they had hoped in an earlier sitting of an examination, and would not be attending courses that would generally be regarded as falling in the non-formal sector. That public examinations are not generally considered relevant to non-formal education is due largely to differences in the goals of formal and non-formal education, differences in characteristics of participants in the two sectors, and a reluctance on the part of many non-formal educators to get involved in formal assessments of students’ achievements. Although most participants in non-formal education will not wish to enter the formal sector, for those who do formal arrangements to establish equivalences between qualifications obtained in the formal and non-formal sectors should be put in place.

2.6. The quality of public examinations in Africa

41. Many studies and official reports have identified inadequacies in current examinations. First, since most examinations, at both primary and secondary level, are limited to pencil and paper tests, they ignore a variety of areas of knowledge and skill, often specified in curricula, which cannot be measured by these tests. Second, examinations place a high degree of emphasis on the achievement of cognitive skills (particularly language and mathematics at the end of primary schooling), and pay very little attention to practical skills. Third, there is evidence that items in many examinations measure achievement at a low taxonomic level (involving the recall or recognition of factual knowledge), rather than the achievements students are going to need in later life, such as the ability to synthesize material or apply knowledge to new situations (to make inferences, to develop a logical sequence of steps to solve a problem, to argue a case). Fourth, many examinations contain very little reference to the everyday life of students outside the school. They deal with scholastic topics and applications for the most part, rather than, for example, trying to find out if a student can use money in the market place or knows how to deal with health problems in the home. Fifth, the quality of questions in examinations is often poor: they are poorly phrased, alternatives in multiple-choice tests are unsatisfactory, or scoring criteria lack clarity (ERGESE, 1986; Kellaghan & Greaney, 1992; Kelly, 1991; Little, 1982; Oxenham, 1983; Somerset, 1996).

42. There are two major implications of this situation. First, issues arise about the validity of the examinations, in particular the extent to which they are biased towards the testing of competencies needed by students continuing their education into the next cycle, and so fail to adequately reflect the goals of curricula for students (a majority in most countries) who will not proceed to secondary education. And secondly, since teachers focus their teaching on what is assessed in an examination which has important consequences attached to performance, serious concerns have been raised about the character and quality of teaching and learning in schools. We shall return to this issue below.

2.7. Improving the quality of public examinations

43. Over the years, various attempts have been made in many countries to improve the quality of examinations. The improvements were often technical, designed to improve efficiency. As student numbers and the problem of processing very substantial numbers of examination scripts grew, many adopted a multiple-choice format. Some used a mixture of multiple-choice and short constructed responses; while some changed to multiple-choice, which was later abandoned in favor of the constructed short-answer format (Bude & Lewin, 1997; Lewin & Dunne, 2000).

44. Other reforms involve much more radical changes in examinations, particularly ones relating to their content. During the 1980s and 1990s, among the many countries that engaged in examination reform were Uganda (1983), Kenya (1985), Zimbabwe (1990), Tanzania (1994) (Bude, 1997), and Namibia (1991-). Following the World Bank (1988) policy paper, *Education in Sub-Saharan Africa*, which recommended “a renewed commitment to academic standards, principally through strengthening examination systems” (p. 93), the first plenary meeting of the Donors to African Education (DAE) (now the Association for Educational Development in Africa – ADEA), reflecting the mood and developments in many countries, addressed the need to improve educational quality. This was followed by the establishment of a Working Group on School Examinations (WGSE) in recognition of the important role examinations could play in quality improvement, which in turn was followed in 1988 and 1989 by studies of examinations in fourteen countries (Kellaghan & Greaney, 1992). The objectives of the studies were to contribute to the improvement of educational quality in a cost-effective way through adjustment of inputs to examination systems, and to help develop institutional capacity for this purpose in Sub-Saharan countries.

45. Among the reforms in examining which have been recommended in various sources are: broadening the scope of examinations to reflect the curriculum of schools; using diverse modes of assessment (written, practical, oral); redesigning examinations to include items which test higher-order thinking skills; assessing students’ ability to apply their knowledge and skills in situations outside the school, as well as in scholastic contexts; shifting testing from a norm-referenced to a more criterion-based approach (that is, measuring pupils’ success or failure in relation to criteria which represent competence rather than with reference to the performance of other examination candidates); incorporating into students’ examination grades information from teachers’ assessment of students; incorporating information into public examinations from pupils’ records and profiles of achievement to reflect the acquisition of competencies that are different in kind from those that can be assessed in a terminal written examination; and ensuring that the examinations differentiate between students on the basis of characteristics relevant to opportunities being offered (Bude & Lewin, 1997; Kellaghan & Greaney, 1992). All these changes in themselves would seem desirable, and should result in more appropriate and valid assessments of students’ achievements.

2.8. Using data from examinations to describe student achievements

46. Given the prevalence of public examinations in Africa, it seems reasonable to ask if the data obtained from them can be used to provide evidence on student achievements that may be useful to policy-makers and teachers. One might also ask if examination data can provide evidence that standards of achievement are improving, remaining the same, or deteriorating. We describe three sources of information about examination performance: chief examiners’ reports, other feedback mechanisms, and analysis to identify categories of pupils who perform poorly.

2.8.1. Chief examiners' reports

47. Analysis of student performance on examinations is a feature of many examination systems. The results are typically presented in chief examiners' reports in which the strengths and weaknesses of a candidature, together with perceived deficiencies in teaching, are described. Reports vary considerably in quality, in the level of detail that they provide, and in the relevance of the information they yield for individual schools and teachers. However, the extent to which the information in these reports is considered by key stakeholders is unclear.

48. Examples of the kind of information provided in examiners' reports, which is not untypical of the kind of information provided in other countries, is to be found in chief examiners' reports in Swaziland. The reports are based on what students wrote in primary and junior certificate examinations. They provide comments on broad areas of strength and weakness in students' achievements; on teacher competence and curriculum implementation in schools; and on how students' language competence affected their performance (*Box 1*).

Box 1. Chief examiners' reports in Swaziland

In reports on the Primary Certificate Examination, the poor levels of examinees' vocabulary and comprehension in English were noted. Problems were also identified in examinees' ability to construct sentences in writing, as well as in their use of grammar, spelling, and handwriting. In reading, there was evidence of poor ability in scanning and in locating relevant information in passages. Pupils' facility in English was relevant not just to their performance in the English examination, but also in other subjects. Thus, the performance of candidates in Science, Social Studies, Agriculture, and Home Economics was considered to have been affected by their inability to understand some words or questions.

In the examiners' reports on the Junior Certificate Examination, the issue of students' proficiency in English was a persistent theme, and was again raised not just in the context of performance on the English examination (on which candidates' poor sentence construction and use of tenses, poor punctuation and spelling, as well as problems in answering questions involving inference, discussion, and symbolism were noted), but in the context of many other examinations. For example, in mathematics, students' failure to understand questions was noted, while performance on the Integrated Science examination was considered to have been affected by students' limited ability in English which led to failure to understand questions and to difficulty in composing responses.

Examiners' reports on the Swazi examinations were not limited to an analysis of candidates' performance. Examiners also formed impressions about curriculum implementation in schools and teaching. For example, they concluded that the curriculum had not been fully covered in the Integrated Science course in some schools, which led to candidates avoiding Physics and Chemistry questions. Similarly, many areas of the Metalwork curriculum (sheet-metal work, design, oblique projection) were considered not to have been taught. In Integrated Science, there was evidence that in some schools students had actually been given information that was factually incorrect. The need for students to have access to a greater range of books and other materials was identified in the review of the Geography examination (Kellaghan, 2002).

49. The role of language in instruction, and, by extension, in examinations, raised in the Swazi report has received considerable attention in recent years in both anglophone and, perhaps even more so, in francophone countries (Box 2). There can be no doubt that many pupils are at a disadvantage when instruction is in a language, often a language of wider communication (LWC), in which they are not proficient (Naumann & Wolf, 2001). At issue is the extent to which language proficiency contributes variance to candidates' scores that is not a function of their competence in the subject being assessed (e.g., science, history). On the basis of chief examiners' reports for Swazi examinations, there is a danger that this was happening in the case of several subjects. Further evidence is available from a study in Burundi, in which students' performances in reading, mathematics, science, and agriculture in tests in French and Kirundi were compared. Students obtained higher scores in Kirundi (except in mathematics). Furthermore, the performance of the most able students was most affected by being tested in French (Eisemon, 1990).

Box 2. Language of instruction

The issue of language is not just one of examining, but of the use of a language of wider communication (LWC) for instruction, especially in subjects which have great practical importance for the majority of primary-school leavers who do not continue their studies. Most African community languages are not used in teaching in many countries, in some countries not at all (Ouane, 2003). The issue is of significance since in the Monitoring Learning Achievement (MLA) project, countries in which instruction was in the mother tongue of students outperformed others in most learning areas (Chinapah et al, 2000).

Since poor proficiency in the language of instruction may be impeding children's learning, a number of countries have introduced experimental programs in which instruction is in children's mother tongue (e.g., Mali, Niger, Nigeria). Preliminary evaluation points to lower rates of repetition and dropout, improved performance on primary-school leaving examinations, and higher transition rates for children in the programs (ADEA, 1999; Bamgbose, 1991). However, since class size in experimental schools in the Niger study was much smaller (22-25:1, compared to 50-90:1 in traditional classes), it is not possible to attribute the undoubted benefits solely to language of instruction.

Reasons for the use of LWC are complex. They may reflect concern about national integration, cost, and international communication (Clayton, 1998). LWC may also have practical advantages in linguistically heterogeneous situations.

2.8.2. Other feedback mechanisms

50. In addition to chief examiners' reports, a number of countries have developed systems of analysis and feedback to schools on how pupils performed in examinations taken at the end of primary schooling. A reform of examinations in Kenya provided "guidance information" based on an analysis of the performance of candidates on individual items in the examination (Kyalo, 1997; Rharade, 1997; Somerset, 1987, 1988, 1996). Concrete advice was also provided on how to address difficulties that candidates exhibited in their responses. The information came in a newsletter sent to all schools and in workshops, seminars, and school visits (particularly to schools with poor results). The Kenya National Examinations Council continues to produce a newsletter, but lack of financial resources precludes it from sending it to all schools. It can, however, be purchased from the Council (J.M. Mwachihi, personal communication, May 2003).

51. Feedback has also been a feature of other examination systems. In Zambia, every school was sent a copy of the Chief Examiner's report. In Lesotho, performance item statistics for the population of examination candidates was sent to all schools (Kellaghan & Greaney, 1992). In Uganda, newsletters were sent to schools in which problem areas in examination performance were identified, and teaching strategies to deal with them suggested.

2.8.3. Identification of categories of pupils who perform poorly

52. Examination results can be used to identify differences in performance between boys and girls, between regions, between locations (urban-rural), and between schools. It is not unusual to find schools in which practically all students are successful in examinations and ones in which practically no students are. Although the reasons for these differences are complex, and implicate sociocultural and student background factors that may not readily be manipulated by political action, nevertheless the provision of guidance, support, and resources to poorly performing schools or districts by educational authorities can lead to improvement. Analyses that reveal gender and regional differences can also provide a basis for a review of examination procedures to ensure they are not biased against any subgroups in candidatures.

2.8.4. Limitations of the data provided by examinations

53. While analysis of examination results can provide insights into the achievements of students in the education system, as well as identifying schools in which student performance is weak and so may require additional attention, there are limitations to the information that they can provide. One limitation is that public examinations usually provide information on students' achievements in only limited areas of a curriculum. This is partly because only a limited number of subjects are examined. It is also partly because even within these subjects, the focus of examinations is on curriculum content and associated competencies that will maximize discrimination between students who will be selected and those who will not. Thus, the achievements of lower performing students may not be adequately represented.

54. The results of examinations are sometimes used to monitor student achievements in the education system over time to determine whether standards are rising, falling, or remaining constant. This use, however, is problematic. First, a different examination is set each year and there usually is no way of knowing whether the difficulty level of the examinations remains constant over time. Secondly, the results reported for examinations are often standardized and norm-referenced, reflecting their selective function, in which case the reported mean is fixed from year to year, and so would not reveal a change in standards. Thirdly, as educational provision continues to expand (as is expected under EFA), and as more students sit for public examinations, the characteristics of examinees will change, which in turn might affect the average level of achievement of examinees. However, it is difficult to assess the impact of increased participation (Greaney & Kellaghan, 1996a; Kellaghan, 1996a).

2.9. Using examinations to improve student achievements

55. Analysis of examination results can provide some useful, if limited, information on the achievements of students in the education system. However, a much more radical and proactive role is assigned to examinations when it is proposed that they be used, not just as a means of obtaining information about education systems, but as a lever of reform. Proposals to use examinations in this way received considerable impetus from the observation that *when high stakes are attached to examination performance* (that is, when performance has important consequences for students, and often for teachers), an alignment will take place between what is taught in schools and the objectives of the examinations. The case for what came to be called “measurement-driven instruction” had been made in the United States by Popham (1983, 1987) who argued that changes in assessment do not have to mirror changes in curriculum. In fact, the assessment changes could come first and these would be expected to drive curriculum implementation in schools. For Popham, it was important that “the achievement tests we now hope will serve as instructional targets contain a reasonable number of assessment foci, so teachers can organize their instructional efforts around a reasonable number of instructional foci” (p. 25). The approach, it will be noted, runs counter to the more traditional view that the content of an assessment should reflect the curriculum, a view, for example, that underlies changes to the national examination and assessment system in Namibia which were considered necessary because of reforms to the curriculum (ADEA, 1999).

56. Several commentators on education in developing countries have seen a value in measurement-driven instruction as a means of changing classroom practice, given the high stakes attached to examination performance. The formula was relatively simple: Make sure that examinations cover important and relevant content and assess important knowledge and skills, and teachers will adjust their teaching accordingly. In this vein, Eisemon, Patel, & Abagi (1987) regarded the strategy of improving instruction through improved examinations as basically sound, while according to Little (1982), “examination improvements could help turn the education system into one which encourages, rather than stultifies, desirable outcomes” (p. 228). Furthermore, changing examinations was not only a powerful means of influencing the quality of teaching and learning in schools, it was also inexpensive (Heyneman & Ransom, 1992).

2.10. Effects of changes in examinations

57. What is the evidence that changing public examinations will improve the quality of students’ learning? We will attempt to answer this question by considering evidence relating to three more specific questions which relate to the objectives of reforms. Will a change in the content areas examined result in a shift in the content to which students are exposed in class? Is a change in examinations likely to result in an improvement in the level of student achievements? And will a change in examinations result in change in students’ cognitive processing skills?

58. Will a change in the content areas examined result in a shift in the content to which students are exposed in class? There is considerable evidence from a variety of countries that changes in examinations will affect the content to which students are exposed in class – the subjects that receive attention, and the topics within subjects that are taught. This happened as a result of changes in chemistry examinations in Australia; in physics, mathematics, and chemistry examinations in Ireland; in the primary school examination in Belgium; in modern languages in the New York Regents examination;

and in mathematics on the College Entrance Examination Board achievement tests in the United States (Madaus & Kellaghan, 1992).

59. In Kenya, the introduction of Kiswahili and practical subjects to the Kenya Certificate of Primary Education in the 1980s was reported to have resulted in a dramatic increase in the coverage of these subjects in schools in spite of a great many difficulties relating to facilities, textbooks, and teacher competence (Eisemon, 1990). Also in the 1980s, the replacement of multiple-choice items on sentence style and structure by an essay-writing component in the Common Entrance Examination (taken at the end of primary schooling) in Trinidad and Tobago had the effect of increasing the amount of writing tasks assigned by teachers during the year to provide students with experience in formulating arguments and applying their knowledge to new problems and issues. Although data in support of the conclusion are not provided, London (1997) reported that “essay writing has now been actively taught in the schools for almost a decade ... most teachers ... express a sense of relief that essay-writing ... is being given its fair share of time within day-to-day classroom exercises” (p. 144).

2.10.1. Is a change in examinations likely to result in an improvement in the level of student achievements?

60. There is little empirical evidence either to support or to challenge the claim that a change in examinations will result in an improvement in the level of student achievements.

61. The most frequently cited study in this context is that of Somerset (1987, 1988) who examined the effects of reforms in the examination administered at the end of the basic education cycle in Kenya in the 1970s. The aims of the reform were (i) to improve the efficiency of the examination as a selection instrument; (ii) to give less privileged pupils (rural, low income, girls) a better chance of showing their abilities and hence gain access to secondary school; (iii) to encourage and assist teachers to provide all pupils (especially those for whom basic education will be terminal) with a more relevant set of cognitive skills; (iv) to improve the overall quality of primary education; and (v) to reduce differences among districts and schools. To achieve these aims, the content of examinations was changed substantially; a school merit order list was published to put pressure on schools; and information on how pupils had performed on examinations was provided in a newsletter.

62. Unfortunately, data are not available that would allow us to determine if the performance of pupils in general improved as a result of the reforms, as scores each year were standardized (to a mean of 50 and a standard deviation of 15). A comparison of raw scores would probably not be very meaningful either. Information on the proportions “passing” might have provided a basis for year-to-year comparisons if one could be sure that criteria did not change over time. However, the data available on performance did permit comparisons over time between the relative performances of candidates in different districts. While the initial impact of the reforms was to widen achievement differences between districts, this trend was reversed after the system had been in operation for four years: nearly all districts in which performance had been relatively poor showed striking gains relative to performance in other districts (Somerset, 1987).

2.10.2. Will a change in examinations result in change in students' cognitive processing skills?

63. A key objective of proponents of reform is to use examinations to develop higher-order cognitive processing skills in students. A problem with current examinations is that they are largely confined to assessing lower-level skills (such as recall and recognition), with the result that these are what are taught in schools. If, however, examinations required students to display higher-order skills, it is argued that teachers would shift the content and methods of their teaching to meet that demand.

64. Evidence on the effect of examinations on the teaching and on students' acquisition of higher-order skills is mixed. While data from Kenya experience on this issue are not available, Eisemon et al's (1987) inability to discern a greater emphasis in primary school instruction on problem-solving, reasoning, and explanation following the introduction of the new types of item would not lead one to expect any great change in students' cognitive processing skills.

65. There is further evidence that teachers do not always react to the demands of examinations, which in turn has implications for the skills that students acquire. In a study carried out in Swaziland (Rollnick, Manyatsi, Lubben, & Bradley, 1998), teachers' approaches to the teaching of chemistry were related to the kinds of questions used on two examinations. In one examination (the Junior Certificate Examination taken at the end of the junior cycle of secondary education), most items were at the level of recall and knowledge, while in the examination which followed (O-level at the end of the senior cycle of secondary schooling), three cognitive levels (recall, knowledge with understanding, handling information) were assessed. However, teaching in preparation for both examinations was found to be basically the same. In both situations, teachers were mainly concerned with the acquisition of factual knowledge, which was reflected in their use of low-level cognitive questions in class and in tests. Given that O-level teachers followed the approach used by teachers at junior secondary level and did not teach and test at higher cognitive levels as required by the examination, it is not surprising that over a third of students failed Science (Physics, Chemistry) in the 2001 O-level examination (Kellaghan, 2002). There was, however, according to Rollnick et al, some evidence that inservice provision could be effective in changing the approaches adopted by teachers.

66. There is further evidence that when guidance is provided to teachers in the preparation of students for examinations, and teacher understanding of the demands of examinations is developed in inservice training, the inclusion of tasks requiring higher-order skills in examinations may lead to greater emphasis on teaching these skills in classrooms. In a study carried out in Standard 8 in Nairobi primary schools, teachers were asked to prepare pupils for two sets of mock examination questions. One set had been prepared by the Kenya National Examinations Council; the other was specially designed to elicit higher-order cognitive skills. In the latter, students were required to make inferences, rather than to recognize the correct answer. The specially designed paper resulted in significant changes in the ways teachers prepared pupils. Furthermore, pupils who had been prepared for this paper performed better than students who had not on both examinations (Eisemon, 1990).

2.10.3. Mechanisms to bring about change in school practice and student achievement

67. Simply changing an examination, the content that is examined, and the kinds of cognitive processing skills that are examined may be sufficient in itself to alter school practice and the knowledge and skills that students acquire. This is most obvious when the content areas of an examination are changed (e.g., when a practical element is introduced or students are required to write an essay as well as to respond to short-answer questions). In these situations, teachers might be expected to respond to the syllabus and the demands of the examination and provide appropriate experiences for students.

68. However, some teachers may require additional motivation to change their practice or may not be able to adjust to change. For example, teachers may not feel adequately prepared to teach practical elements of a science curriculum, or to change their method of teaching to facilitate the development of higher-order thinking skills. In these situations, examination reformers and school authorities have resorted to a number of procedures designed to induce change in teacher behavior. Some of these procedures rely on introducing or reinforcing competition between schools; others on providing information to schools; yet again, others provide additional guidance and/or resources.

2.10.4. Competition between schools

69. In some examination systems, an attempt is made to influence teacher behavior simply by making public how students in individual schools performed on examinations. The publication of mean performance statistics for each school and for each district in “league tables,” making it possible for schools to see where they stood with respect to other schools in a district, and for districts to compare themselves with other districts, was a key feature of the Kenya examination reform, in which this kind of information was called “incentive information” (Somerset, 1987). The underlying idea was that dissemination of information about school performance would create competition between schools which would motivate teachers to change their instructional practices (Chapman & Synder, 2000). Furthermore, it was considered that such information might serve to create a “market” situation in which parents “shopped around” for schools, armed with “evidence” about the performance of schools in situations in which a choice of school existed.

70. There can be little doubt that assessment data published in league tables that show how well or how poorly schools are doing can impact on the behavior of schools. In Senegal, in the 1990s, a results-oriented management system, in which information on the performance of schools was published in the press, was introduced. Between 1995 and 1998, the success rate on the examination at the end of primary school rose from 30% to 48%. Furthermore, the enrolment rate of girls rose from 40% to 77% (ADEA, 2002). It is not possible, however, to attribute these improvements solely to publication of results since other reforms were implemented at the same time (job descriptions were drawn up; school inspections were increased; seminars and open school days were held).

71. Despite the use of league tables in Kenya (now discontinued), Senegal, and elsewhere, several factors indicate that their use is more complicated than might appear at first sight, is far from free of problems, and may even be misleading (Box 3). Apart from the issues listed in Box 3, individuals wishing to assess the contribution a school might make to the development of students should bear in mind that student achievements are dependent on a variety of influences other than those brought to bear by the school. If one

is interested in a school's contribution to student achievements, then the problem of how to make adequate allowance for factors outside the control of the school has to be addressed. Thus, it is reasonable to ask if students performed poorly because they were not motivated or had little home support, or because the school was inadequately resourced, or because teachers taught badly. Furthermore, if students differ from school to school in their level of achievements when entering a school, a measure of achievement at a later date that does not take this into account will be inequitable and misleading in that it will not adequately reflect a school's success in moving students from their initial entry level to their present level of achievement as reflected in a public examination (Kellaghan & Greaney, 2001b).

2.10.5. Provision of guidance to schools

72. Students' performance in an examination may be analysed, and the results of the analysis provided to schools. As we have already seen, such information may be provided in traditional chief examiners' reports, and, more recently, has been provided in newsletters to schools. The provision of what was called "guidance information" in which key topics and skills which were causing pupils difficulty, together with concrete advice to teachers on how to deal with them, was a key feature of the newsletter in the Kenya examination reforms (Somerset, 1987, 1988). According to Somerset, the competitive aspect of the reforms was not enough; there was also a need to provide support and guidance so that teachers could reorient their efforts.

Box 3. Problems with league tables

- The performance of schools (and thus their position relative to the performance of other schools) may vary depending on the school outcome that is used. Thus, school rankings will differ depending on whether they are based on public examination results, on performance in basic curriculum areas, on an assessment by school inspectors, or on an overall measure of school effectiveness.
- Even rankings based on the same assessment can vary depending on the criterion of "success" that is used. Thus, a ranking based on the proportion of students who obtain high grades (honors or distinction) may differ from a ranking based on the proportion of students who achieve a more moderate but acceptable level of performance (e.g., who pass).
- Lack of precision in even well-constructed assessment instruments means that small differences between schools in scores on outcomes will be due to chance. However, a difference of a few points on an outcome score could result in a very large difference in a school ranking in a large school system.
- Whatever outcome measure is used, results can vary from year to year, as has been shown in Britain, where among the "top" 250 schools, the rank of a school could change by as much as 123 places from one year to the next. This might be due to differences in measuring instruments, in cohorts of students, or between teachers.
- Competition cannot operate when only one school is available in an area.

2.10.6. Support for underperforming schools

73. In addition to the provision of information about performance, "underperforming" schools may be identified, the reason for poor performance investigated, and assistance provided. These efforts will involve attempts to carefully explain the meaning of results (with reference, for example, to achievement objectives

and the nature of student errors), which in turn will lead to the provision of practical advice regarding skills to be developed, sequences to be implemented, and types of exercises to be set. The relevance of the advice to the objectives being pursued should be visible in concrete terms. Efforts are more likely to be successful if teachers can hope for a direct “return” in the form of improved performance on the next examination (Grisay & Mählch, 1991).

74. If it is concluded that poor performance is due to lack of textbooks, then the matter may be relatively straightforward. However, low achievement usually stems from more complex problems (Chapman & Snyder, 2000) which may require complex interventions. In recognition of this, district education offices in Kenya, on the basis of results, advised schools on pedagogy and other issues (Somerset, 1988).

75. Even when resources are provided, their use to improve student achievement may still be problematic for some teachers. If teachers have not had resources in the past, they will have adapted their teaching to that situation, and it cannot be assumed that when resources become available, they will automatically change their teaching methods to make optimal use of them.

76. A further issue in dealing with low performing schools is that they are often in the most inaccessible locations, serving minority populations who have little influence on national politics (Chapman & Snyder, 2000).

2.10.7. Linkage with curriculum

77. At a broader level, the examining authority, inspectors of the ministry of education, and staff at curriculum centers may discuss results and their implications for curriculum development and implementation, teaching, and the preparation of textbooks, as happened in Kenya (Somerset, 1988). Such discussion reflects an appreciation of the need for a close relationship between an examining authority and those responsible for curricula (Eisemon, 1990).

2.10.8. Particular problems when high stakes are attached to assessment information

78. We have seen that proponents of examination reform have argued that if the content coverage of examinations is extended and the taxonomic level of the cognitive processes required to respond to items raised, these reforms should be reflected in how and what students are taught in school. Echoing this position, Somerset (1996) has said that “there is no intrinsic reason why the backwash effects of examinations on teaching and learning should be harmful; they are harmful only when the examinations are of inadequate quality” (p. 282). This view, while recognizing that the power of examinations to change school practice resides in the fact that high stakes are attached to performance, fails to take into account something else that has long been recognized: that whatever the quality of examinations, negative, if unintended, consequences can follow if sanctions are attached to performance (Kellaghan & Greaney, 1992; Madaus, 1988; Madaus & Greaney, 1985; Madaus & Kellaghan, 1992). These consequences are outlined in this section (see *Box 4*).

Box 4. How teachers are influenced by examinations

The extent to which examinations dominate teaching can be seen in the description of the not untypical behavior of teachers in Burundi. Among the characteristics of teaching that were noted were: an increase in instructional time beyond that laid down for primary schools (e.g., on Saturday afternoons); regular review of material that was considered essential; frequent testing to identify performance deficits and to improve test-taking skills; diagnosis of sources of pupil errors followed by a selection of appropriate instructional methods; and the use of class periods allocated to prevocational skills which were not examined (e.g., in agriculture) to teach the academic subjects which were examined (Eisemon, 1990). Many of these features have been observed in examination systems in other countries when examination results are an important determinant of future education and life chances, or are used to evaluate teachers (Kellaghan & Greaney, 1992).

79. While the alignment of teaching with examinations, in which teachers mirror the foci of examinations in their instruction, is proposed as one of the advantages of high-stakes examinations, the downside of this is that only a subset or sample of an entire achievement domain is assessed in an examination. By focusing on what is examined, subjects that are not examined will receive less attention in schools than their place in the curriculum might suggest they merit, while within subjects, teachers will strive to increase the overlap between instructional and test content, leading to a narrowing of the curriculum and to a situation in which the examination becomes the manifest definition of the domain (Le Mahieu & Leinhardt, 1985).

80. A further disadvantage of examinations to which high stakes are attached is that they tend to affect teaching strategies, learning strategies, students' involvement in learning tasks, and their attitudes to learning. Teachers will tend to rely on drill, and may expect students to engage in learning strategies that are superficial or short-term (such as memorizing, rehearsing, and rote learning). Students will respond accordingly, and may focus on social comparisons (doing better than others), something that may be reinforced by the common practice of clapping when students provide an acceptable response in class. It has been found that when high stakes are attached to performance, students tend to be less successful in acquiring higher-order and transferable skills; learning tasks are perceived as not inherently interesting and, if a reward is removed, students will be less likely to engage in a task (Kellaghan, Madaus, & Raczek, 1996). Thus, some of the basic goals of education reform may be frustrated.

81. A further problem that is associated with high-stakes examinations is that considerable effort and time will be put into drill-dominated test preparation by teachers and students. Thus, time is lost to instruction (e.g., in sitting "mock examinations"). This can amount to as much as a week per month in the year students take examinations in Guinea and Mali. Test preparation activities can also distort teaching and learning. For example, if an examination is made up of selection-type questions, normal teaching may take the form of statements accompanied by a range of options from which students are required to select a response. The focus on test-preparation skills may also serve to make students direct their efforts towards mastering strategies to help them over the examination hurdle, rather than towards developing mastery of subject matter and honing lasting competencies.

82. A further worrying aspect of high-stakes examinations is that their efforts may not be limited to the educational experience of students at the grade at which examinations are administered, though they may be most intense at that level. At lower grades, the subjects in which examinations are taken are likely to receive greater emphasis, probably at the expense of other curriculum goals. Even the format of examinations may affect teaching. For example, the multiple-choice format has been observed not only in classroom tests, but in the teaching methodology of teachers in the early grades of primary school.

83. At a more general level, the use of high stakes tests limits the ability to make inferences about improvements in student achievements. We should not be surprised to find that students' scores on the examination will rise when teachers focus on content that is limited to the scope of an examination. However, the increase will be in areas in which curriculum objectives overlap with test content (Le Mahieu & Leinhardt, 1985). In this situation, goal displacement occurs since preparation for examinations diverts efforts from attaining curriculum objectives to meeting the requirements of the examination. Its effect is to distort the inferences that can be made from actual test performance with the result that high performance or an increase in scores cannot be interpreted as evidence of high achievement in the broader domain of achievement envisaged in the curriculum (and which presumably an examination should measure). Thus, the examination may be invalidated as a vehicle for assessing overall achievement (see also Le Mahieu, 1984; Linn, 1983, 2000).

84. Also at a general level, high stakes may be associated with malpractice. In their efforts to obtain high scores, students (and sometimes teachers and others) may resort to various forms of cheating, designed to give a candidate an unfair advantage over others (Greaney & Kellaghan, 1996b). Malpractice takes many forms: copying from other students during an examination, collusion between students and examination supervisors, use of material smuggled into the examination hall, bribing or intimidation of examiners, and purchasing of examination papers (see Box 5). The pressure to comply with malpractice can be great. A case has been reported in which invigilators were shot when they refused to co-operate. While the extent of malpractice in examinations is not known, it is probably considerable. In a recent primary certificate examination in Uganda, the results of 3.16% of candidates were cancelled on grounds of malpractice. Given that malpractice can erode the credibility of an examination system, a variety of practices to prevent it (e.g., having examination papers printed outside the country, keeping markers in isolation while scoring papers), to detect it (e.g., matching the response patterns of students who sit close to each other in an examination), and to punish it (e.g., by a prison sentence) are in place in many countries.

Box 5. Some forms of examination corruption

Prior to examination

Leakage of paper content by officials, paper setters, moderators, or school administrators; improper assignment of candidates by officials to targeted centers.

During examination administration

Impersonation; external assistance (e.g., from helpers, cellular phones); smuggling of material (e.g., in garments); copying and collusion among candidates; intimidation of supervisory staff (e.g., by candidates, external helpers, government agencies, politicians, journalists, and teacher unions); substitution of scripts; use of "ghost" (non-existent) centers.

After examination

Substitution of scripts; intimidation of markers aided by corrupt officials; markers seeking reward from candidates' parents; collusion between the candidate and the marker (e.g., candidates give home address); falsification of data files and results sheets; issuance of fake diplomas.

85. Finally, we should be aware of how high-stakes examinations can conflict with aspects of educational policy. One way in which this can happen is if teachers, because their reputations depend on how well their pupils perform in examinations, focus their efforts on pupils who are most likely to succeed. If this happens, then it is likely to inhibit attainment of the Education For All goal that all pupils complete basic education of good quality.

86. A focus on achieving a high success rate on examinations may also lead to high rates of repetition and dropout, which are associated with high expenditure and low levels of internal efficiency (see Eisemon, 1997; N'tchougan-Sonou, 2001. In Kenya, the low transition rate between standards 6 and 7 was partly explained by the fact that schools discouraged weaker pupils from taking the Kenya Certificate in Primary Education examination for fear it would lower mean scores in published league tables (Ackers, Migoli, & Nzomo, 2001). Pupils may also be retained in the belief that repetition enhances pupils' chances of doing well in examinations, thereby improving their chances of going to secondary school. Repetition of course, is not limited to Kenya and, indeed, is a particular feature of francophone countries (Burkina Faso, Burundi, Cameroon, Chad, Comoros, Guinea, Guinea-Bissau, Niger, Rwanda, Senegal). Government policies of automatic promotion (which incidentally exist in Kenya) are likely to be inhibited by high-stakes examinations.

2.10.9. Guidelines to improve public examinations

87. Disadvantages will continue to be associated with public examinations as long as high stakes are attached to performance. However, it is unlikely that there will be any diminution in their role or importance in most countries until education systems can accommodate many more students. In this situation, the goal should be to reduce negative effects associated with poor quality instruments, even if the negative effects associated with high stakes assessment cannot be entirely eliminated. A number of guidelines, which if followed should serve to improve the quality of public examinations, are set out in *Box 6*.

Box 6. Guidelines to improve public examinations

- Assessments should reflect the curriculum, and since not everything in a curriculum can be assessed in an examination, the areas that are assessed should be ones that are considered important. If the expectation is that student achievement will align itself with the content of examinations, it is critically important that the stated objectives and content be carefully developed. Objectives should reflect the contribution of knowledge and skill which they embody to the long-term growth and development of students (Le Mahieu, 1984).
- Insofar as possible, modes of assessment (e.g., written, practical, oral) should be diverse to reflect the goals of curricula.
- Examination items should not be limited to the measurement of recall or recognition of information, but should attempt to measure higher-order outcomes defined in terms of more complex cognitive processes (e.g., understanding, synthesis, application).
- Examinations should assess students' ability to apply their knowledge, not just in scholastic contexts, but also in situations outside school.
- Examinations, both in content and in difficulty level, should reflect their certification function and provide an adequate assessment of the competencies of pupils who will not obtain a higher-level qualification.
- Examination performance should be analyzed to provide feedback to schools and other stakeholders (curriculum authorities, etc.).
- As many teachers as possible should be involved in the setting and scoring of examinations, as these activities provide valuable insights into the demands of examinations which can be applied in teaching.
- The use of teachers' assessments to contribute to grades that their students are awarded in public examinations merits investigation.
- The content and form of examination items should be free of gender, ethnic, and location biases. Thus, for example, the language used should not exhibit gender bias (e.g., in the use of names), and topics that are included should be of interest to both girls and boys and reflect situations that would be equally familiar to both genders. It should be noted that there is evidence that the multiple-choice format tends to favor boys

(Kellaghan, & Greaney, 1992).

3. NATIONAL ASSESSMENTS

88. While public examinations are a long-standing feature of education systems, national assessments (sometimes called system assessments, assessments of learning outcomes, and less appropriately learning assessments) are relatively new (Greaney & Kellaghan, 1996a; Kellaghan, 2003; Kellaghan & Greaney, 2001a,b). A national assessment may be defined as an exercise designed to describe the level of achievements, not of individual students, but of a whole education system, or a clearly defined part of it (e.g., fourth grade pupils or 11-year olds). The centerpiece of the assessment is the collection of data in schools. Usually students respond to assessment instruments and questionnaires in groups. Teachers may also be requested to complete questionnaires in which they provide information considered relevant to an interpretation of their students' achievements. The main elements of a national assessment are listed in *Box 7*.

Box 7. Main elements of a national assessment

- The Ministry of Education (MOE) appoints an implementing agency either within the ministry or an independent external body (e.g., a university department or a research organization) and provides funding.
- Policy needs to be addressed in the assessment are determined by the ministry, sometimes in consultation with key educational stakeholders (e.g., teachers' representatives, curriculum specialists, business people, parents).
- The MOE, or a steering committee nominated by it, identifies the population to be assessed (e.g., fourth grade students).
- The domain of achievement to be assessed is determined (e.g., reading, mathematics).
- The implementing agency prepares achievement tests and supporting questionnaires and administration manuals.
- The tests and supporting documents are pilot-tested and subsequently reviewed to determine curricular and technical adequacy.
- The implementing agency selects the targeted sample (or population of schools/students), arranges for printing of materials, and establishes communication with selected schools.
- Test administrators (classroom teachers, school inspectors, or graduate university students) are trained by the implementing agency.
- Survey instruments (tests and questionnaires) are administered in schools on a specified date.
- Survey instruments are collected, returned to the implementing agency, cleaned, and prepared for analysis.
- Analyses are carried out, and a report prepared.

(Kellaghan & Greaney, 2001b, p. 35).

89. National assessments were introduced in realization of the fact that the educational data on inputs to education which had typically been collected in the past were often of little relevance or use to educational planners (Kudjoh & Mingat, 1993). National assessments would address this issue by providing information on the "products" or "outcomes" of schooling (e.g., student achievements, inequalities in the system) which, it was hoped, could be used in conjunction with input data to provide a sounder basis for policy and decision-making. Thus, national assessments would provide

policy and decision-makers with relevant and reliable information about the state of the education system, its achievements, and its problems, which would be amenable to analysis and interpretation. The kind of information they would provide would differ from that available from other sources. It would differ from that available from public examinations which, while providing data on outcomes, do so only for those who take examinations. Information from national assessments would also differ from that provided by research and education sector studies, which are generally short-term, while national assessments hold out the possibility, if integrated into the overall management and administration of the education system, of providing information relevant to education sector analysis on a continuing basis.

90. The longest running and best known national assessments are in the United Kingdom, which has operated a system in one form or another since 1948, the USA, in which an assessment was first conducted in 1969, and France, which introduced a system in 1979. Also of long standing is the national assessment in Chile, which was first administered in 1978. These systems are all quite elaborate and collect information on a frequent basis. Assessment systems in most other countries are more recent and less elaborate (Kellaghan & Greaney, 2001b).

91. Developments in national assessment did not occur in less economically developed countries until the 1990s. Most seem attributable to the Declaration of the World Conference on Education For All which stressed that it was important to know to what extent students actually learned as a result of educational opportunities provided (the outcomes of education) and to have better information as an input to more effective planning and management. Since such knowledge was not currently available, a need for national assessments, and development of the capacity to carry them out, was indicated.

3.1. Information sought in national assessments

92. All national assessments seek answers to one or more of the following questions.

- How well are students learning in the education system (with reference, for example, to general expectations, EFA goals, the aims of the curriculum, or preparation for life)?
- Is there evidence of particular strengths or weaknesses in the knowledge and skills students have acquired?
- Do the achievements of subgroups in the population differ? Are there, for example, disparities between the achievements of boys and girls, of students in urban and rural locations, of students from different language or ethnic groups, of students in different regions of the country, or students who drop out early or are repeating grades?
- To what extent is achievement associated with the characteristics of the learning environment (e.g., school resources, teacher preparation and competence, type of school) or with students' home and community circumstances?
- Do the achievements of students change over time? This can be particularly important at a time of major change in the system (e.g., when numbers are increasing; when new subjects or curricula are being implemented). (Kellaghan, 2003; Kellaghan & Greaney, 2001b)

3.1.1. Uses of national assessment information

93. The data provided by a national assessment have been used for a variety of purposes. These include:

- to bring to the notice of politicians, policy makers, and the public the need for more effective education (and perhaps additional resources) to promote social and economic development;
- to justify the reallocation of discretionary resources; it should be realized, however, that there may be little flexibility in how resources are allocated (e.g., when most resources go to teachers' salaries);
- to support policy decisions (e.g., regarding grade repetition);
- to constrain "bad" policy decisions (e.g., withdrawal of funding for inservice education), which often focus on one dimension of policy, by placing issues in a broader context;
- to suggest more efficient resource allocation, prompted by the findings of analyses that identify relationships between alternative inputs and student achievements;
- to improve management efficiency through increased accountability (and in some cases competition);
- to use the backwash effect of the assessment (as has been proposed in the use of public examinations) to ensure that teachers teach certain subject matter. This is only possible when all schools participate in an assessment and when sanctions are attached to performance (e.g., in Chile; see Himmel, 1996; Schiefelbein, 1993).

3.2. Variation in national assessments

94. While all national assessments have as their objective the description of students' achievements in the education system, there is a good deal of variation in how they go about this (see *Box 8*). Different approaches have implications for the way in what information derived from an assessment can be used. For example, it would not make sense to attempt to attach high stakes to a school's performance (e.g., by publishing results) if the assessment was based on a sample of schools. Again, it would not make sense to use data from an assessment to make judgments about the performance of individual students or teachers if each individual student took only a fraction of a test (as low as 1/9 in some assessments).

3.3. National assessment activity in Africa

95. Four major categories of national assessment can be identified in Africa. Three involve similar activities in several countries: the Monitoring Learning Achievement (MLA) project; the Southern Africa Consortium for Monitoring Educational Quality (SACMEQ) project; Programme d'Analyse des Systèmes Educatifs des Pays de la CONFEMEN (PASEC). In a fourth category, national assessments which were not related to any of these projects were carried out in individual countries.

Box 8. Differences in national assessment procedures

- Most assessments are based on a sample of schools/students; in some (e.g., Chile, England, France), all or most students at a targeted age or grade are assessed.
- Each student may take only a fraction of a large number of assessment tasks, allowing for extensive curriculum coverage without requiring students to spend a lot of time responding to tasks (e.g., Brazil, Ireland, U.S.A.), or all students may respond to the same tasks (e.g., Uganda, Zambia).
- An assessment may or may not be designed to provide information about individual schools (or even individual teachers and students). When information is available about individual schools, an assessment may become a high-stakes operation like a public examination (e.g., in England).
- Assessments differ in the complexity of the way in which students' performance is described. Results may be reported as simple percentage correct scores or students' scores may be scaled using Item Response Modelling.
- Assessments differ in the detail in which they describe performance (e.g., by subdomains within the broader domain of mathematics or numeracy) and in the extent to which they attempt to establish relationships (some of which may be interpreted causally) between student achievement and a variety of factors (e.g., school type, community characteristics).
- Assessments differ in the extent to which the data obtained from them are integrated into other aspects of the education system.

96. MLA, which began in 1992, is a joint UNESCO/UNICEF initiative, and is part of the EFA assessment which has as its objective constructing a comprehensive picture of the progress of countries towards EFA goals (Chinapah, 1997). In particular, it is a response to the need to monitor the extent to which students actually acquire useful knowledge, reasoning ability, skills, and values (Article 4 of the *World Declaration on Education for All, 1990*).

97. In MLA I, the achievements of grade 4 pupils were assessed in literacy (reading/writing), numeracy, and life skills (relating to awareness and knowledge of health, nutrition, sanitation, hygiene). In MLA II, grade 8 pupils' achievements were assessed in mathematics and science; common tests were used in all countries. In both MLA I and II data were collected on students' backgrounds, school characteristics, and family backgrounds.

98. To date, MLA assessments have been carried out in more than 70 countries, 47 of which are in Africa (UNESCO, 2003a). Forty African countries participated in MLA I and 11 in MLA II. In Nigeria, results were presented by state. A regional approach was adopted in implementation to capitalize on local expertise and to develop capacity in participating countries. By March 2003, reports had been published on the MLA I assessments of 18 Sub-Saharan countries, and on the MLA II assessments of two.

99. In addition to national reports, a separate report on MLA I for 11 countries has been prepared (Botswana, Madagascar, Malawi, Mali, Morocco, Mauritius, Niger, Senegal, Tunisia, Uganda, Zambia) (Chinapah et al, 2000). Only four of these countries had met their Jomtien learning target (i.e., 80% of learners should attain the defined learning competencies) for fourth grade pupils in life skills; two in literacy; and none in numeracy. Gender differences were small in all countries. With the exception of

Mauritius, pupils in private schools performed better than students in public schools in all three learning areas. The ability of parents to assist learners in doing school work was related to student achievement in most countries.

100. SACMEQ is a collaborative voluntary grouping of 15 ministries of education in Southern and Eastern Africa, working in close collaboration with the International Institute for Educational Planning (IIEP) in Paris to build institutional capacities through joint training to carry out co-operative educational policy research (Ross et al, 2000; UNESCO, 2003b). Between 1995 and 1998, eight education ministries collected information in SACMEQ I on baseline indicators for educational inputs, general conditions of schooling, equity assessments for human and material resource allocations, and the literacy levels of grade 6 students. Teachers as well as students were tested, except in Mauritius and South Africa. Fifteen countries participated in SACMEQ II between 1999 and 2002: Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania (mainland), Tanzania (Zanzibar), Uganda, Zambia, Zimbabwe. Seven national reports have been published, and 14 are in preparation.

101. A major aim of SACMEQ is to promote capacity building by equipping educational planners in member countries with the technical skills needed to effectively monitor and evaluate schooling and the quality of education. In providing valid and accessible information systems as a basis for decision-making, it also seeks to promote stakeholder involvement and greater transparency. A particular feature of the approach is its learning-by-doing training for educational planners by involving them in the conduct of studies.

102. As in the case of MLA, the results of SACMEQ assessments indicated that education systems were failing to meet the performance standards of ministries. For example, less than 30% of sixth grade pupils met specified minimum literacy standards in Namibia and Zambia (UNESCO, 2000a). Although over 50% met “minimum” standards in Zimbabwe, the figure did not change over time (Machingaidze, Pfukani, & Shumba, n.d.). Significant achievement differences were usually found between regions and between types of school within countries. Gender differences were not significant (Mioko, 1998).

103. PASEC, which was established in 1991 as a response of francophone countries to the Jomtien conference, acts as a network for sharing information on educational evaluation instruments and results. It encourages the involvement of senior decision-makers and of other stakeholders in the identification of policy issues, and emphasizes the need to base decisions on reliable data, followed by a realistic agenda for action (including time-frames and cost estimates) (Kulpoo & Coustère, 1999). Initially, pupils in grades 2 and 5 were assessed in French and mathematics. Now pupils are assessed in all grades from 2 through 6. Since 1994, the same tests have been used in all countries. Data are also collected from pupils and teachers on a variety of school and background factors. Twelve countries have participated in PASEC: Burkina Faso, Cameroun, Congo (Brazzaville), Côte d’Ivoire, Djibouti, Guinea, Madagascar, Mali, Niger, République Centrafricaine, Senegal, Togo. It has been estimated that PASEC and SACMEQ cost about US\$50,000 per country (Kulpoo & Coustère, 1999).

104. PASEC, differs from other national assessments in Africa in that in some countries pupils are assessed near the beginning (November) and end (May) of the academic year. Thus, it is possible to obtain some indication of growth or of the “value” added during the course of the year, though this is possible only for pupils who survive in the system. The approach reflects the strong research orientation of the program which was in evidence at its inception, and attempts to identify “causal relationships” on which to base policy designed to improve the efficiency and effectiveness of education. Efforts

have been made to determine the “impact” of inschool factors (teacher training, class size, textbook availability) as well as of nonschool factors (parental education, distance to school, home language).

105. In addition to these three programs to support national assessment, several countries have carried out national assessments which vary in their technical adequacy (e.g., Burundi, Eritrea, Mali, Namibia, Senegal, Uganda, Zambia). In Namibia, the National Learner Baseline Assessment measured student achievement in English and mathematics at grades 4 and 7 (Namibia Ministry of Education and Culture/Florida State University/Harvard University, 1994). Reading and listening comprehension were assessed in Oshindonga in a subset of northern schools. The assessment was designed to help policymakers allocate resources to underachieving schools. The results suggested that the expected level of competence in English was too high, and that curriculum materials might need to be revised.

106. Eritrea’s Department of General Education (1999), as part of the EFA 2000 assessment, tested pupils in six regions to determine if they had mastered the basic skills as laid down in the official curriculum in their mother tongue (in grade 1) and in English and mathematics (in grade 4). The assessment identified aspects of the curriculum that were causing particular problems (e.g., place value, word problems in mathematics); found that boys generally outperformed girls; and identified implications for teacher education and teaching methodologies.

107. All the national assessments in Africa share a number of features. All involve education planners and managers, and all are designed to provide information for policy making. This was not true of PASEC in its early stages when it was perceived primarily as a research/technical activity. However, in time, ministry officials became interested and involved. All the assessments view capacity building and the strengthening of the policy-maker/researcher nexus as major objectives. All are sample-based. All adopt an input-output model of the education system and attempt to identify factors associated with achievement.

3.4. Impact of national assessments

108. Impact has been reported by assessment teams or close associates for all the assessment programs in a number of areas. In policy debate and formulation, SACMEQ results have featured in presidential and national commissions in Zimbabwe and Namibia; prime ministerial and cabinet reviews of educational policy in Zanzibar; national education sector studies in Zambia; and reviews of national education master plans in Mauritius (Ross et al, 2000). MLA results were used in preparation of education reform programs in Malawi and Uganda, and PASEC results in country analysis reports in Burkina Faso and Niger.

109. The experience provided by the programs was also considered to have contributed to capacity building. Following MLA, a number of countries carried out their own assessments (Morocco, Mauritius, Mali, Madagascar). In Kenya and Zimbabwe, data cleaning methods used in SACMEQ were adapted for school census data. In PASEC countries, improved capacity in test construction and in the design and execution of surveys was observed.

110. Information from national assessments regarding grade repetition, class size, and the availability of textbooks was used to support policy in PASEC countries. To what extent this was justified, however, is another matter. Reanalyses of data for Senegal attribute less importance to textbook availability than was attributed in original analyses (Naumann & Wolf, 2001).

111. Information from national assessments has also given rise to national debates. For example, in Mauritius, SACMEQ data were used in a debate on strategies to mitigate the damaging role of the Certificate of Primary Education and to support an improvement in the pupil-to-book ratio. In the Seychelles, SACMEQ results initiated a debate on streaming.

112. In a number of countries, SACMEQ results caused education managers to reappraise the adequacy of the resources provided for education. Kenya introduced benchmarks for the provision of facilities (desks per pupil; books per pupil). In Zimbabwe, repairs to school buildings followed, special funds were provided for classroom supplies, and training programs were initiated on the management of school libraries.

113. As a result of experience with MLA, some education systems are paying more attention to life skills in curricula. However, there is little reported evidence of an impact on school practice.

114. Assessment results have also been used as a rationale for major donor support to address identified system weaknesses. In Guinea, for example, the results of the PASEC assessment prompted government and the World Bank to develop a program to promote instruction in local languages in the early grades, following which a gradual transition to French would take place.

115. The effects of national assessments should not, however, be over-estimated. In Mauritius, an external review team reported that relatively little action had followed policy suggestions made in three separate national assessments. In response to a series of questions on the use of national assessment results as part of the present study, senior education personnel in six countries (Ethiopia, Malawi, Niger, Nigeria, South Africa, and Uganda) indicated that findings were sometimes covered in the media. Results did not feature in parliamentary debate in any country. Results were used in only one country to justify granting additional resources to schools. In four countries, results were shared with curriculum authorities, but feedback to teachers/schools was provided in only two, and feedback to textbook publishers in one. Respondents in two countries (Ethiopia, Nigeria) said that national assessment results had not been used in the formulation of educational policy.

3.5. Issues in national assessments

116. A decision to carry out a national assessment gives rise to a number of issues. Some need to be resolved before the actual assessment can begin; others relate to use and interpretation of data. The issues are identified in this section, and comments are provided on the basis of international experience.

- *Competence domain.* A decision has to be made about the competence domain that will be assessed. The domains most commonly assessed in assessments throughout the world are pupils' language of instruction and mathematics, and less frequently science and social studies. An issue that merits consideration is whether the competencies assessed are based on the prescribed curriculum or on an appraisal of the knowledge and skills students will need when they leave school.
- *Population.* Some national assessments involve pupils at one or two grades in primary school, others may involve pupils at all grades. Assessments at secondary school are less common but do exist in a number of countries.
- *All students or a sample of students.* In some national assessment systems (e.g., Chile), all (or practically all) students in the relevant grades are assessed. Thus, information is available about all schools and teachers, as well as about all students.

This procedure is used when action at the school (or teacher) level is planned, such as the publication of results or the provision of support and resources to poorly performing schools. Selection of a sample of students for an assessment, which is considerably less expensive than testing all pupils in the system, is more common. However, the information obtained in such an assessment will be useful in guiding policy formation and system-wide initiatives, not intervention in individual schools.

- *Comparisons between regions or districts.* Many countries are interested in comparing the performance of students in regions or districts. Often, however, samples are too small to allow meaningful comparisons.
- *The same test taken by all students or the test taken by each individual student to contain only a fraction of items.* In several national and international assessments, students respond to only a fraction of the total number of items used in an assessment so as to increase coverage of the curriculum, or to allow for extended passages in the assessment of reading comprehension, without placing too great a burden on individual students. In other assessments, all students respond to the same set of items. While there are advantages associated with having students respond to only a fraction of items, there are also disadvantages, particularly for countries beginning a national assessment program. Administration is more complex, as is scoring and the scaling of scores, while analyses involving individual student or school data are problematic.
- *Standards.* The establishment of “standards of competence”, “proficiency levels”, or “mastery levels” (“minimal” or “desirable”), even when assessment instruments are described as criterion-referenced, is more problematic than much use of the terms would imply. There is no obvious basis for deciding on the point at which a standard can be said to be satisfied, or for saying the students who score just above a dividing point differ substantially in their achievements from students who score just below it. Since the actual setting of standards is usually based on the judgments of individuals, which will always to some extent be arbitrary (see Cizek, 2001), and since the procedures involved can vary, the method used in standard setting should always be described. There is also a need to consider how criteria of competence in different assessment systems are related. Different standard-setting procedures usually produce different cut-points, even if each has a rational basis. Data from public examinations and national assessments may reveal considerable discrepancies. Concern has been expressed in Senegal over inconsistencies between students’ scores on PASEC tests and teachers’ promotional practices. Some students who did relatively well on PASEC were not promoted, while others who did relatively poorly were (B. Moll, personal communication, May 2003). There would also appear to be differences between the judgments made on the basis of performance on public examinations (Kellaghan & Farrell, 1998) and judgments based on performance in national assessments. In Lesotho, where four out of five pupils passed the primary certificate examination, less than one in six scored at the minimum level of mastery in a national assessment of literacy. In Malawi, close to four out of five pupils passed the primary certificate examination, but only one in five achieved minimum mastery in a national assessment. In Uganda, about 70% passed the certificate examination, but only about a third achieved minimum mastery in a national assessment. It should be noted that the figures for the examinations and national assessments are not based on the same cohorts of students. However, the discrepancies are so large that it is unlikely that they do not represent real differences in the standards applied in public examinations and national assessments.
- *Monitoring over time.* If student achievements are to be monitored over time, a number of conditions have to be met: the assessment instruments should be unchanged (or, if changed, it is necessary to establish that instruments are equivalent); probability samples of adequate size (if the total population is not

assessed) should be employed; and the definition of “standards” should be transparent. It should be noted that the equivalence of instruments cannot always be maintained (e.g., if a curriculum is changed to include content that is more relevant to students’ lives, or to improve levels of understanding, analysis, and application).

- *Interpreting relationships.* There is a tendency in some reports of national assessments to interpret relationships between inputs to the education system (e.g., school or student characteristics) and student achievements causally. However, causal relationships on the basis of cross-sectional survey data are rarely justified. Furthermore, the fact that the numbers of schools/pupils in some categories (e.g., in districts, in rural areas) are small and that methods of analysis are inappropriate may mean that estimates are unreliable.
- *Capacity building.* Experience to date suggests that retaining trained competent personnel in national assessment activity is likely to be difficult. Such individuals are often promoted to more senior posts.
- *Stakeholder involvement.* In some education systems, it may be difficult to secure support from the broader stakeholder community (e.g., teacher educators, teacher unions, minority groups) in overseeing and implementing a national assessment. However, opposition from a stakeholder group can undermine an assessment.

3.5.1. Conditions and practices required to develop and institutionalize the capacity to carry out a national assessment

117. A number of conditions and practices are required to carry out and institutionalize a national assessment (Chapman & Mählck, 1993; Kellaghan & Greaney, 2001b; Marope, 1999; Samoff, 1999).

- The need for information (including information about student achievements), and the need to develop the capacity to provide it, should be recognized. Those who are in a position to make decisions about a national assessment, and to provide the resources required to carry it out (e.g., a minister or permanent secretary), must believe that the information will be relevant and useful in identifying problems and in informing policy and practice. Unless they understand that a capacity gap exists, and are prepared to address it, efforts to improve capacity to carry out a national assessment and to maintain the activity will fail.
- The knowledge and skills required to design, manage, and interpret a national assessment must be available. Training (supported by adequate funding) may be required to achieve this.
- An adequate level of support by stakeholders, especially policy makers, is required. Signs of a lack of institutional commitment include moving an individual who has undertaken training for national assessment activity to another job, or failure to provide adequate time or resources to individuals carrying out an assessment.
- A national assessment should be integrated into policy environments, existing structures, policy and decision-making processes, and channels of resource allocation. It should not be seen as a project or separate activity that will cease when the assessment has been carried out. This should help ensure that the information from a national assessment reaches those who have a role to play in formulating and implementing policy, and does not suffer the fate of some sector analysis studies, the findings of which often do not reach key government personnel.
- Every effort should be made to ensure that national assessments are aligned with other instructional guidance mechanisms in the education system, not only with other

assessment systems (public examinations, classroom assessment), but with curricula, teacher education, school capacity building, and measures to address inequities.

- Educational Management Information Systems (EMIS), where they exist and are operating, should be adapted to include data on the quality of student achievements, as well as on inputs. This would help institutionalize assessment data and locate it in an environment in which it will come to the attention of policy makers and managers.
- Education authorities should establish systems and strategies of communicating information to individuals outside the ministry, especially teachers, who will have a role in implementing policy based on national assessment data.
- Proposals regarding the use of national assessment data should be made in the context of a realistic understanding of the role of quantitative data in planning and management, since resource allocation and planning always have to take cognizance of a variety of social, economic, and political factors.

4. INTERNATIONAL ASSESSMENTS

118. International assessments share many procedural features with national assessments, although they differ from them most obviously in the fact that they have to be designed to allow administration in more than one country (Beaton et al, 1999; Greaney & Kellaghan, 1996a; Kellaghan & Greaney, 2001b).

119. As in national assessments, instruments are developed in international assessments to assess students' knowledge and skills. However, instead of representing the curriculum of only one education system, the instruments have to be considered appropriate for use in all participating systems. The age or grade at which the instruments are to be administered has to be agreed, as have procedures for selecting schools/students. International studies have all been based on samples of students.

120. Since the 1960s, over 60 countries have participated in international studies of achievement in reading, mathematics, science, writing, literature, foreign languages, civic education, and computer literacy. The studies give some indication of where students in a country stand relative to students in other countries. They also show the extent to which the treatment of common curriculum areas differs across countries, which may lead a country to reassess its curriculum policy. Indeed, the main impact of international studies seems to have been on curriculum (see, e.g., Burnstein, 1993; Kellaghan, 1996b; Romberg, 1999).

4.1. International assessments in Africa

121. Few African countries have participated in major international studies. Ghana, Nigeria, and Zimbabwe participated in the IEA science study in 1983-1984 (Postlethwaite & Wiley, 1992). No African country participated in the IEA reading literacy study; one (South Africa) participated in the Third International Mathematics and Science Study (TIMSS), three (Morocco, South Africa, Tunisia) in TIMSS-R, and one (Morocco) in the Progress in International Literacy Study (PIRLS). This is not surprising since the studies were designed for industrialized countries. Participation by African countries in international studies that are pitched to conditions and standards in the industrialized world would seem to be of little value.

122. Students in South Africa performed poorly in both TIMSS and TIMSS-R. In TIMSS-R, the mean score recorded in mathematics at grade 8 was lower than the mean of the least proficient pupils in almost all participating countries. The South African results were used to compare performance by province (Howie, 2001) and to relate achievement scores to background variables that included home language, socioeconomic status, race, language of instruction, and student attitudes (Howie, 2002).

123. While MLA, PASEC, and SACMEQ were designed as national assessments, results have been reported in a way that permits international comparisons. However, to allow valid comparisons, instruments, target populations, sampling, and analyses would have to be identical in all countries. It is not clear that this was the case. For example, MLA involved the construction of "country specific instruments" (Chinapah et al, 2000, p. 4). If instruments varied from country to country, interpretation of the mean scores for 11 countries (presented in Chinapah et al, 2000, p. 82) is problematic.

4.2. Problems associated with international assessments

124. A number of problems have been identified in international studies (Kellaghan & Greaney, 2001b). These exist whether the studies are carried out in Africa or elsewhere, though some problems are more likely to arise in developing than in industrialized countries. The first problem that can be identified is that it is difficult to design an assessment procedure that will adequately measure the outcomes of a variety of curricula. Although there are common elements in curricula across the world, particularly at the primary school level, there are also considerable differences between industrialized and developing countries in what is taught and in expected standards of achievement. The greater the difference between the curricula and levels of achievement of countries participating in an international assessment, the more difficult it is to devise an assessment procedure that will suit all countries.

125. A second problem arises if it is necessary to translate instruments into one or more languages. This problem can also arise, of course, within countries. In either case, if comparisons are to be made between performances assessed in different languages, it should be realized that differences that may emerge may be attributable to language-related differences in the difficulty of assessment tasks. It is very difficult to ensure that the way questions are phrased and the cultural appropriateness of content are equivalent in all language versions of an assessment task.

126. A third problem associated with international assessments relates to the equivalence across countries of the populations and samples of students that are being compared. This problem is most obvious where retention rates differ from one country to another, and so again is particularly relevant in studies in which industrialized and developing countries participate. There may also, however, be more subtle differences related to population and sampling when, for example, students in special education programs or with learning difficulties are excluded in one country, but not in another.

127. A fourth problem arises when the primary focus in reporting the results of an international assessment is on the ranking of countries in terms of the average scores of their students, usually the main interest of media. However, rankings can be misleading when the statistical significance of mean differences in achievement is ignored. Besides, rankings in themselves tell us nothing about the many factors that may underlie differences between countries in performance (see Mislevy, 1995).

128. Finally, while it might be argued that an examination of relationships between class-room inputs and student achievement in some countries may be relevant in other countries, one cannot assume that practices associated with high achievement in one country will show a similar relationship in another. Relationships between inputs, processes, and outcomes need to be examined in the context of individual countries (Chapman & Mählck, 1993).

129. Although there are many problems associated with international studies (see *Box 9*), some countries may still be interested in participating in them. Certainly, there is much to be said for the experience in test construction, sampling, analysis, and report writing that an international comparative study can provide. Development and analytic costs may also be less since they are shared. Furthermore, in South Africa, experience with international assessment was considered to have provided benefits in the form of increased private sector contributions to education, baseline information on standards in mathematics and science, and capacity building in educational research. However, it was not clear that the results of the assessment led to significant policy shifts or contributed to

the development of sustainable national capacity in educational measurement since members of the TIMSS and TIMSS-R teams moved to other positions (Howie, 1999).

Box 9. South Africa's experience with international assessments

South Africa's experience with TIMSS and TIMSS-R underlines the problems facing implementers of international assessments. Howie (1999) noted that deadlines imposed by organizers can be difficult, if not impossible, to meet in situations where there may be no mail or telephone services or funds for travel to schools. Other problems include lack of accurate population data on schools; poor management skills; insufficient attention to detail, especially in editing, coding, and data capture; lack of funding to support project workers; and difficulty in securing quality printing on time. Instructions to test administrators, for example to walk up and down the aisle, are obviously inappropriate when classrooms do not have an aisle.

130. If countries are interested in obtaining cross-country comparative data, there is much to be said for limiting participation to countries in which levels of economic and social development are similar. It should also be noted that many of the benefits of international co-operation can be obtained without going down the road of collecting international comparative data.

5. CLASSROOM ASSESSMENT

131. The assessment of students' learning in the classroom (both by teachers and by students themselves) is an integral component of the teaching-learning process. Much of this kind of assessment is subjective, informal, immediate, on-going, and intuitive, as it interacts with learning as it occurs, monitoring student behavior, scholastic performance, and responsiveness to instruction. In addition to ongoing teacher observation, it involves classroom questioning and dialogue, the marking of homework, and the use of portfolios. Its function is primarily formative. It occurs during learning (rather than when learning is presumed to be complete) and is designed to assist or improve students' acquisition of knowledge and skills. Its role is to determine students' current level of knowledge, skill, or understanding, to diagnose problems they may be encountering, to make decisions about the next instructional steps to take (to revise or to move on), and to evaluate the learning that has taken place in a lesson (see *Box 10*).

Box 10. Classroom Assessment

An official British publication reminds us of some of the basic roles of assessment in teaching and learning:

[Assessment] can provide a framework in which educational objectives may be set, and pupils' progress charted and expressed. It can yield a basis for planning the next educational steps in response to children's needs. By facilitating dialogue between teachers, it can enhance professional skills and help the school as a whole to strengthen learning across the curriculum ... (Great Britain. Department of Education and Science and Welsh Office, 1988).

132. A strength of teachers' assessment practices is its focus on "performance". This has several advantages: it does not decontextualize knowledge and skills; it provides evidence of student learning in authentic settings; it allows an assessment of students' ability to think critically, to co-operate, to solve problems, to communicate; and it can contribute substantially to advancing students' learning and understanding.

133. Classroom assessment may on occasion be more formal, as when teachers administer a quiz or end-of-term examination. Such assessment is more objective, and will have a summative function when, for example, the information derived from it is used to make a decision about the retention of a student in a grade or promotion to the next grade. The information may also be reported to pupils, to parents, and to other teachers and individuals who may need to know about a student's progress. Further, it may be used to evaluate the appropriateness of curriculum, methodologies, classroom organization, and textbooks, as well as to provide guidance in matching the curriculum to students' needs and abilities.

5.1. The quality of classroom assessment

134. Despite its central role in the teaching-learning process, we do not know a great deal about how teachers assess their students. There is evidence, however, that the quality of those practices may be deficient in many ways. Problems that have been identified include the use of poorly focused questions, a predominance of questions that require short answers involving factual knowledge, the evocation of responses that involve repetition rather than reflection, and a lack of procedures designed to develop students' higher-order cognitive skills (Black & Wiliam, 1998; Madaus & Kellaghan, 1992).

135. Observations of practice in African classrooms do not present a very different picture. While they should not be interpreted as providing a description that is applicable to all schools in Africa, it is implicit in many of them, if not actually stated, that they apply to many schools.

136. One set of observations relates to questioning, which is an important element in assessment. In a Kenyan study, it was observed that in many lessons pupils were asked no questions at all. When questions were asked, they were closed – a form of questioning which is not likely to facilitate the development of higher-order thinking skills. Further, there was little by way of assessment of pupils' understanding before moving on to the next part of a lesson (Ackers et al, 2001).

137. In Swaziland, the vast majority of questions in higher secondary classes were described as either rhetorical or at a low cognitive level (Rollnick et al, 1998). The low taxonomic level of questioning has also been noted in primary classes in Tanzania where questions were described as requiring pupils to recall facts, which they did individually or in chorus (O-saki & Agu, 2002).

138. Homework provides the opportunity for teachers to assess students' proficiency as well as to provide feedback on problem students may exhibit. However, little homework was given in schools observed in Tanzania. The reasons given were that textbooks or exercise books were not available in rural schools, and that large class sizes in urban schools made it difficult to correct individual students' work. Some teachers rarely marked pupils' exercise books, losing the opportunity to provide feedback. In fact, only a few exercise books contained teacher comments that might have provided reinforcement of good work or identified problems in poor work (O-saki & Agu, 2002).

139. Several commentators attribute the assessment procedures that are prevalent in schools to the nature of the teaching-learning situation, which is invariably described as one in which the teacher is dominant and pupils are passive. Teachers have been described as talking all the time throughout a lesson, leaving little room for pupil activities (Bude, 1993). In a study involving classroom observation and interviews in primary schools in Tanzania, teachers were described as standing in front of the class and teaching pupils with expositions (O-saki & Agu, 2002). A classroom interaction study of 102 lessons in Kenyan schools also found that teachers adopted a "transmission" approach, which was very similar for mathematics, English, and science, and in which there was very little opportunity for pupils to become actively engaged (Ackers et al, 2001). Lessons in science in upper secondary school in Swaziland have also been characterized as teacher-centered, with the teacher asking questions, and pupils answering in chorus or individually (Rollnick et al, 1998).

140. Other explanations that have been offered for existing assessment practices include poorly qualified teachers, large class sizes, poor facilities, and a shortage of learning materials (including books) and of a place to store them. For example, teachers in Guinea were considered to be poorly trained in assessment techniques, the reality being "far from the continuous evaluation procedures recommended by official programmes" (Carron & Châu, 1996).

141. The negative influence of public examinations on teaching and assessment has also been noted. Eisemon et al (1987) observed that pupils were taught through drill, recitation, and exercises of a fill-in-the-missing-word type, all of which were designed to impart the factual information and techniques that they would need in a public examination. The use of "mock" examination papers (based on public examinations) was also considered to limit the scope of instruction to what teachers thought would be examined, leading to incoherence in lessons. Although repetition and dropout are

complex in terms of their causes and effects (see N'tchougan-Sonou, 2001, it is likely that teachers' assessment practices, as well as public examination, contribute to them.

5.2. Improving classroom assessment

142. If the observations on classroom assessment that we have outlined represent general practice, or even a significant amount of practice, in schools in Africa, then a sea-change, not only in assessment practices but also in teachers' general approach to instruction, will be required if assessment is to contribute significantly to the improvement of student learning. This change will require teachers to increase their efforts to seek the active participation of students in their learning, to involve them more in the instructional-learning process, and insist that they share responsibility for their own learning. At a more fundamental level, teachers may need to appreciate that learning is more than improved examinations results, more than the acquisition of information, and that learners should not only acquire, but also generate, master, develop, and create knowledge (Samoff, 1999). Anything that can move education systems – preservice, inservice, or reformed public examinations – towards these goals is to be welcomed.

143. With some notable exceptions in francophone countries, classroom assessment has received little attention in reforms that propose the use of assessment to improve student learning. That, as we have already observed, may be because it is difficult and expensive to do. However, given its central role in the teaching-learning process, and the disadvantages associated with high-stakes public examinations, improving classroom assessment practices should be accorded high priority in any reform that has as its objective the improvement of students' learning (see *Box 11*). Furthermore, approaches to improving classroom assessment should be co-ordinated with reforms in other assessment systems.

144. Given the complexity of classroom assessment, and evidence relating to teachers' skills and practice in this area, there is an obvious need for development of an infrastructure to support improvement of its quality. The infrastructure would involve all stakeholders (all who wield influence on classroom teaching and assessment practices, such as school inspectors, curriculum developers, those who set examinations, and authors of textbooks), preservice and inservice training for teachers, as well as possibly the provision of assessment materials to schools (which might include tests and item banks). For example, in South Africa, Assessment Resource Banks (ARB), which comprise a set of tasks designed to assess specific learning outcomes, are provided to schools in areas serving pupils from low socioeconomic communities. Use of the ARBs involves keeping adequate records of pupils' progress in individual learner profiles and the provision of regular reports to school principals, who in turn provide regular reports to district education offices (Kanjee, 2003). In Swaziland, materials provided to schools include general information about classroom assessment, item banks, tests, item specifications, and remediation and enrichment materials (Mazibuko & Magagula, 2003).

145. The preparation of students to assess their pupils should be an integral part of a teacher education program. Students should be introduced to concepts underlying assessment, and should learn about the use, interpretation, and appropriateness of a range of formal and informal techniques designed to assess pupils' progress and to diagnose difficulties they may be encountering. They should also learn about strategies to address pupils' difficulties. Courses should be provided in the context of close co-operation between college staff in psychology, education, subject methods, and teaching practice.

Box 11. Suggestions to improve classroom assessment

The following are a number of suggestions designed to improve teachers' classroom procedures. They may also be used to inform teacher education courses.

- Assessment should be an integral and frequent aspect of teaching, in which questions that focus on meaningful aspects of learning are used.
- Teachers should develop reasonable, but challenging, expectations for all pupils, using a variety of methods in a variety of situations (e.g., essays, homework, and projects).
- The focus in assessment should be on diagnostic and formative aspects, rather than normative aspects (i.e., assessments that rank students on the basis of results).
- Teachers should ask questions that allow students display higher-order thinking skills (not just recall) and that require inferential and deductive reasoning.
- Pupils' understanding of the general principles of a curriculum domain should be assessed, as well as their ability to use appropriate methods and strategies in solving problems.
- Readily understood and prompt feedback should be provided to students.
- Students' processes (how they approach/analyse issues), not just products, should be assessed.
- Assessment should help students reflect on their own learning.
- Questions should require students to explore/expand on issues, not just repeat information.
- The results of assessments, when appropriate, should be communicated to parents and other interested parties (e.g., other teachers).
- The use of criterion-referenced tests can enrich teachers' classroom assessment practice. Periodic administration (every few weeks) of the tests will provide information on what students have learned, when there is a need for further teaching, and identify students in need of additional help.

5.3. The use of school-based assessment in public examinations

146. So far, we have been considering classroom assessment in the context of day-to-day teaching in which information derived from assessments is used primarily for formative purposes. We now consider the use of assessment data generated in the classroom for summative purposes (usually referred to as school-based assessment), in which the data contribute to the grade students are awarded in an external public examination.

147. Among the criticisms made of public examinations are that they are limited in the time-frame in which they are administered, in the knowledge and skills that are assessed, and in the techniques that are used. Any of these limitations could result in students not being given the opportunity to demonstrate their true level of competence.

148. To address this situation, several examination systems in Africa have introduced, or are planning to introduce, an element of school-based assessment to their public examinations. Indeed, some would hope ultimately, especially at the primary-school level, to replace external examining completely with school-based assessment, perceiving that to be the only way in which the range of competencies specified in curricula can be validly assessed, and in which the negative effects of external examinations on teaching and learning can be removed. However, such a step is problematic for a number of reasons, not least of which is the fact that it may not be perceived to provide an “objective” means of selection in a situation in which only some pupils can be accommodated in secondary schools.

149. Reasons have been advanced both in favor of and against the use of the results of school-based assessment to determine student grades in what is primarily an external examination (Bude, 1997; Heyneman, 1988; Kellaghan & Greaney, 1992; Pennyquick, 1990a, 1990b; Wasanga, 1997). The following reasons have been advanced in support of the use of school-based assessment. First, since school-based assessment is carried out over time and by a person who knows students well, it is likely to provide a more valid and reliable appraisal of a student’s achievements than is possible in a single external terminal examination. Second, school-based assessment permits an extension of the range of curriculum topics which are examined. Aspects of achievement that cannot be satisfactorily assessed in a terminal examination include a student’s ability to plan and organize a project and persevere with it over time. While the assessment of oral and practical skills may be carried out in an external examination, inevitably it will be limited, artificial, and expensive.

150. Third, school-based assessment reduces the undesirable backwash effects of external examinations, since grades are not determined solely on students’ performance on the examination. Fourth, school-based assessment can make allowance for an untypically poor performance of a student on an external examination, due for example to illness. Fifth, school-based assessment, if spread over the year, can increase the level of pupil motivation and application throughout the year.

151. Several reasons have also been advanced why school-based assessment may not be appropriate in examinations when results are used for certification or selection. First, the use of school-based assessment for certification or selection can change the nature of the relationship between teachers and students from an essentially supportive and collaborative one to a judicial one. Second, the competence in assessment of many teachers is considered to be poor, or even if it is not, teachers often feel that they do not know exactly how to translate their informal judgments into more formal and public ones, in which case they may fall back on using poorly constructed test items or the multiple-choice items contained in some textbooks.

152. Third, the standards used to grade students in school-based assessment are likely to vary, both between schools and between classes within schools. Teachers tend to assess their pupils with reference to other pupils in their class, rather than with reference to pupils in other schools. However, public examination results have to be based on comparisons that are valid across schools. To address this issue, school-based assessment results may be “moderated” or scaled against written examination results (e.g., in Namibia). This, in effect, privileges the external assessment by making the school-based results conform to the standards and distributions displayed in the written examination.

153. Fourth, school-based assessment can subject teachers to considerable parental pressure, particularly in small and closely-knit communities. Fifth, school-based assessment requires teachers to devote considerable time to assessment and record-keeping. It also may be considered by teachers to involve too much work. Sixth, school-based assessment gives rise to a variety of administrative problems for schools, such as

what to do when students are absent for tests, or when students transfer from one school to another. Seventh, teachers' assessments are subject to a variety of biases relating to students' gender, socioeconomic background, and personality characteristics. Finally, it is difficult, in some cases impossible, to apply school-based assessment to non-school-based candidates.

154. It is hardly surprising in light of these observations that the implementation of school-based assessment as a component of public examinations has proved problematic in several countries (e.g., Lesotho, Namibia, Nigeria, Swaziland, Tanzania, Uganda). While the aspiration and motivation to introduce it have been high, practical difficulties have on more than one occasion resulted in failure, postponement, or the limitation of the school-based element to a minimal, almost token, amount. For example, the school-based element contributes 5% to students' grades on a public examination in Swaziland (though it is hoped to increase it to 50%) (Kellaghan, 2002). However, in Namibia, the school-based element contributes between a third to a half, depending on the subject (V d Merwe, 1999).

155. The privileged status accorded to external assessment is not limited to the way school-based assessments are moderated against performance in the external component of an examination. Because of problems associated with teachers' competence in assessment and the interpretation of their assessments, there has been a tendency in countries which have incorporated school-based assessments into their public examinations to focus on the more formal aspects of assessment in schools rather than on the informal aspects. Schools may be supplied with written tests or item banks which they administer to their pupils in a formal test situation, which essentially mimics the external examination under the guise of school-based assessment. The only advantage associated with this practice would seem to be that pupils are being given more than one occasion on which to display their competence in a written examination. However, the approach ignores the primary purpose in incorporating judgments based on school-based assessment into public examination grades, which is to provide the opportunity for individuals (teachers) who know students well to assess their competence in a variety of situations, using a variety of modes, over an extended period of time, in areas of the curriculum that cannot be assessed in a written examination.

6. FACTORS THAT MAY INHIBIT THE USE OF INFORMATION DERIVED FROM ASSESSMENTS AND EXAMINATIONS IN CLASSROOMS

156. There is much to commend in efforts to use data from examinations and national assessments at the macro level to provide feedback on student achievements to policymakers, planners, managers, curriculum developers, textbook writers, and the general public. However, there is also a need to move from macro to micro level in the use of the information that is generated. Although it may be widely agreed “that educational reforms live or die by the success of their implementation at the school level” (Verspoor, 1992, p. 23), educational plans invariably fail to pay adequate attention to implementation issues. This may be due, at least in part, to the fact that getting information to teachers, and effecting changes in their behavior, poses many challenges. Expectations that information from assessments and examinations (in publication of results, newsletters, workshops) will radically alter the culture of schools and substantially raise the achievements of all students need to be tempered by a consideration of the factors that may serve to frustrate the intentions of reformers. Unless these factors are recognized and addressed, policies involving assessment may be very limited in their impact on the quality of student learning. The factors discussed in this section are relevant to a consideration of reform proposals based either on public examinations or national assessments.

6.1. Mode of intervention

157. The most common mode of intervention is based on administrative structures in which decisions regarding the input required to bring about desired behaviors in schools are centrally made. This approach can work when problems are well structured and amenable to relatively standardized solutions. It is less likely to be appropriate when problems are not clearly defined and when their solution will require a multiplicity of responses depending on local circumstances. This is the case when problems relate to instructional objectives, equity, and quality, in which case more indirect and interactive strategies will be required. Such strategies typically rely on enabling and framework-setting legislative and administrative measures together with financial and professional support targeted to support desired local actions (Verspoor, 1989, 1992).

6.1.1. Relevance of information

158. There are at least two inherent limitations in the use of data derived from examinations and national assessments to guide the practice of individual teachers. First, the data, for example in chief examiners’ reports or in national assessment reports, relate to general standards of performance in the education system. Thus, they may or may not be relevant to a consideration of conditions and student achievement in any individual school. Secondly, items in examinations and assessments, and the data derived from them are limited in the extent to which they can provide insights into the knowledge structures or cognitive strategies that they call on. Further, since items are not designed to be individually diagnostic, analyses based on results will not provide much in the way of instructionally relevant information about individual students. Thus, the extent to which they can provide specific guidance to teachers will be limited.

6.1.2. Teacher competence

159. Assumptions that all teachers teach in a coherent and organized way (e.g., that they focus on the instructional targets of an examination) and are capable of change, and that conditions in classrooms allow teachers to teach in a way that will meet the objectives of reform, may be based on an unrealistic perception of classrooms.

160. How is one to expect adaptation in classrooms, Carron and Châu (1996) ask, in which “one finds a chaotic reality: teachers and pupils are often absent, the teacher does not have the competence required to teach well, he/she does not follow a precise work plan, merely reads from the textbooks, does not use the blackboard” (p. 202). Lack of teacher competence has been observed by other commentators. Many teachers may have a poor command of the language that is the medium of instruction and poor knowledge of subject matter (especially science and agriculture). Close to half of South African mathematics teachers do not possess a formal qualification in mathematics (Arnott & Kubeka, 1997). In the MLA regional study, fewer than one in four teachers in Malawi, Madagascar, and Mauritius had any post-secondary education (Chinapah et al, 2000). As already noted, a Chief Examiner’s report on the Junior Certificate Integrated Science Examination in Swaziland said that there was evidence in examinees’ scripts that in some schools students had actually been given information that was factually incorrect. These observations may relate to extreme and exceptional cases, but the problems described are precisely the ones that will need attention in any reform effort.

6.1.3. Teachers’ understanding of implications of changes

161. The gulf between what is required in reforms and what teachers understand and actually do can be very large. It cannot be assumed that teachers will know what is really involved in reform proposals, or what they should do on the basis of assessment results. Even when provided with inservice and guidance, the meanings attributed to reform by policy-makers and by teachers may be disparate, since guidance is likely to be interpreted in the context of teachers’ existing understandings and practice relating to curriculum and instruction, which may be very different from those of reformers (Chapman & Snyder, 2000; Grant, Peterson, & Shojgreen-Downer, 1996).

162. In Uganda, for example, teachers had difficulty in understanding the implications of changes in examinations for day-to-day classroom practices and could not adjust to the changes. They were afraid to take chances with new instructional strategies, and so did not change (Snyder et al, 1997).

163. Similar findings are reported from Trinidad and Tobago. A study involving classroom observation and interviews with teachers following changes in public examinations revealed that some teachers that emphasized memorization and group recitation did not know how to change their teaching to develop student competence in problem solving. In this situation, they struggled to understand the implications of the test changes for their day-to-day classroom practice, as did students to understand what, if any, implications the changes might have for them (see *Box 12*).

Box 12. Teachers' difficulties in adjusting to changes in examinations

The main finding of a study in Trinidad and Tobago was that teachers could not necessarily adjust to the examination changes. Even if they understood the examination requirements at a cognitive level, they were often unable to make the necessary changes in the classroom to improve their students' performance. Moreover, experimenting with new pedagogical approaches was threatening to teachers, who felt under enormous pressure to look good and show good local test results ... For many teachers, the route to improved examination scores seemed too uncertain and taking chances with new teaching techniques seemed far too risky.

(Chapman & Snyder, 2000, p. 362)

6.1.4. Complexity of teaching

164. It hardly needs to be said that teaching is an extremely complex activity in which teachers continually interact with a large number of learners who vary in their levels of achievement, aptitude, motivation, and interest. The rules that govern this interaction are not simple, linear, or obvious (see *Box 13*).

Box 13. The complexity of teaching

The key components of classroom instruction are not distinct, static behaviors, but rather dynamic patterns of behavior that follow the ebb and flow of classroom activities and atmosphere. The connection between changing tests to changing instruction is not a technical relationship in which a shift in test format, content, or item complexity necessarily signals any particular change in that dynamic process. Moreover, changes in individual teacher behaviors do not necessarily translate into changes in student learning.

(Chapman & Snyder, 2000, pp. 462-463)

165. Furthermore, teachers may have to employ a multiplicity of strategies in addressing problems associated with poor achievement. Even for students who register the same level of achievement on a test, the nature of that achievement, the reasons for it, and the precise learning difficulties that are implicated can vary very much from student to student, and require different kinds of treatment (see, e.g., Buly & Valencia, 2002).

166. Particular problems arise when examinations, as test designers are increasingly striving to do, place emphasis on measuring higher-order skills rather than, or in addition to, general minimum lower-level skills. Evidence is available to indicate that teaching higher-order skills differs in many ways from teaching lower-level ones. High-level skills cannot be taught by rote methods; they take longer to teach, develop gradually over time, are less amenable to direct instructional approaches, are often difficult to locate in a curriculum, and may be too diffuse to drive instruction (Airasian, 1988).

167. This does not mean that higher-order skills cannot be taught. There is evidence from a study in Nigeria, that when primary school pupils were taught social studies using a problem-approach method, not only did they acquire more facts, they also

comprehended material better, and were able to apply their knowledge to the solution of new problems and to evaluation activities (Ogundare, 1988).

6.1.5. Classroom context

168. The fact that many teachers teach in very large classes with limited resources (textbooks and other aids) is an important consideration when reforms that involve change in teaching style are being proposed.

169. Because of a shortage of textbooks, students may seldom study from printed texts. Teachers may orally summarize the texts, transforming passages into class drills and assignments to be copied into pupils' exercise books (Eisemon et al, 1987). In Kenya, technical subjects (crafts, domestic) were added to general subjects in the primary school curriculum in 1983 with the general objective of facilitating the teaching of techniques, knowledge, and skills that would be useful for economic growth, enabling pupils to proceed to vocational training or start their own business. However, by 1990, only 3,000 of the 13,000 required laboratories had been built (Rharade, 1997).

6.1.6. Opposition based on the perception that a change will involve a risk to pupils.

170. Parents and teachers may perceive changes in classroom practice, and in assessment procedures in particular, as threatening the advantage which pupils have under the existing system (Chapman & Snyder, 2000). Though supportive of steps that might improve the quality of education, they may be more concerned to maintain the comparative advantage which their children enjoy under existing conditions, fearing that they will do less well in an alternative system, or will suffer while teachers are adapting to it.

7. CONCLUSION

171. Improvement of the quality of education in developing countries has been a key feature of reform proposals of governments and donor agencies over the past decade. While the term quality may be applied to many aspects of education, outcomes or student achievements (the acquisition of “useful knowledge, reasoning ability, skills, and values”) were assigned a central position in judgments about quality following the Declaration on Education For All at Jomtien in 1990.

172. Four areas of assessment, all of which have figured in reform proposals, though not to the same extent, were described in this paper: public examinations, national assessments, international assessments, and classroom assessment. Two of these (public examinations and classroom assessment) are long-standing and important features of education systems in Africa. Both also have assessment of the achievements of individual students as their focus. National and international assessments, on the other hand, appeared on the educational scene in Africa only in the last decade. By contrast also with public examinations and classroom assessment, their primary focus is the assessment of the achievements of the education system, not of individual students.

173. There are a number of implications of differences between assessments which are relevant to a consideration of the role they might play in improving the quality of education. First, the use of public examinations and classroom assessment as vehicles of reform will primarily involve the modification of existing practices, while the use of national (and international) assessments will require new procedures. Secondly, the likelihood of impacting on the learning environment of the classroom and on the achievements of individual students is higher in the case of public examinations and classroom assessment than in the case of national (and international) assessments, since the influence of the latter is more remote and will need to be mediated through policy and managerial decisions. Thirdly, classroom assessment differs from the other forms in that it is essentially “private”. Most of the information derived from it remains within the walls of the classroom, though some may become more public when it is conveyed to parents or, in some education systems, when teachers’ assessments contribute to the grade that students are awarded in a public examination. Finally, although classroom assessment may have the greatest potential to influence students’ achievements, it is also the most difficult to alter by an outside agency.

174. While the primary purpose of a system or national assessment is to describe students’ achievements, the role envisaged by reformers who propose using assessment to improve educational quality is not limited to description. In the case of national assessment, it is envisaged that information about strengths and weaknesses in the knowledge and skills students have acquired, and about how achievement is distributed in the population (e.g., by gender, location), would be used to inform policy and decision-making (relating, for example, to resource allocation). It is also envisaged that a national assessment would provide information that is relevant to curriculum developers, textbook writers, politicians, and indeed the general public. Following the communication of information, assessment could become a lever of reform as well as a means of simply describing conditions.

175. It is public examinations, however, rather than other forms of assessment, that have received the most attention in proposals to use assessment as a lever of reform to improve student achievements. It has long been known that the content and form of public examinations impact immediately and directly on what is taught and learned in schools when important consequences are attached to performance for students (and

teachers). On this basis, several commentators over the past two decades have argued that examination reform could improve the level and nature of students' achievements.

176. The reforms proposed for examinations have four major thrusts. The first is to improve administrative efficiency by, for example, the introduction of computerization and optical scanning. The second is to improve the quality of examinations by removing obvious deficiencies (such as limited curriculum coverage and a focus on recall and recognition). A third reform is to include topics in examinations that would be useful to the many pupils who do not go on to secondary school, and so remove the bias in examinations towards their selection function. A fourth is to change the content and taxonomic level of questions to include items that measure higher-order thinking skills and the ability to apply knowledge and skills in new situations, a reform that should benefit both students who leave school after the examination and those who will continue with their formal education. Since teachers "teach to the test", implementation of these reforms should be reflected in what and how teachers teach, which in turn should be reflected in students' achievements. To strengthen the impact, feedback based on an analysis of candidature performance on examinations, in which areas of achievement that are in need of attention are identified, may be provided to schools.

177. There can be no argument against these proposals, which all examination systems would do well to implement. Besides, many of the needed reforms could be achieved with little or no additional cost. Implementation can be assisted in a number of ways, including twinning relationships with overseas agencies to obtain technical assistance, and closer working relationships between examination boards, which already exist in some regions in Africa, to facilitate staff exchange and monitoring of standards. Regional capacity development initiatives could also explore ways of maintaining standards, combating malpractice, and training in new assessment and scoring techniques.

178. While a variety of reforms and innovations may be expected to improve the quality of examinations, it should not be assumed that better assessment will automatically enhance the curriculum and raise standards, just because poor assessment can narrow the curriculum and depress standards (Torrance, 1997). The available evidence suggests that if the content areas of examinations are changed (e.g., a new subject or a new component of a subject, such as essay writing, is examined), the content to which students are exposed in class will indeed change. The evidence regarding an improvement in students' overall achievement levels or in their cognitive processing skills is less clear, but we might expect them to be relatively modest. The present situation is that a great many students perform poorly on examinations. It is not simply the case that the scope of their achievements is limited or the taxonomic level of their cognitive processes low. This state of affairs can hardly be attributable to the quality of the examinations, which are likely to have less influence on student learning than lack of teacher competence, large classes, a lack of resource material, and the difficulty of teaching higher-order skills. It is unrealistic to expect that changing examinations will over-ride the influence of these factors.

179. The fact that some examination reforms are being driven by mechanisms involving high stakes (e.g., selection, competition, publication of results) raises questions about the negative (if unintended) consequences that might be expected. While it is true that examinations to which high stakes are attached are likely to impact on teaching and learning, there is a danger that the greatest benefits will accrue to high achieving students. This will happen if teachers focus their efforts on pupils who are most likely to succeed so as to increase their success rate; if low achieving pupils are retained in grade to prevent them sitting the examination; and, if discouraged, students leave school before completing primary education. These are all serious issues in the context of EFA policy that has as its objective ensuring that "recognized and measurable learning outcomes are

achieved by all, especially in literacy, numeracy and essential life skills.” The conflict between the traditional function of examinations, which was to select the highest achieving students for further education, and the goals of EFA, which express concern for *all* students, needs to be recognized, and steps taken to ensure that examinations do not frustrate attaining the goals. A first step would be to ensure that content relevant to all levels of student achievement is included in examinations, particularly at the end of primary schooling.

180. A great deal of activity relating to national assessment was generated in Africa during the 1990s. This would seem to be due primarily to the impetus and resources provided by international agencies and other donors following Jomtien. If this is indeed the case, then questions arise about the extent to which policy makers and planners in individual countries perceived a need for assessments, as well as about their sustainability. The fact that several countries have had as many as three separate national assessments (each sponsored or supported by a different agency) would suggest that the assessments were not in response to locally perceived needs, and are not integrated into the normal structures and activities of ministries. Given the need for such integration, the cost of the activities, and problems in recruiting personnel with the required competence to carry them out, it is clear that there is an urgent need for rationalization.

181. All national assessment activities in Africa have followed a similar pattern: pupils’ achievements are assessed in a sample of schools in basic curriculum areas and background information is collected from teachers and students, which is then related to the pupils’ achievements. (There are variations from this general pattern; for example, teachers have taken tests in some studies.) The input-output model represented in this approach, and which follows in the tradition of the original international studies of achievement (see Husén & Postlethwaite, 1996), has led to a tendency to accord the research function of assessments (e.g., to help identify or “understand” the causes of achievement) a higher status than its management function (e.g., to provide immediate feedback to decision makers about the condition of the education system). It has also led to a tendency to interpret associations between input factors and student outcomes in terms of cause and effect. This can result in unwarranted, and possibly incorrect, conclusions for at least three reasons. First, causality cannot normally be inferred from data collected in a cross-sectional study. Secondly, the instruments used to collect background information, and the way variables are constructed for analysis, may not identify the most important aspects of the environment that are related to achievement (e.g., the “cultural” and “social” capital of the home and community, rather than family socioeconomic status). And thirdly, the type of statistical analysis (e.g., hierarchical linear modelling) required to explore relationships is likely to be unfamiliar to many personnel involved in national assessments. For these reasons, it would seem more appropriate that the data obtained in a national assessment be considered as an input to educational planning conceived as a tool of strategic management (Verspoor, 1992). In this approach, assessment results would be interpreted and used in a “managerial” context, providing evidence for decisions (e.g., regarding the nature of inservice needed), while the more complex exploration of “causes” of achievement would be left to traditional research.

182. There is evidence that the findings of national assessments have found their way into government deliberations, reviews, and documents, though whether or not this happens would seem to vary from country to country. We may assume that it is more likely to happen when senior policy makers and managers have contributed to the initiation of a national assessment and to the questions which the assessment was designed to address. There is less evidence that the information from national assessments has impacted on school practice. That is not surprising. National assessments have not been in existence for very long and one would expect that it would take some time before they would have an impact. Besides, getting results to affect school practice is not a

trivial task. Although efforts have been made in a number of countries, their impact has not been assessed.

183. When considering the role of national assessments, it is important to bear in mind that their successful execution requires a considerable degree of technical competence. The development of capacity to achieve this competence is essential if the information they generate is to be valid and reliable, and so provide firm evidence for the decisions of policy makers and managers. In the absence of good quality design, instrumentation, sampling, and analysis, the effort may not be worth the expense. To ensure that it is, adequate budgetary provision by governments and donors is required.

184. Few African countries have participated in the several international assessments of student achievement which have been designed for industrialized countries. That is understandable and reasonable. Some of the national assessments which have been carried out in Africa do, however, allow some comparisons to be made between performance in different countries. There are indications that some countries are interested in developing this capacity further. However, the main value in international collaboration would seem to reside in the opportunities it provides to pool resources to develop the skills required to carry out assessments at the national level rather than in comparing the performances of countries.

185. We have already noted that teachers' assessment practices in the classroom have the greatest potential to enhance students' achievements. We have also referred to evidence that indicates that these assessment practices are of poor quality and, in particular, are unlikely to foster the development of higher-order and problem-solving competencies in students. Unfortunately, improving teachers' assessment practices is also more difficult and expensive than improving or developing other forms of assessment. It seems inevitable that additional resources will be required to improve teacher preservice education and to provide additional inservice education. It is particularly important that the many teachers who have been hastily hired to meet the EFA agenda are provided with training in classroom assessment.

186. Limitations of public examinations, which arise from the restricted range of knowledge and skills they assess and from the fact that high stakes are attached to student performance, are to some extent addressed by incorporating teachers' assessments into the grades awarded in public examinations. While several examination systems have taken or are taking steps to do this, there are many problems associated with the practice which render it unlikely that school-based assessment can entirely replace external examinations in selecting pupils as long as the number of places available at higher levels of education systems remains as limited as it is in many African countries. However, the incorporation of a component of school-based assessment into examination grades does have certain advantages. For one thing, the inservice and materials that teachers are provided with should have a beneficial effect on their more general assessment practices. This, in turn, may impact on lower achieving students.

187. A conclusion from the review presented in this paper is that while assessment information can improve policy and the management of resources in education and can shape teachers' instructional practice, success is not assured. Success in the first place will be dependent on the political will of governments to support the effort. Effort is required at two levels. At one level, national assessment systems should be institutionalized and integrated into the structures and processes of government policy making and decision making. This will require the integration of information from assessments into Educational Management Information Systems, and the establishment of procedures that will ensure that the information is provided to all involved in policy and management. In developing the capacity to achieve these objectives, countries will be able to draw on the assistance of the National Education Statistical Information Systems

(NESIS), a program initiated by the Working Group on Education Statistics (WGES) of the Association for the Development of Education in Africa (ADEA). (To date, 47 countries have participated in NESIS activities.) It will also be important to ensure that national assessments are aligned with other major instructional guidance systems in the education system – other assessment systems (including the alignment of standards), curricula, teacher education, school capacity building, and measures to address inequities.

188. Action will also be required at the level of the school. This will require opening adequate channels of communication to inform teachers of changes. In some cases, this may not be too difficult. Interventions may be direct and relatively straightforward when they involve the provision of space or materials. They are less likely to be so when problems relate to instructional objectives, equity, and quality (Verspoor, 1989, 1992), in which case teachers may need considerable and continuing support in interpreting reforms and in devising appropriate teaching strategies based on their interpretation. Inspectors, supervisors, and principal teachers, all of whom may require training, have a role to play in providing this support.

189. More than one observer of education in Africa has commented that educational reforms have not worked very well in the past. Certainly, it would seem true to say that single-focus isolated efforts have usually failed to fulfil their promise. Will proposals to use assessment to improve the quality of education fare any better? The answer would seem to be no if assessment reforms are not integrated into the structures and processes of government policy and decision making and are not aligned with other major instructional guidance systems. If, on the other hand, they are, there are grounds for some optimism.

190. There are other grounds for optimism, despite the many difficulties that face education in Africa which were outlined in the early part of this paper. Following Dakar, the formulation of the Millennium Development Goals (MDG) and the emphasis on national Poverty Reduction Strategy Papers (PRSP), policies, implementation plans, and activities have been, or are being, designed to instil a sense of urgency and to create a climate of accountability. Furthermore, assessment reform should provide objective information for decision making that has not been available in the past. Assessment reform efforts will also benefit from the commitment among ADEA partners to the development of a “culture” of finding solutions and policy responses within the African context to address problems and constraints (ADEA, 1999). If these conditions operate as promised, an important role for assessment in enhancing the quality of student learning can be envisaged.

7.1. Recommendations

191. The following recommendations are proposed for the consideration of Ministries of Education and other relevant education authorities and stakeholders.

- Ministries of Education, in collaboration with stakeholders, should develop overall policies for assessment, in which the goals of assessment systems (public examination, national assessments, classroom assessment) are aligned.
- Assessment systems should be designed to be mutually supportive in improving the quality of student learning.
- Assessment policies should be sensitive to and supportive of policies governing other components of the education system (including teacher education, school capacity building, measures to address inequities) to provide systemic, multifaceted, and coherent approaches to improving student learning.

- The information derived from assessments should be integrated into policy and decision-making structures of the Ministry of Education, to serve as an input to education planning and strategic management, and should also be conveyed to those responsible for curricula and textbooks.
- School curricula and public examinations should be reviewed to determine their relevance to the everyday lives of students and the needs of the economy, and required adjustments made. The review should have in mind the goals of Education For All.
- Efforts to reform public examinations to improve curriculum coverage and the quality of student learning should continue.
- Public examinations should be reviewed to ensure that they adequately reflect their certification function as evidenced in their content and difficulty level, and adjustments, if required, made.
- The scope of public examinations should be broadened so that certificates awarded students include data on aspects of student development not assessed in the written terminal examination.
- National assessments should be a response to the information needs of Ministries of Education and other stakeholders. Activities should be rationalized in countries in which more than one system operates.
- Ministries of Education should seek to meet high technical standards in national assessments by ensuring that the technical means and organizational capacity to monitor student achievements are developed.
- In recognition of the central role of teachers in improving the quality of student learning, steps should be taken
 - to improve the assessment skills of teachers in preservice and inservice courses, which accord adequate regard to the conditions in which teachers work (class size, availability of resources);
 - to provide guidance in the use of assessment information in making decisions regarding grade promotions;
 - to provide examples of good assessment practices at the end of chapters in textbooks;
 - to develop means of communicating information on assessment and reforms to classroom teachers, and to provide assistance in interpreting the significance of the information for classroom practice.
- Work should be taken to establish equivalences between achievements / qualifications awarded in the formal and non-formal sectors of the education system.

8. REFERENCES

- Ackers, J., Migoli, J., & Nzomo, J. (2001). Identifying and addressing the causes of declining participation rates in Kenyan primary schools. *International Journal of Educational Development*, 21, 361-374.
- ADEA (Association for the Development of Education in Africa). (1999). *What works and what's new in education: Africa speaks*. Paris: Author.
- ADEA. (2002). *Reaching out, reaching all. Sustaining effective policy and practice for education in Africa. Fighting HIV/AIDS*. Proceedings of the ADEA Biennial Meeting, Arusha, Tanzania, October 7-11, 2001. Paris: Association for the Development of Education in Africa.
- Airasian, P.W. (1988). Measurement driven instruction: A closer look. *Educational Measurement: Issues and Practice*, 7, 6-11.
- Arnott, A., & Kubeka, Z. (1997). *Mathematics and science teachers. Demand, utilization, supply and training in South Africa*. Johannesburg: Edusource.
- Bamgbose, A. (1991). *Language and the nation: The language question in Sub-Saharan Africa*. Edinburgh: University Press.
- Beaton, A.E., Postlethwaite, T.N., Ross, K.N., Spearritt, D., & Wolf, R.M. (1999). *The benefits and limitations of international achievement studies*. Paris: UNESCO: International Institute for Educational Planning; International Academy of Education.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7-74.
- Bude, U. (1993). Strategies for using information to improve learning conditions and instructional practices at the school level. In D.W. Chapman & L.O. Mählck (Eds.), *From data to action: Information systems in educational planning* (pp. 68-91). Paris: UNESCO: International Institute for Educational Planning.
- Bude, U. (1997). End of primary school examination in eastern and southern Africa: An overview. In U. Bude & K. Lewin (Eds.), *Improving test design. Vol 2. Assessment of science and agriculture in primary schools in Africa; 12 country cases reviewed* (pp. 12-30). Bonn: Education, Science and Documentation Centre.
- Bude, U., & Lewin, K. (1997). Introduction. In U. Bude & K. Lewin (Eds.), *Improving test design. Vol 1. Constructing test instruments, analysing results and improving assessment quality in primary schools in Africa* (pp. 6-11). Bonn: Education, Science and Documentation Centre.
- Buly, M.R., & Valencia, S.W. (2002). Below the bar: Profiles of students who fail state reading assessments. *Educational Evaluation and Policy Analysis*, 24, 219-239.
- Burnstein, L. (Ed.). (1993). *The IEA study of mathematics III: Student growth and classroom processes*. Oxford: Pergamon.
- Carron, G., & Châu, T.N. (1996). *The quality of primary schools in different development contexts*. Paris: UNESCO Publishing/International Institute for Educational Planning.
- Chapman, D.W., & Mählck, L.O. (1993). Improving educational quality through better use of information. In D.W. Chapman & L.O. Mählck (Eds.), *From data to action: Information systems in educational planning* (pp. 1-22). Paris: UNESCO: International Institute for Educational Planning.
- Chapman, D.W., & Snyder, C.W. (2000). Can high stakes national testing improve instruction? Reexamining conventional wisdom. *International Journal of Educational Development*, 20, 457-474.
- Chinapah, V. (1997). *Handbook on monitoring learning achievement. Towards capacity building*. Paris: UNESCO.

- Chinapah, V. et al. (2000). *With Africa for Africa. Towards quality education for all*. Pretoria: Human Sciences Research Council.
- Cizek, G.J. (2001). *Setting performance standards. Concepts, methods, and perspectives*. Mahwah NJ: Lawrence Erlbaum.
- Clayton, T. (1998). Explanations for the use of languages of wider communication in education in developing countries. *International Journal of Educational Development*, 18, 145-157.
- Eisemon, T.O. (1990). Examination policies to strengthen primary schooling in African countries. *International Journal of Educational Development*, 10, 69-82.
- Eisemon, T.O. (1997). *Reducing repetition: Issues and strategies*. Paris: UNESCO: International Institute for Educational Planning.
- Eisemon, T.O., Patel, V.L, & Abagi, J. (1987). Read these instructions carefully: Examination reform and improving health education in Kenya. *International Journal of Educational Development*, 8, 55-66.
- ERGESE (Evaluative Research of the General Education System in Ethiopia). (1986). *A quality study: Summary report presented to the Executive Committee of ERGESE*. Addis Ababa: Author.
- Eritrea. Department of General Education. (1999). *Competence test report 1998/99*. Asmara: Author.
- Grant, S., Peterson, P., & Shojgreen-Downer, A. (1996). Learning to teach mathematics in the context of system reform. *American Educational Research Journal*, 33, 500-543.
- Greaney, V., & Kellaghan, T. (1996a). *Monitoring the learning outcomes of education systems*. Washington DC: World Bank.
- Greaney, V., & Kellaghan, T. (1996b). The integrity of public examinations in developing countries. In H. Goldstein & T. Lewis (Eds.), *Assessment: Problems, developments and statistical issues* (pp. 167-188). New York: Wiley.
- Great Britain. Department of Education and Welsh Office. (1988). *Task Group on Assessment and Testing. A report*. London: Her Majesty's Stationery Office.
- Grisay, A., & Mählick, L. (1991). *The quality of education in developing countries: A review of some research studies and policy documents*. Paris: International Institute for Educational Planning.
- Heyneman, S.P. (1988). Improving university selection, educational research, and educational management in developing countries: The role of examinations and standardized testing. In S. Heyneman & I. Fägerlind (Eds.), *University examinations and standardized testing* (pp. 197-216). Washington DC: World Bank.
- Heyneman, S.P., & Ransom, A.W. (1992). Using examinations and testing to improve educational quality. In M.A. Eckstein & H.J. Noah (Eds.), *Examinations: Comparative and international studies* (pp. 105-120). Oxford: Pergamon.
- Himmel, E. (1996). National assessment in Chile. In P. Murphy, V. Greaney, M.E. Lockheed & C. Rojas (Eds.), *National assessments: Testing the system* (pp.111-128). Washington DC: World Bank.
- Howie, S.J. (1999). *National assessment in an international context: Building capacity and expertise in South Africa*. Paper presented at World Bank Conference on Human Development, March 3-5.
- Howie, S.J. (2001). *Mathematics and science programmes in grade 8 in South Africa 1989/99. TIMSS-R South Africa*. Pretoria: Human Sciences Research Council.
- Howie, S.J. (2002). *English language proficiency and contextual factors influencing mathematics achievement of secondary school pupils in South Africa*. PhD dissertation, University of Twente.
- Husén, T., & Postlethwaite, T.N. (1996). A brief history of the International Association for the Evaluation of Educational Achievement (IEA). *Assessment in Education*, 3, 129-141.

- Kanjee, A. (2003). Using assessment resource banks to improve the teaching and learning process. In *Improving the quality of primary education: Good practices and emerging models of district development* (pp. 59-71). Pretoria: District Development Support Programme/Research Triangle Institute.
- Kellaghan, T. (1992). Examination systems in Africa: Between internationalization and indigenization. In M.A. Eckstein & H.J. Noah (Eds.), *Examinations: Comparative and international studies* (pp. 95-104). Oxford: Pergamon.
- Kellaghan, T. (1996a). Can public examinations be used to provide information for national assessment? In P. Murphy, V. Greaney, M.E. Lockheed & C. Rojas (Eds.), *National assessments: Testing the system* (pp. 33-48). Washington DC: World Bank.
- Kellaghan, T. (1996b). IEA studies and educational policy. *Assessment in Education*, 3, 143-160.
- Kellaghan, T. (2002). *Assessment in the Swazi system of education*. Dublin: Educational Research Centre.
- Kellaghan, T. (2003). Local, national, and international levels of system evaluation: Introduction. In T. Kellaghan & D.L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 873-882). Dordrecht: Kluwer Academic.
- Kellaghan, T., & Farrell, E. (1998). *A study of the feasibility of obtaining, and the utility for describing learning outcomes of, data relating to public examination performance at the end of primary schooling*. Report submitted to UNESCO on behalf of the International Association for Educational Assessment.
- Kellaghan, T., & Greaney, V. (1992). *Using examinations to improve education. A study in fourteen African countries*. Washington DC: World Bank.
- Kellaghan, T., & Greaney, V. (2001a). The globalisation of assessment in the 20th century. *Assessment in Education*, 8, 87-102.
- Kellaghan, T., & Greaney, V. (2001b). *Using assessment to improve the quality of education*. Paris: UNESCO: International Institute for Educational Planning.
- Kellaghan, T., & Madaus, G.F. (2000). Outcome evaluation. In D.L. Stufflebeam, G.F. Madaus & T. Kellaghan (Eds.), *Evaluation models. Viewpoints on educational and human services evaluation* (2nd ed.; pp. 97-112). Boston: Kluwer Academic.
- Kellaghan, T., Madaus, G.F., & Raczek, A. (1996). *The use of external examinations to improve student motivation*. Washington DC: American Educational Research Association.
- Kelly, M.J. (1991). *Education in a declining economy. The case of Zambia, 1975-1985*. Washington DC: World Bank.
- Kudjoh, A., & Mingat, A. (1993). Toward a better understanding of the functioning of school systems for better decision-making: The case of primary schools in Togo. In D.W. Chapman & L.O. Mählck (Eds.), *From data to action: Information systems in educational planning* (pp. 147-174). Paris: UNESCO: International Institute for Educational Planning.
- Kulpoo, D., & Coustère, P. (1999). Developing national capacities for assessment and monitoring through effective partnerships. In *Partnerships for capacity building and quality improvements in education. Papers from the ADEA 1997 Biennial Meeting, Dakar, Senegal* (pp. 131-138). Paris: Association for the Development of Education in Africa.
- Kyalo, F.K. (1997). The use of examination results for monitoring performance of schools, districts and provinces. In U. Bude & K. Lewin (Eds.), *Improving test design. Vol. 1. Constructing test instruments, analysing results and improving assessment quality in primary schools in Africa* (pp. 138-144). Bonn: Education, Science and Documentation Centre.
- Le Mahieu, P.G. (1984). The effects on achievement and instructional content of a program of student monitoring through frequent testing. *Educational Evaluation and Policy Analysis*, 6, 175-187.

- Le Mahieu, P.G., & Leinhardt, G. (1985). Overlap : Influencing what's taught. A process model of teachers' content selection. *Journal of Classroom Interaction*, 21(1), 2-11.
- Lewin, K., & Dunne, M. (2000). Policy and practice in assessment in anglophone Africa: Does globalisation explain convergence? *Assessment in Education*, 7, 379-399.
- Linn, R.L. (1983). Testing and instruction: Links and distinctions. *Journal of Educational Measurement*, 20, 179-189.
- Linn, R.L. (2000). Assessments and accountability. *Educational Research*, 29(2), 4-16.
- Little, A. (1982). The role of examinations in the promotion of the 'Paper Qualification Syndrome'. In *Paper Qualifications Syndrome (PQS) and unemployment of school leavers. A comparative sub-regional study*. Addis Ababa: International Labour Office.
- London, N.A. (1997). A national strategy for systems-wide curriculum improvement in Trinidad and Tobago. In D.W. Chapman, L.O. Mählc & A.E.M. Smulders (Eds.), *From planning to action: Government initiatives for improving school-level practice* (pp. 133-146). Paris: International Institute for Educational Planning; Oxford: Pergamon.
- Machingaidze, T., Pfukani, P., & Shumba, S. (n.d.) *The quality of education: Some policy suggestions based on a survey of schools, Zimbabwe*. Paris: International Institute for Educational Planning; Harare: Ministry of Education and Culture.
- Madaus, G.F. (1988). The influence of testing on the curriculum. In L.N. Tanner (Ed.), *Critical issues in curriculum*. Eighty-seventh Yearbook of the National Society for the Study of Education (pp. 83-121). Chicago: University of Chicago Press.
- Madaus, G.F., & Greaney, V. (1985). The Irish experience in competency testing: Implications for American education. *American Journal of Education*, 93, 268-294.
- Madaus, G.F., & Kellaghan, T. (1992). Curriculum evaluation and assessment. In P.W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 119-154). New York: Macmillan.
- Marope, P.T.M. (1999). Capacity development through ADEA working groups: Applicable practices and lessons. In *Partnerships for capacity building and quality improvements in education. Papers from the ADEA 1997 Biennial Meeting, Dakar, Senegal* (pp. 139-174). Paris: Association for the Development of Education in Africa.
- Mazibuko, E.Z., & Magagula, R. (2003). *Quality of education in Sub-Saharan Africa: A literature review with specific reference to Swaziland c 1991-2001*. A review financed by ADEA through the Educational Network in Eastern and Southern Africa (ERNESA).
- Mioko, S. (1998). *Gender vs socioeconomic status and school location differences in grade 6 reading literacy in five African countries*. Harare: Ministry of Education and Culture.
- Mislevy, R.J. (1995). What can we learn from international assessments? *Educational Evaluation and Policy Analysis*, 17, 419-437.
- Namibia. Ministry of Education and Culture/Florida State University/Harvard University. (1994). *How much do Namibia's children learn in school? Findings from the 1992 National Learner Baseline Assessment*. Windhoek: Ministry of Education and Culture.
- Naumann, J., & Wolf, P. (2001). The performance of African primary education systems: Critique and new analysis of PASEC data for Senegal. *Prospects*, 31, 373-391.
- N'tchougan-Sonou, C. (2001). Automatic promotion or large-scale repetition – Which way to quality? *International Journal of Educational Development*, 21, 149-162.
- Ogundare, S. (1988). Curriculum development: A description of the development of the national curriculum for primary social studies in Nigeria. *Educational Studies*, 14, 43-50.

- Omolewa, M., & Kellaghan, T. (2003). Educational evaluation in Africa. In T. Kellaghan & D.L. Stufflebeam (Eds.), *Handbook of educational evaluation* (pp. 465-481). Dordrecht: Kluwer Academic.
- O-saki, K.M., & Agu, A.O. (2002). A study of classroom interaction in primary schools in the United Republic of Tanzania. *Prospects*, 32, 103-116.
- Ouane, A. (Ed.) (2003). *Towards a multilingual culture of education*. Paris: UNESCO: International Institute for Educational Planning.
- Oxenham, J. (1983). What examinations test and emphasize: An analysis of the Lesotho Primary and Junior Secondary Examinations. In *The education sector survey. Annexes to the report of the Task Force*. Maseru: Government of Lesotho.
- Pennyquick, D. (1990a). Factors influencing the introduction of continuous assessment systems in developing countries. In D. Layton (Ed.), *Innovations in science and technology education, Vol. III* (pp. 139-152). Paris: UNESCO.
- Pennyquick, D. (1990b). The introduction of continuous assessment systems at secondary level in developing countries. In P. Broadfoot, R. Murphy & H. Torrance (Eds.), *Changing educational assessment. International perspectives and trends* (pp. 106-110). London: Routledge.
- Popham, W.J. (1983). Measurement as an instructional catalyst. *New Directions for Testing and Measurement*, 17, 19-30.
- Popham, W.J. (1987). The merits of measurement driven instruction. *Phi Delta Kappan*, 68, 679-682.
- Postlethwaite, T.N., & Wiley, D.E. (1992). *The IEA study of science II: Science achievement in twenty-three countries*. Oxford: Pergamon.
- Rharade, A. (1997). Educational reform in Kenya. *Prospects*, 27, 163-179.
- Rollnick, M., Manyatsi, S., Lubben, F., & Bradley, J. (1998). A model for studying gaps in education: A Swaziland case study in the learning of science. *International Journal of Educational Development*, 18, 453-465.
- Romberg, T.A. (1999). The impact of international comparisons on national policy. In G. Kaiser, E. Luna, & I. Huntley (Eds.), *International comparisons in mathematics education* (pp. 189-199). London: Falmer.
- Ross, K.N. et al. (2000). *Translating educational assessment findings into educational policy and reform measures: Lessons from the SACMEQ initiative in Africa*. Paris: UNESCO.
- Samoff, J. (1999). Cooperation, but limited control and little ownership. In *Partnerships for capacity building and quality improvements in education. Papers from the ADEA biennial meeting, Dakar, Senegal* (pp. 175-204). Paris: Association for the Development of Education in Africa.
- Schiefelbein, E. (1993). The use of national assessments to improve primary education in Chile. In D.W. Chapman & L.O. Mählck (Eds.), *From data to action: Information systems in educational planning* (pp. 117-146). Paris: UNESCO: International Institute for Educational Planning; Oxford: Pergamon.
- Snyder, C.W., Prince, B., Lohanson, G., Odaet, C., Jaji, L., & Beatty, M. (1997). *Exam fervor and fever: Case studies of the influence of primary leaving examinations on Uganda classrooms, teachers, and pupils*. Washington DC: Academy for Educational Development.
- Somerset, A. (1987). *Examination reform in Kenya*. Washington, DC: World Bank.
- Somerset, A. (1988). Examinations as an instrument to improve pedagogy. In S.P. Heynemann & I. Fägerlind (Eds.), *University examinations and standardized testing. Principles, experience, and policy options* (pp. 169-194). Washington DC: World Bank.
- Somerset, A. (1996). Examinations and educational quality. In A. Little & A. Wolf (Eds.), *Assessment in transition. Learning, monitoring and selection in international perspective* (pp. 263-284). Oxford: Pergamon.

- Takala, T. (1998). Making educational policy under influence of external assistance and national politics. A comparative analysis of the education sector policy documents of Ethiopia, Mozambique, Namibia and Zambia. *International Journal of Educational Development*, 18, 319-335.
- Torrance, H. (1997). Assessment, accountability, and standards: Using assessment to control the reform of schooling. In A.H. Halsey, H. Lauder, P. Brown & A.S. Wells (Eds.), *Education: Culture, economy, and society* (pp. 320-331). Oxford: University Press.
- UNESCO. (1990). *World Declaration on Education for All. Meeting basic learning needs*. New York: Author.
- UNESCO. (2000a). *The Dakar Framework for Action. Education for All: Meeting our collective commitments*. Paris: Author.
- UNESCO. (2000b). *Education for All. Status and trends 2000. Assessing learning achievement (2000)*. Paris: Author.
- UNESCO. (2002). *Education for All. Is the world on track? EFA global monitoring report*. Paris: UNESCO.
- UNESCO. (2003a). *Monitoring Learning Achievement (MLA) Project. Update*. Paris: Author.
- UNESCO. (2003b). Southern African Consortium for Monitoring Educational Quality (SACMEQ). Harare: Author.
- V d Merwe, I.F.J. (1999). *Case study on the establishment of a national examinations and assessment system for school examinations in Namibia*. Windhoek: Ministry of Basic Education and Culture.
- Verspoor, A. (1989). *Pathways to change. Improving the quality of education in developing countries*. Washington DC: World Bank.
- Verspoor, A. (1992). *Challenges to the planning of education*. Washington DC: World Bank.
- Wasanga, P.M. (1997). Testing and monitoring procedures developed for primary schools. In U. Bude & K. Lewin (Eds.), *Improving test design. Vol 1. Constructing test instruments, analysing results and improving assessment quality in primary schools in Africa* (pp. 145-157). Bonn: Education, Science and Documentation Centre.
- World Bank. (1988). *Education in Sub-Saharan Africa. Policies for adjustment, revitalization, and expansion*. Washington DC: Author.
- World Declaration on Education for All*. (1990). Adopted by the World Conference on Education for All, Meeting Basic Learning Needs. New York: UNDP/UNESCO/UNICEF/World Bank