

Evaluation and Accountability in Asian and Pacific Countries

RAMON MOHANDAS

National Institute for Research and Development, Jakarta, Indonesia

MENG HONG WEI

China National Institute for Educational Research, Beijing, The People's Republic of China

JOHN P. KEEVES

Flinders University Institute of International Education, Australia

1. INTRODUCTION

The use of both evaluation and accountability in education systems has been of long standing, although the ways in which they are used have changed markedly over the past century. Where schools and universities are run as private, fee paying organisations, the clients, who are the students and their parents, must make their own judgments of the effectiveness and value of the education provided. However, where schools and universities are paid for out of public funds the need arises to evaluate and account for the spending of public monies. With the expansion across the world of the education systems in each country, province and municipality, the call for evaluation and accountability of those systems has increased, because of the substantial and rising expenditures involved. Over the past 100 years, not only has the size of the school-aged population grown markedly, but the demand has also grown, initially for universal primary education, subsequently for widespread secondary and technical education, and more recently for greatly increased provision of higher education and training in technology and commerce. The costs of providing these services are substantial and often relatively larger in those countries that can least afford to provide such services, but have the greatest need. This article considers the nature of both evaluation and accountability and examines how they are practised and developed in the region.

2. DEVELOPMENT OF EDUCATION SERVICES IN THE REGION

The expansion of education in the countries of the Asia-Pacific region was initially undertaken by the colonial powers that ruled those countries, namely, Great Britain (in the Indian subcontinent, Australasia, many of the Pacific Islands, Hong Kong, and Singapore, as well as countries like Malaysia where the British influence was strong), France (in the Indo-China, New Caledonia and Tahiti regions), the Netherlands and Portugal (in the Indonesian archipelago), and the United States (in the Philippines, the Hawaiian islands, and American Samoa). In addition some influence of Russian education is to be found in China, and such countries as Vietnam, North Korea and the Central Asian republics. Consequently, it is not surprising to find that the curricula and instructional practices employed in the countries of the Asia-Pacific region should reflect the traditions of Western Europe, and in particular, Great Britain, the Netherlands, Russia and France, as well as the United States. Nevertheless, it must be acknowledged that in many countries of the Asia-Pacific region there was a well-established education system, largely of a private nature, in operation in the eighteenth and nineteenth centuries prior to the introduction of influences from Western Europe. Today these education systems maintain longstanding traditions, even though Western influences have introduced many changes and produced a global uniformity of educational provision.

The explanation of the development of a largely common curriculum across not only the countries of the Asia-Pacific region, but also across the whole educational world is rather more complex than the simple process portrayed here in terms of the influence of dominant countries in Western Europe and the United States. Meyer, Kamens and Benavot (with help from Asian colleagues Yun-Kyung Cha and Suk-Ying Wong) (1992, p. 172) have examined the world-wide institutionalisation of primary school curricula and have considered the possibility of world-wide influences from: (a) evolving educational knowledge and theory, (b) general changes in world society, and (c) the interests and qualities of the world powers that have been politically and educationally dominant. In addition, there has been the influence of certain key individuals whose ideas, as is shown in this article, have been accepted across the world in many different and modified forms.

In Western Europe evaluation and accountability of the school systems were initially achieved by such policies as 'payment by results' to school principals. These policies were transferred to systems in the region (for example, see Hearn, 1872). Moreover, the examination system that the Jesuit Order introduced into their schools and colleges across Europe in the sixteenth century in order to lift the standards of education provided in European universities (Madaus & Kellaghan, 1992, p. 121) was carried by them to other parts of the world including Asia and Australasia, where schools of a largely academic nature were set up. The examination programs of these schools grew into national examination systems primarily for the purposes of selection into universities, but these examinations also served a function of certification as the school systems expanded.

Furthermore, the public examination systems in each country also provided a basis for evaluation and accountability, in so far as schools could be and were judged by their successes in the examinations that were conducted at the different levels of schooling.

With the expansion of education that took place in different countries at different times during the twentieth century, the examination systems in some countries collapsed under their own weight, since with rapidly growing numbers of candidates it became increasingly difficult to operate a system of public examinations. It was in the affluent United States during the early decades of the twentieth century where alternatives first emerged. This article is primarily concerned with the procedures that have been set up and are being introduced in the Asia-Pacific region to provide for the evaluation of the schools and school systems in the countries of this region.

3. EVALUATION, ACCOUNTABILITY, AND ASSESSMENT

It is necessary to distinguish between the meanings of three terms that are often used in ambiguous ways, namely 'evaluation', 'accountability', and 'assessment' before discussing in this article the procedures for evaluation and accountability that are employed.

3.1. *Evaluation*

In general, the use of the term 'evaluation' is reserved for application to entities, such as programs, curricula and organisations. The word 'evaluation' implies a general weighing of value or worth. Moreover, evaluation commonly involves the making of comparisons with a standard, or against criteria derived from stated objectives, or with other programs, curricula or organisations. Evaluation is primarily an activity that is associated with research and development. It generally involves consideration of the attainment of identifiable outcomes by both individuals and groups. Consequently, its potential importance for the improvement of educational practice is widely recognized, but fierce controversy sometimes surrounds the methods that should be employed in the conduct of an evaluation.

3.2. *Accountability*

Accountability is a process that involves the responsibility of systems and organisations to submit at regular intervals an account of the tasks they have performed to the body or bodies that have authorized or delegated to the system, or organisation the duty to perform those tasks. All educational organisations and systems have some form of accountability to a body or bodies that may sanction or reward the organisation or system (Neave, 1988, p. 19). In the field of education, schools, colleges, organisations and systems are accountable to a wide range of stakeholders. These include: (a) students, (b) teachers, (c) parents,

(d) school principals, (e) school councils, (f) school district personnel, (g) other educational organisations, (h) state, regional, and national educational agencies, (i) the national or regional parliament, and (j) the public at large. In so far as provision for education is made by many agencies with considerable expenditure involved, these agencies may require that an evaluation study should be conducted as part of the accountability process or may simply require that the organisations to which responsibility is delegated should report periodically. Alternatively, these agencies may require that an assessment study should be carried out to provide information on the performance of those individuals who benefit from the provision of educational services. However, the process of accountability generally involves much more than the supplying of information obtained through student assessment, since the use made of the inputs provided, must be considered as well as the outputs obtained.

3.3. *Assessment*

In general, the employment of the term 'assessment' in the field of education is reserved for use with reference to people, since it involves estimating the amount of learning or development that has occurred, or the level of performance attained. It may involve the administration of tests that relate to specific outcomes. Alternatively, it may simply involve the activity of grading or classifying student performance with respect to specified criteria. The evaluation of a curriculum or program may require the collection of information through the assessment of student performance with regard to the objectives of the curriculum or program, but it is the value of the curriculum or program that is being considered in an evaluation study, and not the value or worth of those students whose performance is assessed. It is, nevertheless, unfortunate that the term 'student evaluation' is now being widely used as a consequence of the growing emphasis on the evaluation of educational programs.

Education is primarily concerned with the learning and development achieved by individuals in organisations and systems. At the core of the educational process is the curriculum provided by these organisations and systems for individuals. Consequently, any consideration of evaluation and accountability in education must focus on the curricula of educational organisations and systems and on the instruction provided by those organisations.

4. CURRICULUM, INSTRUCTION AND ASSESSMENT IN SCHOOLS

In the early decades of the twentieth century the industrialized countries faced the tasks of (a) expanding secondary education, (b) meeting the vocational needs of a technical revolution, and (c) reforming education to make it more efficient, as was required of industry and commerce. Cubberley (1917) in the United States argued that no longer could the efficiency of the school system be solely determined by personal opinion and inspection. A better method for the evaluation of the work of schools and their teachers had been developed that involved

the use of standardized tests. The outcomes of education could be measured by such tests, and school superintendents could use the results of testing to evaluate the instructional program of a school (Munroe, 1917). Moreover, these tests would provide teachers with a standard by which they might judge the performance of their students in comparison with the performance of other students in the same school or the same school district. The idea of measuring the outcomes of education using the assessment of student performance was further developed during the first four decades of the twentieth century to meet the changing nature of the schools that were influenced by the Progressive Education Movement.

In 1933, a major study was set up to examine the effects of this reform movement in secondary education in the United States with R.W. Tyler as Research Director of the Evaluation Staff. From the work of this study, the Eight Year Study, a new approach to the evaluation of the work of schools emerged, that was formulated in *The Basic Principles of Curriculum and Instruction* (Tyler, 1949). These ideas were further developed by Bloom and his colleagues in the *Taxonomy of Educational Objectives Handbook 1, Cognitive Domain* (Bloom, 1956) and *Handbook 2, Affective Domain* (Krathwohl, Bloom and Masia, 1964). In 1971, these ideas were disseminated widely through a workshop in Sweden conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA) at which teams of six people from selected countries involved in curriculum development and evaluation studied together for a period of six weeks. Asian teams came from Thailand and Malaysia.

Thus, this new approach to curriculum development and evaluation spread throughout both the developed and developing countries of the world. Bloom directed this workshop and the *Handbook of Formative and Summative Evaluation of Student Learning* (Bloom, Hastings & Madaus, 1971) was the set text for the workshop. Tyler attended for all six weeks as a key speaker and resource person with his monograph on the basic principles of curriculum and instruction being reprinted for the occasion. The Tyler model of evaluation has gradually become a guiding model for developments in evaluation and accountability in many countries of the world, although the model has undergone many changes over time (see Tyler, 1986).

4.1. *The Tyler Model*

The Tyler Model for evaluation in education may be shown as a triangle, with curriculum objectives at the apex of the triangle, that lead to instruction and the provision of learning experiences, which in turn lead to an evaluation of the extent to which the objectives are realized. The relationships in this triangle are reciprocal in nature, in so far as evaluation feeds back to instruction and to the curriculum objectives. The curriculum triangle is shown in Figure 1 (Tyler, 1949).

While evaluation refers to both the curriculum objectives and to the learning experiences provided, at the student level the process involves the assessment of

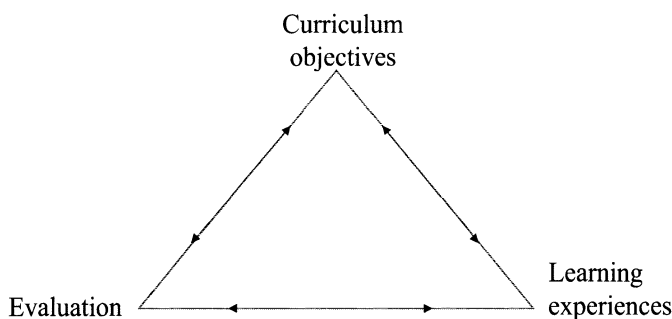


Figure 1. The curriculum triangle

student performance. Although the approach that Tyler proposed was a model for developing curricula, it has tended to become a framework for test development, with the provision of learning experiences frequently overlooked. Tyler's approach was reduced to seven steps: (a) formulating objectives, (b) classifying the objectives, (c) refining and defining the objectives, (d) identifying ways of assessing student performance with respect to the objectives, (e) trying out appropriate ways of assessing student performance, (f) improving methods of assessment, and (g) interpreting the results with a reexamination of the objectives. Tyler (1986) subsequently questioned the over-emphasis on testing and assessment by psychologists and psychometricians at the expense of examining the curriculum objectives and the learning experiences employed in schools.

A further widespread shortcoming in the use of this approach has been a reliance on assessment using relative performance with respect to other students in a specified reference group, rather than the employment of standards of performance to be attained. Thorndike (1918) recognized the need to specify degrees of proficiency in the assessment of student performance. However, several decades passed before Mager (1962) emphasized the setting of standards as a step associated with the specification of curriculum objectives. Bloom (1968) in the consideration of learning for mastery, and Tyler (1973) in detailing the place of testing in programs of student assessment have both subsequently acknowledged the importance of identifying standards of performance.

The curriculum triangle in Figure 1 requires modification and extension in order to (a) introduce the standards of performance; (b) indicate that assessment takes place with students; and (c) that evaluation and accountability apply to all four other vertices of the curriculum-evaluation diamond as presented in Figure 2. The idea that evaluation only relates to the learning outcomes and to student assessment introduces a fallacy, which is unfortunately all too common, and is detrimental to the best interests of educational policy and practice. Research workers in Australia (Keeves, 1999, pp. 128–130) have argued for the recognition of underlying scales associated with the curriculum objectives, the standards of performance, and the assessment of student outcomes and have shown how measurements can be made to test the measured structures of the

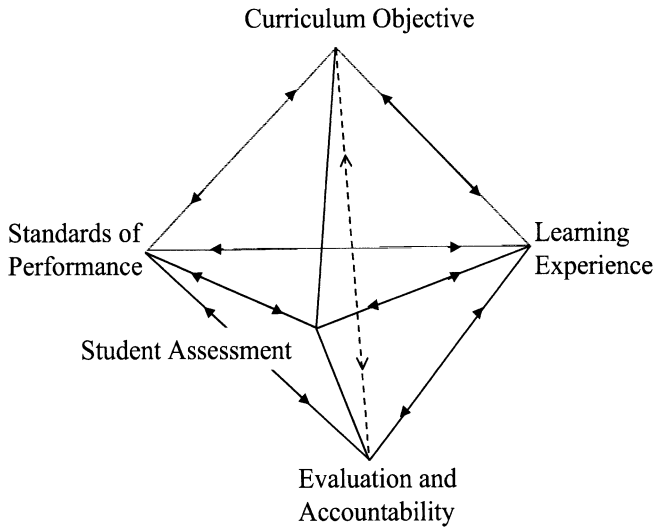


Figure 2. The curriculum-evaluation diamond

curriculum objectives, the performance standards, and the outcomes achieved by students. Teachers in planning instruction for classroom groups of students should examine the structures that are measured, and proceed with the provision of appropriately sequenced learning experiences.

4.2. *The Evaluation of the Mastery Learning Project in Korea*

The Mastery Learning Project that directly involved standards of performance was conducted in Korea and was evaluated within the Tylerian framework and Bloom's notions of a model of school learning to examine the effectiveness of strategies of mastery learning. This large-scale evaluation study was associated with one of the most systematic and most successful innovations in Korean education during the second half of the twentieth century. The introduction of mastery learning strategies that has had a marked impact on learning in Korean schools (Kim, 1975, pp. 13–22) was built upon a soundly conducted evaluation study.

5. PUBLIC EXAMINATIONS AND ACCOUNTABILITY

Public examinations, whether conducted at the national, systemic, or regional levels provide an important form of accountability and direct evidence of the effectiveness of a school or organisation in terms of the number and level of preparation of students, who are successful in the examinations. In general, the success rate of a school becomes public knowledge, and judgments can be made about the effectiveness of the school and the organisation's contribution to student success. However, the use of examination results in this way suffers from

two major shortcomings. First, schools differ in the quality of the intake that they receive, either because they use selection procedures in recruiting students or they have catchment areas in which students of higher quality live. Second, schools differ in their retention rates, and even if schools had intakes of students of similar quality, their graduating classes would differ in quality because of differing retention rates. Consequently, it would seem appropriate for a school to be judged by what it achieves with the students it recruits. Nevertheless, these shortcomings are largely ignored in the comparisons that are made about a school's performance from a consideration of the successes of its students in public examinations. There is no doubt that public examinations provide evidence, which is widely used in the evaluation of schools and organisations (Keeves, 1994).

5.1. *The Certification of Graduating Students*

Public examinations provide for the certification of students at the terminal stages of schooling as well as selection for further education. Consequently, public examinations at the middle and upper secondary school levels are sometimes maintained in order to award successful students with a graduation certificate. However, the functions of selection and certification that are associated with the provision of public education are to some extent incompatible. This sometimes leads to the conduct of examinations at the terminal stage of schooling in two forms, namely a certification examination that is held at a regional level and a selection examination that is conducted by or on behalf of the institutions of higher education. China, India, Indonesia, Japan, and the Republic of Korea maintain different examinations to serve these two purposes of selection and certification. Nevertheless, where such public examinations are conducted there is a clearly identified form of accountability that publicly provides evidence for the evaluation of both schools and their teaching methods, as well as the curriculum that is laid down for the schools by the examination system.

5.2. *Large-Scale Testing Programs*

Australia, New Zealand and American Samoa, are countries where public examinations are, in general, not held in any form during the years leading up to the terminal secondary school stage. Moreover, at the terminal stage, teachers through their assessment of students' work during the final year of schooling contribute to the gradings assigned for the purposes of selection and certification, sometimes in conjunction with, or moderated by public examination scores. Consequently, in these countries there is seen to be a serious lack of evidence for accountability purposes that was formerly provided by a system of public examinations at levels below the terminal stage of schooling. This has led to the widespread use of standardized tests for the purposes of accountability that are prepared in New Zealand by the national Department of Education, and in the territories of the United States largely by commercial publishing houses. In Australia, each of the States and Territories of the Commonwealth of Australia

conducts its own state-wide testing programs for all students at Grades 3 and 5 and increasingly at Grade 7 in order to provide evidence for accountability purposes to the different stakeholders in the Australian school systems. While the tests conducted in the different States are not strictly comparable, work is in progress to provide comparable results as part of an accountability program that compares the performance of students in different regions of Australia.

6. NATIONAL APPROACHES TO PUBLIC EXAMINING – AN EXAMPLE

In 1973, the Indonesian Government through the *President's Instruction Number 10, 1973* started an educational program to provide better opportunities for learning by school-aged children in elementary schools. The success of this six-year program of compulsory education inevitably led to the next educational program, namely the commencement of the nine-year compulsory basic education program. This program was officially declared by the President of the Republic of Indonesia on 2 May 1994. However, this rapid expansion of education has been accompanied by a demand to ensure that standards were being maintained, which has led to the development of an extensive examination system.

The structure of the school system in Indonesia consists of the Elementary School (6 years), the Junior Secondary School (3 years), the Senior Secondary School (3 years), and the University. At the end of each cycle of schooling in Primary and Secondary Education (Grade 6 of the Elementary Schools, Grade 9 of the Junior Secondary Schools, and Grade 12 of the Senior Secondary Schools), a final examination is administered. For Elementary and Junior Secondary Schools, the final examination serves two purposes, namely for certification and selection to a higher level of education. However, for Senior Secondary Schools, the final examination only serves as certification, because universities, both state and private, do not use the final examination results for selection into the universities, they administer their own entrance examinations.

The final examinations in Indonesia have experienced several major changes in recent decades. Until the early 1970s, the final examination for the Primary and Junior Secondary Schools in Indonesia was administered using a system referred to as 'State Examination.' In this system, a national committee at the central level prepared examination papers for all subject-matter areas tested throughout Indonesia. Students from all over Indonesia at the same level of schooling, taking the same school subjects, would take the same set of items for the final examination.

From the early 1970s to the early 1980s, there was a major change in the system. The State Examination was changed into the Schools Examination. In this new system, every school was given the authority to construct its own tests, to score them, and to decide the passing grade for the students taking the tests. This system was referred to as the 'Schools Examination'. This change gave greater freedom to the schools across the country at a time of marked expansion.

From the 1980s up to the present, there was a further change in the system, referred to as *EBTANAS* (this is an acronym in Indonesian language which means 'National Final Examination of Students' Achievement'). These changes were introduced in an effort to raise the standards of education across the country during the period of continued expansion of the school system. In this new examination system, the test specifications for the subject matter to be tested at the end of each stage of schooling in both Primary to Junior Secondary Education were prepared by a national team in Jakarta.

Thus for both primary and secondary schools, the administration of the teaching-learning process and the conduct of the examinations have been the responsibility of the Directorate General of Primary and Secondary Education. However, the Examination Development Centre, as a research and development institution within the Office of Educational and Cultural Research and Development, has been responsible for the design and development of the examination papers. Among its specific tasks in the development of the examination system have been the construction of a national item bank, as well as the provision of technical guidance to teachers in item writing, both tasks serving the purpose of raising the quality of the examination papers employed across the country.

Formerly for the Primary and Junior Secondary Schools, the provincial committees developed three to seven test forms for each subject-matter area. The Provincial Offices were given the authority to review, finalise, and administer these tests in their respective provinces and to undertake the scoring of students' test papers. The passing score for the test, however, was determined by the Directorate General of Primary and Secondary Education in Jakarta. For the Senior Secondary Schools, the provincial committees developed one set of items for each subject-matter area. These test items were then sent to Jakarta for further review by the national team, which developed the test specifications. The national team was responsible for the preparation of seven sets of items out of the 26 sets received from the 26 provinces. The final versions of the sets were then sent back to each province for duplication and for administration in the final examination. Each province was given the authority to make decisions on the results of this examination.

Beginning in the 2000/2001 school year, there was a slight change in these arrangements. By the decree of the Minister of National Education issued in February 2001, the Examination Development Centre is now given full authority to administer the examinations. However, the examination at the end of Primary Schooling was subsequently abandoned. This Centre is now responsible for the construction and administration of the test items for all subject matter areas in Junior Secondary, and Senior Secondary Schools. The construction of the test items at both levels of schooling are carried out primarily by choosing for each subject matter area at each level appropriate items that are available in an Item Bank which is managed by the Examination Development Centre.

For entry to a higher level of schooling (from Junior Secondary Schools to Senior Secondary Schools), the scores of *EBTANAS* are used as criteria for

selection. These scores, referred to as *NEM* (*Nilai EBTANAS Murni*, meaning Purely *EBTANAS* Scores), are recorded on separate sheets as a supplement to the certificates. As the name suggests, the scores obtained by the students in *EBTANAS* are recorded without change on the separate sheets.

In this system, comparability of test scores becomes a very important issue in order to make fair decisions in selection. If the different test forms were not comparable, the students who happened to take a more difficult test form would be disadvantaged compared to those who might take an easier test form. From the way the different test forms were constructed without pretesting, it would be methodologically wrong to assume that the different test forms would automatically provide comparable results.

In the 1994 administration of the *EBTANAS*, the Examination Development Centre introduced the use of common items in the five different test forms for six subject matter areas in the Junior Secondary Schools. The six subject matter areas are: (a) Pancasila Moral Education; (b) Indonesian Language; (c) Social Science; (d) Science; (e) Mathematics; and (f) English. The purpose of using these common items is to make the equating of the different test forms in each subject possible. By using these common items for linking the different forms, Rasch equating can be employed to obtain comparable scores across the different test forms.

In the current selection system, the scores of the students in the *EBTANAS*, regardless of which forms they are taking, play a major role in the decision to admit or to reject students into the schools they have chosen to attend. In most cases, the favourite schools, which are naturally the better schools, set their criteria of selection in such a way that only those students with high scores are admitted to the schools. Since there are different forms of tests used in the *EBTANAS* for security reason, the comparability of test scores is essential in order to make fair decisions about which students to admit and which students to reject. This comparability must be maintained even though the students take different test forms in the *EBTANAS*.

In an effort to raise the standards of education across the country, there should also be reliable information about the achievement of the students in different places and at different times. The information about the increase or the decrease of the students' achievement over the years in different parts of the country can only be obtained when the measures of their achievement are also comparable over time and place. Comparability of students' scores in *EBTANAS*, to a certain extent, also serves the purpose of monitoring the achievement levels of students over the years and the different locations, so that this information can also be used in an effort to improve the quality of education provided by the schools.

7. SURVEY STUDIES OF EDUCATIONAL ACHIEVEMENT

The introduction of survey studies involving testing programs at the primary and lower secondary school levels in order to assist with the advancement of

education at those stages of schooling has become recognized as a way in which the work of the schools can be evaluated at less expense than is involved in conducting nationwide testing programs (see *Case Study of Learning Achievement in South Asia*). These survey types of testing programs have been developed over the past 20 years.

In 1980, the International Association for the Evaluation of Educational Achievement (IEA) and the National Institute for Educational Research (NIER) in Japan organised training seminars for teams of evaluators drawn from the Office of Educational and Cultural Research and Development in Indonesia, the National Education Commission in Thailand, and the Ministry of Education in Malaysia. The seminars were held in two stages. At the first seminar, general instruction was provided in evaluation techniques, and detailed plans of evaluation were developed by the six team members from each country. At the second seminar held 18 months later in 1982, data analyses were undertaken and national reports prepared. While full national reports were written on each of the two survey studies conducted by each team in the national languages, short reports were also prepared for journal publication (Amara, Abu & Jiyono, 1982). This program provided highly effective training for research workers to carry out evaluation studies, using survey research procedures and the monograph prepared on this program indicated the type and quality of work in educational evaluation in countries in the Asia-Pacific region. Such training programs have been replicated from time to time in the different countries in the Asia-Pacific region.

8. SCHOOL-BASED INDICATORS OF EFFECTIVENESS

In 1995, the Human Resources Development Working Group of the Asia-Pacific Economic Cooperation (APEC) organisation initiated a study of school-based indicators of effectiveness. Nine countries participated in the study, namely Brunei Darussalam, People's Republic of China, Chinese Taipei (Taiwan), Hong Kong, New Zealand, Singapore and Thailand from the Asia-Pacific region and Canada and the United States who were from outside the region, although they were APEC members. The economic growth and social development of the APEC member states had created an increased interest in monitoring and improving the quality of education provided. In order to improve the educational system of each country, it was considered necessary to improve the performance of individual schools and to strengthen the effectiveness of each school. The aim of the study was to provide an overview of existing indicators of school effectiveness in the Asia-Pacific region through a survey and through a series of case studies. More specifically the study looked at:

- (a) policy issues in the evaluation of school effectiveness,
- (b) a historical description of the study of school effectiveness,
- (c) existing indicators and programs of school effectiveness, and

(d) how these indicators and programs catered for needs at the school level.

The project sought to prepare through a cooperative study, a guidebook that described the methods, programs of supervision of school effectiveness, and use of school-based indicators that were valid and reliable, and that identified common approaches and strategies to answer several questions.

1. Whether the resources devoted to education were achieving the desired effects?
2. What standards of achievement should be expected of schools?
3. How could it be shown that standards had improved over time?

The orientation of the study was for schools to be responsible for evaluating their own effectiveness and reporting to their stakeholders including their teachers, students and parents on how each school performed.

Each country participating in the project undertook a case study of practices within the country and submitted a report for publication. In addition, consultants presented invited papers on:

- (a) monitoring school effectiveness: conceptual and practical possibilities and dilemmas in developing a framework (Cheng, 1997a);
- (b) framework of indicators of education quality in Hong Kong primary schools: development and application (Cheng, 1997b); and
- (c) developing indicators on the performance of education systems in an international context: The OECD Education Indicators (Schleicher, 1997).

The report of the study (Meng Hong Wei, Zhou Yigun & Fang Yihua, 1997), found that member states saw educational outcomes as an important indicator of school effectiveness, with growing concern for the physical and emotional development of students. A critical issue was the need to maintain a balance between the school's internal development and accountability to the public through the monitoring of school effectiveness. Consequently, there was a need to manage the balance in both evaluation and accountability between the school level and the system level. The information collected through this project also showed that most schools had little systematic knowledge of the nature and extent of their effectiveness and used few indicators of institutional performance.

9. CONCLUSION

In most of the countries of the Asia-Pacific region educational evaluation studies have had only a short life of no more than 30 years. During that time the conduct of educational evaluation studies has changed markedly throughout the world. The advance of the electronic computer has for many investigators changed the way in which evaluation studies could be carried out because large bodies of data could now be conveniently stored and readily processed. Under

these circumstances the vision is not merely the conduct of on-off studies of evaluation and accountability, but the undertaking of a planned sequence of related investigations that monitor the changes that are occurring. Consequently it should be possible over the coming decades to provide advice for policy and practice, from an ongoing series of studies involving both evaluation and accountability, that examine national systems of education and monitor the changes that are occurring. Questions must be asked not only about the direction of change, but also the rate of change and the factors that influence both direction and rate. Unfortunately governmental policies vary markedly from occasion to occasion, in a reaction to perceived changes that have occurred and new situations that have arisen. What is needed is a longer term vision of the development of an education system with perceptive monitoring of the changes taking place.

REFERENCES

- Aikin, W. M. (1942). *Adventure in American Education, Volume 1, The Story of the Eight-Year Study*. New York: Harper.
- Amara Sawadisevee, Abu Bakar Nordin & Jiyono (1982). Six educational evaluation projects in South East Asia. *Evaluation in Education*, 6(1), 1–158.
- Bloom, B. S. (Ed.) (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1, Cognitive Domain*. New York: McKay.
- Bloom, B. S. (1968). Learning for mastery. *University of California Comment*, 1(2), 1–12.
- Bloom, B. S., Hastings, T. J., & Madaus, G. F. (1971). *Handbook of Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.
- Cheng, Y. C. (1997a). Monitoring effectiveness: Conceptual and practical possibilities and dilemmas in developing a framework. In Meng Hong Wei et al. (1997), *School Based Indicators of Effectiveness: Experiences and Practice* (pp. 179–206). Guilin, China: Guangxi Normal University Press.
- Cheng Y. C. (1997b). A framework of indicators of education quality in Hong Kong primary schools: Development and application. In Meng Hong Wei et al. (1997), *School Based Indicators of Effectiveness: Experiences and Practices* (pp. 207–249). Guilin, China: Guangxi Normal University Press.
- Cubberley, E. P. (1917). Editor's introduction. In W. S. Munroe (1924), *Educational Tests and Measurements* (pp. xi–xv). Boston: Houghton Mifflin.
- Hearn, W. E. (1872). *Payment by Results in Primary Education*. Melbourne: Stellwell and Knight.
- Keeves, J. P. (1994). *National Examinations: Design, Procedures and Reporting*. Paris: UNESCO, IIEP.
- Keeves, J. P. (1999). Research into curriculum change. In J. P. Keeves & K. Marjoribanks (Eds.), *Australian Education: Review of Research* (pp. 113–144). Melbourne: ACER.
- Kim, H. (1975). Evaluation of the mastery learning project in Korea. *Studies in Educational Evaluation*, 1(1), 13–22.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of Educational Objectives. The Classification of Educational Goals, Handbook 2, Affective Domain*. New York: McKay.
- Madaus, G. F., & Kallaghan, T. (1992). Curriculum evaluation and assessment. In P. W. Jackson (Ed.) (1992), *Handbook of Research in Curriculum* (pp. 119–154). New York: Macmillan.
- Mager, R. F. (1962). *Preparing Objectives for Programmed Instruction*. San Francisco: Fearon Publishers.
- Meng Hong Wei, Zhou Yigun & Fang Yihua (1997). *School Based Indicators of Effectiveness: Experiences and Practices*. Guilin, China: Guangxi Normal University Press.
- Meyer, J. W., Kamens, D. H., & Benavot, A. with Yun-Kyung Cha and Suk-Ying Wong (1992). *School Knowledge for the Masses*. London: Falmer Press.
- Munroe, W. S. (1917). *Educational Tests and Measurements*. Boston: Houghton Mifflin.

- Neave, G. (1985). Accountability in education. In T. Husén & T. N. Postlethwaite (Eds.), *The International Encyclopedia of Education* (pp. 19–29). Oxford: Pergamon.
- Schleicher, A. (1997). Developing indicators on the performance of education systems in an international context: The OECD education indicators. In Meng Hong Wei et al. (1997), *School Based Indicators of Effectiveness: Experiences and Practices* (pp. 251–264). Guilin, China: Guangxi Normal University Press.
- Thorndike, E. L. (1918). The nature, purposes and general methods of measurements of educational products. In G. M. Whipple (Ed.), *The Measurement of Educational Products* (17th NSSE Yearbook) IL (pp. 16–24). Bloomington: Public School Co.
- Tyler, R. W. (1949). *Basic Principles of Curriculum and Instruction*. Chicago: University of Chicago Press.
- Tyler, R. W. (1973). Testing for accountability. In A. C. Ornstein (Ed.), *Accountability for Teachers and School Administrators*. Belmont, CA: Fearon Publishers.
- Tyler, R. W. (1986). Changing concepts of educational evaluation. *International Journal of Educational Research*, 10(1), 1–113.