



---

A PUBLICATION OF THE CENTER FOR EDUCATIONAL MEASUREMENT

APRIL 1995

VOLUME 6

---

# PHILIPPINE JOURNAL OF EDUCATIONAL MEASUREMENT

Special Issue on the  
1st Asia-Pacific Invitational  
Conference on Educational  
Assessment and Research



Center for Educational Measurement

# PHILIPPINE JOURNAL of EDUCATIONAL MEASUREMENT

---

**APRIL 1995      •      SPECIAL ISSUE      •      VOLUME 6**

---

The Philippine Journal of Educational Measurement publishes empirical papers and nonempirical reports such as theoretical notes, specialized reviews, and commentary on issues related to educational testing, measurement, assessment, evaluation, and research in all of its many aspects.

---

## EDITORIAL STAFF

LENORE LL. DECENTECEO  
Editor

CARLOS H. LOPEZ  
Technical Consultant

MA. CYNTHIA A. ALCANTARA  
VENIA FE E. PEPITO  
Editorial Assistants

MYRNA A. DIAZ  
Layout Artist

---

Address all inquiries and submit manuscripts for publication to:

The Editor  
Philippine Journal of Educational Measurement  
Center for Educational Measurement  
6/F Wing B, Concorde Condominium  
Cor. Salcedo and Benavides Sts.  
Legaspi Village, 1229 Makati City  
PHILIPPINES

Articles and commentary in the Journal represent the views of the authors and do not necessarily reflect official CEM policy.

---

## FOREWORD

This volume is a special issue on the proceedings of the *First Asia-Pacific Invitational Conference on Educational Assessment and Research* held in Manila in February, 1994.

A major aim of the conference is to make each country in the Asia-Pacific region aware of the educational systems of its neighbors, and to organize these countries into an association which will ensure a continual sharing of ideas, experiences, and expertise in the field of educational assessment and research.

The papers reflect a diversity in educational systems. But they also reflect a need to evolve, develop, and define an assessment and research perspective highly responsive to the specific educational contexts of countries in the Asia-Pacific region.

LETICIA M. ASUZANO  
President, CEM

## Table of Contents

Keynote Address: Educational Assessment in the Asia-Pacific Region: Issues and Problems .....	1
<i>Armand V. Fabella</i>	
Organizing for International Assessment of Educational Systems in the Asia-Pacific Region .....	4
<i>Abraham I. Felipe</i>	
Measurement Of Educational Outcomes in Basic Education: Thai Experience .....	9
<i>Sangob Laksana</i>	
Assessment in Elementary Education in the Philippines .....	13
<i>Marcelina M. Miguel</i>	
Assessment for Admission in Higher Education: the Philippine Perspective .....	16
<i>Lucila F. Tibigar</i>	
Assessment in the Singapore School Sytem.....	22
<i>Andrew Poh Sui Hoi</i>	
The Development of National Monitoring System of Educational Achievement in China.....	23
<i>Meng Hong-wei</i>	
Educational Assessment in the Guam Public School System.....	32
<i>Zenaida N. Natividad</i>	
Assessment in the Private School Sector: the CEM Experience*.....	36
<i>Esperanza C. Buen</i>	
An Appraisal of Teaching of Evaluation in Colleges of Education in India.....	40
<i>G. Subramonia Pillay</i>	
Multi-Media Education: Implications for Assessment .....	42
<i>Akira Ichikawa</i>	



Professional Occupations in Thailand: A Caselook of Registration and Licensureship .....	46
<i>Wilars Singhawisai</i>	
The Secondary Education Development Program (SEDP) Evaluation .....	54
<i>Eugenia M. Moraleda</i>	
Assessment in Vocational Education: An Australian Perspective .....	58
<i>Peter Thomson</i>	
Quality Assurance System in Thai Higher Education: Current Developments and Issues .....	67
<i>Amornwich Nakornthap</i>	
Assessment For Admission to Medical Colleges in the Philippines .....	75
<i>Basilio R. Iledan</i> <i>Lenore Ll. Decenteceo</i>	
Issues and Problems In Educational Research in the Philippines .....	80
<i>Jesus A. Ochave</i>	
Panukat ng Pagkataong Pilipino (PPP) .....	82
<i>Annadaisy J. Carlota</i>	
Short Forms of the Marlowe-Crowne Social Desirability Scale (MCSDS): Comparative Perspectives of the Classical and Modern Test Theories .....	91
<i>Werawat Punnitamai</i>	
An Introduction to Item Response Theory Using <i>Quest</i> .....	92
<i>Malcolm Rosier</i>	
The Application of IRT and Optimal Test Designs in the Assessment of Mathematical Abilities .....	99
<i>Dr. Peter Lam Tit Loong</i>	
Directory of Participants .....	111

**KEYNOTE ADDRESS:  
EDUCATIONAL ASSESSMENT IN THE ASIA-PACIFIC REGION:  
ISSUES AND PROBLEMS**

**Armand V. Fabella**  
*Secretary*  
*Department of Education, Culture, and Sports*  
*Chairperson*  
*CEM Board of Trustees*

On behalf of the Philippine government allow me to welcome you all to our country and to the First Asia-Pacific Invitational Conference on Educational Assessment and Research.

Allow me also to take this opportunity to commend the organizers of this conference which hopefully will lead to a better understanding of the different educational systems in the Asia-Pacific region and, in more concrete terms, to greater cooperation in the field of educational assessment and research among those countries represented here.

It is indeed with great eagerness that we wait to hear of how educational assessment is practiced and, for that matter, the extent to which it is accepted in other Asia-Pacific countries.

**Educational Assessment in the Philippines**

In the Philippines, the increasing membership of both public and private schools in the Center for Educational Measurement (CEM) indicates the extent to which educational assessment is now being utilized to improve instruction.

The importance of educational assessment has certainly been accepted by the present government.

With the cut-off score in the National College Entrance Examination (NCEE) lowered substantially and with achievement tests for graduating elementary and high school students now being administered, the emphasis in testing has since shifted from admission, into assessment.

Our aim is no longer to determine who goes where or who does what, but more importantly, to pinpoint weaknesses - and from there to implement remedial measures.

Assessment has also influenced the directions Philippine education has taken since the present government assumed office in June 1992.

Recognition of educational assessment as undertaken by both the government and the private sector has pointed the direction towards which educational reform has been carried out.

Recently adopted policies such as those expanding the school calendar, increasing student contact hours, and allowing private schools to allot more time for language, mathematics, and science - have all been influenced to some extent by test outcomes.

The present government recognizes the need for continuously measuring the performance of our students. Only through such efforts can we institute appropriate reforms, towards turning out graduates capable of competing effectively in an increasingly competitive world and building a competent work force to achieve national development goals.

### **Problems in Educational Assessment**

Those of us who are involved in educational assessment and research, whether in this country or in the Asia-Pacific region or for that matter in the world at large, encounter basically the same problems.

These are the fundamental questions that we struggle with and work long and hard for answers. These are basically questions of accuracy and reliability.

These fundamental questions include the following:

- \* Do the instruments we develop and administer actually measure what they are supposed to measure?
- \* Do they have built-in biases which we are not even aware of?
- \* Do they serve to preserve certain social divisions?
- \* How have some instruments fared in predicting performance?
- \* Are the results which eventually guide policy and operational decisions accurate?

These are primarily the concerns of such testing agencies as the NETRC and CEM. They are, however, also the concerns of policy-makers and of the very people who may benefit or suffer from how educational assessment is practiced and how results are utilized.

Certainly the government and testing agencies have received a certain amount of flak concerning the administration of such tests as the NCEE and the National Medical Admission Test (NMAT).

Even the recently administered achievement tests for graduating elementary and high school students were not spared from criticism, especially from the private sector.

In a country like the Philippines, finding reliable measures may be even more difficult because of the issue of language.

The use of English, a language associated with the country's political and economic elite, has been said to favor these elite over the less advantaged sectors, especially in seeking admission to the more prestigious institutions of higher learning.

Developing truly reliable instruments is by no means an easy task. It certainly cannot be accomplished overnight.

Continuing research will be needed in order to improve existing instruments and to develop new ones which will give us a clearer picture of the performance of our educational system as reflected in the performance of our students.

Surely I do not need to stress the importance of knowing where we are right now in order to arrive at where we want to go. Reform would be meaningless and perhaps even disastrous if they are instituted in the wrong areas for the wrong reasons.

### Challenges For Educational Assessment

The development of accurate and reliable instruments is an ongoing effort being undertaken by both government and the private sector.

Three areas, however, are of vital concern in our assessment efforts in the Philippines. These are the areas of language, mathematics, and the sciences, where weaknesses have been found among Filipino students.

Another area of concern is vocational and technical education, which has not received enough attention thus far from our test development agencies. This area however, is essential for national development, and reliable instruments will have to be developed if we are to build a corps of skilled craftsmen and technicians.

Note that we are not even speaking of advanced skills. For the moment we are more concerned with the *basic skills* that will allow our graduates to learn more and to compete effectively in both the domestic and international markets.

In a region of the world which promises to be the center of academic activity in this decade and in the coming century, educational assessment will certainly have a role to play in maintaining the level of development which some countries have attained and in reaching that level which other countries aspire for.

Competition will not slacken. On the contrary, competition is likely to increase. Through educational assessment, however, each country may be able to determine the direction it should take in terms of education, which would allow it to maintain or increase its competitiveness in its most important resource - its people.

Hopefully, the presentations and discussions in this conference will help make those directions clearer to us all.

## ORGANIZING FOR INTERNATIONAL ASSESSMENT OF EDUCATIONAL SYSTEMS IN THE ASIA-PACIFIC REGION

Abraham I. Felipe  
*President, EDUVISION 2000*

On behalf of the Board of Trustees of the Center for Educational Measurement, I am pleased to welcome you all to this First Invitational Conference on Educational Assessment. I am aware that you have been invited because of expertise, for the intention of the organizers is to get leaders, no less, in the field of educational assessment in the Asia-Pacific countries. We are therefore honored by your presence.

Allow me to immediately call attention to the name of this conference: *First Invitational Conference on Educational Assessment*. The intention therefore is to make this neither the only conference nor the last. It suggests an intention to have more.

I am calling attention to this because to the cynical the first two objectives of this conference, especially the first, may be accomplished after this First conference without having to call a second and a third. But the first statement of objective is more than the usual ritual or protocol of international meetings. We believe that it is a condition that we must satisfy before we will be able to attain the others.

But to assure you about our hopes for continuing meetings, I call your attention to objectives 3 and 4. The third speaks of evolution, and the fourth of association. The organizers definitely have a longer time-frame, and I hope you also share the value of having other meetings.

I would also like to call attention to the subject of this and, hopefully, other conferences: *educational assessment*, we all know, covers many things. It could refer to assessing the individual in all his aspects, for educational purposes. It could refer to assessing the teacher in his subject and pedagogical competence. It could refer to assessing the curriculum in terms of currency, scope and coverage. It could refer to assessing school management in terms of efficiency and effectiveness. It could refer to assessing the whole educational system or any of its subsystems in terms of learning outcomes and other educational and non-, or better still, extra-educational objectives for which they were set.

I suppose that we are not referring to all of these types of assessment. Instead we seem to think of an assessment that, according to the rationale of this Conference, is pertinent to policy making, that bears on the improvement of the quality of educational systems of countries. This excludes therefore assessing the individual *qua* individual, not because this is not important but because the need for that type of assessment can be suitably met by other mechanisms different from international associations and invitational conferences. The rationale also speaks of

testing, which, together with the sponsorship of this conference by the CEM, suggests that this will be on assessments that use educational tests. This effectively excludes assessing pedagogical skills, curriculum adequacy and management effectiveness. We seem to have come here to talk about assessing educational systems with the use of tests.

The need to undertake this type of assessment needs little defense. The view that education is needed for a country to develop is widely shared. That our countries have very large current investments in our educational systems is a well known fact. Therefore, the need to monitor the performance of our educational systems to ensure that our investments do not go to waste, is a duty. Political interference, a practice that reduces the vitality of our schools, is common, but we lack data that could warn about the harm it inflicts. Among the testing professionals, the proposition that educational assessment would help improve education is at the level of conviction.

All these argue for assessment. But it is the nature of the existing technology of educational assessment that the measures we take do not have absolute meanings. Their meanings come from frames of reference that we adopt, against which measurements are compared. Whether scoring is norm-referenced or criterion-referenced, an external reference is needed to provide an interpretation. It is like the process of triangulation in determining one's position, except that it is even less exact.

Thus, even though our countries have undertaken national assessments of their systems of education, their interpretations have been denied that important dimension that an international framework could provide.

If these are acceptable premises, then it will be worth our while to look at what we can do in this conference towards some international projects of educational assessments, even as we remain committed representatives of our respective country's interests.

We can think of other reasons why Asia-Pacific countries would have need for international assessments. As commerce and trade increase among our countries, an inevitable development, we will have the need for systems that could place foreign pupils in school levels according to their preparedness. International assessments, together with the process of establishing curricular equivalency, could be of help. Perhaps something could also be said about the benefit in development that each country would ultimately derive from a collective effort that contributes to improving each other's educational system.

Thus, I believe, to undertake international educational assessments is useful and desirable. To think about it now is also timely.

I do recognize that we differ in our readiness for, and disposition towards, this undertaking. These differences will affect the timing and extent of participation in an international project.



Beyond calling attention to the value of an international project, and the opportunity to give birth to one now that we are together, I do not intend to do more advocacy. I realize that participation in such undertaking will be completely voluntary. It could not be anything else.

So I will just spend the remaining time to ask a few questions about our respective circumstances. The questions are meant to help each one of us assess the degree of our readiness, disposition, and capacity for national educational assessment at the international level. The questions are not meant to be answered aloud. They are meant to help each of us individually assess privately if we can satisfy the requirements of an international assessment in this region of the world.

1. National assessments are useful only if they are undertaken or at least blessed by the government ministry in charge of education. They could not be usefully undertaken by universities, academics, educational testing practitioners, or private individuals completely independent of ministries. These groups tend to attend to issues that are less urgent to those responsible for systems of education. They tend to give too heavy an academic orientation to the exercise. We can each ask ourselves two questions: is there a conscious need within my government to assess my country's system of education? Does my government already have an idea about how to use the results of an assessment?
2. National assessments, as explained earlier, could be enriched with the use of an international framework. The latter provides a dimension national assessments could not otherwise have. Questions: Will the relevant authorities in my country agree to this? Do they ask questions about our respective systems of education that could be answered more adequately through an international study -- questions about performance in general, or performance in specific areas such as science and mathematics, competencies of specific age levels, about the curriculum, etc.? Will they agree that it is feasible to get these information through such study? Will they welcome an international project? Will they support it?
3. Testing technology, among others, can be used in assessing the system. Question: Is the use of tests in assessments acceptable to the authorities in my country?
4. There are many possible oppositions and obstacles to implementing smoothly an internationally framed national assessment. Some are administrative, others political. And some administrative obstacles are motivated politically. Implementation requires sufficient government backing to ensure logistics, to persuade parties to cooperate and to gain acceptance (such as gaining entry in classrooms). Questions: Will the government give it enough importance in such a way as to make it politically acceptable? Will

there be people who could be assigned to the project who have enough know-how and influence to remove administrative stumbling blocks?

Some opposition will not surface at the start of the study but at the end when the results have already been known. The opposition may invoke technical adequacy. Can we involve respected professionals who may insure the results against questions of inappropriateness and/or technical deficiency?

5. It will be difficult to mount an assessment that could address all the concerns of the ministries of education of all countries in attendance in this conference. Even the ETS, an institution with a vast experience in assessment, limited itself to assessing competence in only two areas (science and mathematics) in the first National Assessment of Educational Progress. We, here, should also approach our task in humility. Will our governments understand this constraint, or will they expect more than what a first international project could likely deliver?
6. Many questions could also be asked on the nitty-gritty of day to day administration.
  - 6.1 Do you have an existing organization that has experience in large scale testing, to which day to day affairs could be assigned?
  - 6.2 As mentioned earlier, the role of ministries of education is important. It is desirable to involve ministry personnel in the exercise. Will they welcome this involvement? Are there non-government bodies that could help the government especially on matters that may sometimes not be available in ministries such as the more esoteric requirements of measurement, sampling and research design, test development and construction, and data analyses using computer technology?
  - 6.3 Results of assessments may not be flattering to a ministry. They may indicate deficiencies in areas for which they see personal and official accountability. This usually becomes a source of problems in protecting the integrity of the results, which undermines the usefulness of the study. It is therefore important that ministry personnel feel no threat because of the outcomes of the assessment. Can our governments make firm commitments not to take the results against school administrators, pupils or teachers?

The least of our worry should be the cost. I understand that many an internationally conducted assessment had been covered by simple grants from ministries. This suggests that none of our governments would go bankrupt from supporting a project that gathers data from a couple of thousand of students. Neither am I raising questions on technical



matters -- like item format, norm- versus criterion-referenced scoring, applicability of item response theory, the virtue of contextualized assessment. These are important but there is a proper time to raise them. Now, it is only time to assess the feasibility of making an international project.

Assessing the feasibility of the idea should not only consider conditions back home. It should also consider the conditions in the here and now, and by this I mean our own (that is, this group's) readiness and disposition for the undertaking. It is important to explore the group's readiness or disposition because this group could be a catalyst or a backbone of the exercise.

Again, I am not in a position to provide answers. But I want to ask the questions that could help assess the feasibility of the idea from our own standpoints.

Do you think that an internationally framed national assessment is possible now? If not, do you think that it is desirable to have in the future? What can we do between now and the future, so that possibility may be realized?

Is there a place for an international association such as the one being contemplated in this Conference, that could keep the fire burning, as it were, while we wait for a more opportune time? Should this association be of individual professionals? Should the network include universities and research centers as institutions? To what extent must it have government blessings now? What will it do in the meantime?

# MEASUREMENT OF EDUCATIONAL OUTCOMES IN BASIC EDUCATION: THAI EXPERIENCE

Sangob Laksana

*Department of Curriculum and Instruction Development  
Ministry of Education*

The basic education in Thailand is the 6-3-3 system: six years for primary, three years for lower secondary, and three years for upper secondary education. Education is compulsory and free for all starting from the pre-primary education up to the completion of lower secondary school. (See Figure 1.)

Based on the newly revised 1990 national curriculum for basic education, the primary schools and the secondary schools are empowered to adapt the curriculum and to develop new courses in the local curriculum including the learning system that are responsive to the needs, interests, and abilities of the students.

The goals of the revised curriculum address the following points:

## **Developing oneself**

1. Possessing basic knowledge and skills
2. Being healthy (physically and mentally)
3. Being able to solve problems
4. Being able to develop
5. Possessing good moral ethics, emphasizing self sacrifice for the benefit of the group
6. Possessing good attitudes toward work
7. Possessing knowledge in keeping up with technological advancement

## **Developing one's career**

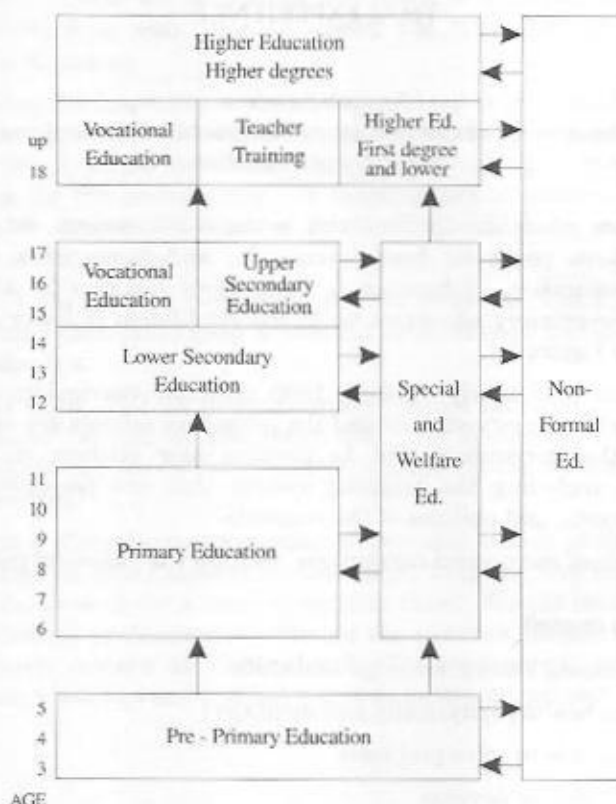
8. Possessing the skills and the right attitudes toward any type of honest occupation or career

## **Developing one's society**

9. Carrying out one's duties as a good citizen of one's family, one's community, and one's nation

These three points are continuous and related to each other. They are also considered as the framework for educational planning, management, instruction and evaluation.

Figure 1  
Outline of the Education System - Thailand



### Standard Setting

The national curriculum development is conducted by the National Committee on Curriculum Development through the coordination of the Department of Curriculum and Instruction Development (CID). The national curriculum is regarded as the educational standard because it specifies clear learning objectives (expected learning outcomes), course descriptions, recommended learning activities, the list of main and supplementary textbooks, and evaluation methods.

### Control of Quality Standard

In order to ensure the standard of education and to maintain national unity and identity, all schools in both the government and private sectors are required to implement the national curriculum and to follow guidelines in teaching-learning activities, learning materials, and measurement of learning outcomes.

To maintain the quality of education, the CID department has enforced the following measures:

1. Regular monitoring and supervision through regional and provincial coordinators are provided at the school levels.
2. National assessment of learning outcomes at primary education (Grade 6), lower secondary education (Grade 9), and the upper secondary education (Grade 12) is annually conducted.
3. Teachers are encouraged to improve the quality of teaching through the in-service training and the provision of prototype innovative materials including the measurement instruments and techniques by the CID. Teachers' self development in innovative delivery systems and the satisfactory learning outcomes are used as credits for professional advancement.

### **Levels of Measurement of Educational Outcomes**

The three levels of measurement of educational outcomes are as follows:

#### **1. Measurement at School Level**

The main purpose is to determine the educational achievement of the various courses taught in schools. The set of learning objectives and the design of measurement are controlled and approved by the school cluster committee that is composed of representatives from ten to twenty schools in the same geographical area. The measurement instruments developed by the teachers are used in assessing both cognitive and noncognitive learning outcomes. Item analysis and item banking conducted at schools are also encouraged. Other measures such as regular formative evaluations and summative evaluations are useful methods for remedial sessions and for placement purposes, respectively.

#### **2. Measurement at Local Level**

Measurement at this level involves the provincial and the regional offices in the assessment of educational outcomes as part of the accountability evaluation. The local authority may assess the learning outcomes of certain subjects at the end of the school year of each grade, or of the complete level (Grade 6, 9, 12). The measurement will provide the feedback for supervision and corrective measures.

#### **3. Measurement at National Level**

The national assessment conducted by the CID Department aimed at providing the information for determining the standard attainment, charting the student's progress and success on significant competencies, and personal development. Standardized

testing instruments have been developed on each well-defined behavior on operational competencies. The responses of students in various competencies are transmuted into proportions based on the proportion of correct or satisfied answers. The cut-off points of judgment according to the specified competency were moderated by the teachers and curriculum experts. These cut-off scores are used for categorizing the students into 3 levels: Level 1 for low achievement, Level 2 for moderate achievement, and Level 3 for strong achievement. The number and percentage of students who met the specified cut-off scores served as the index of competency attainment. Average scores across all competencies transformed into percentages are also used as the overall indicator of standard attainment.

The information regarding the quality of learning outcomes is used by decision makers in resource allocation, development projects, professional development, and corrective measures.

### **Issues and Problems**

In Thailand, problems in technical aspects of measurement and evaluation are minimal. The prototype testing instruments and the teacher guides for self study and practice are available. They also provide computer software for item analysis, item banking, and data processing to the schools.

Some difficulties, however, are encountered in providing corrective measures to the students that showed weaknesses as revealed by their test results. Innovative delivery systems are needed to upgrade the quality of learning among the students, especially for the disadvantaged students.

### **Innovation in Measurement**

There are some new developments in measurement and evaluation of learning outcomes. The utilization of Item Response Theory (IRT) has been used to some extent. The information curves and the true ability scores obtained from one-, two-, or three-parameter item analysis model are observed as efficient.

Equating the test scores among schools that use different tests and equating the progress of certain students who took different tests in successive years have been found quite useful.

Utilizing the statistical moderation for adjusting the grade-point-average obtained from various schools that have different standards is another new development. The scores on Scholastic Aptitude Test have been used as the basis for adjusting the school GPA's into the same and comparable scale. This adjusted GPA's will be used as a part of information for determining the ability of students who want to enter universities.

## ASSESSMENT IN ELEMENTARY EDUCATION IN THE PHILIPPINES

Marcelina M. Miguel  
*Bureau of Elementary Education*

The elementary education program has always provided for the assessment of pupil progress in learning. The measurement and evaluation of pupil achievement have been largely the responsibility of the classroom teacher. Test results are communicated to the pupils to inform them of their achievement in the class. These serve as the basis for deciding later whether they have passed or failed in the subject or grade. No examinations external to the class were given regularly or annually whether at the end of each grade or at the end of elementary schooling by the national, regional, division, district, or school office as basis for giving marks to pupils. It is only this school year, 1993-94 that an order from the national office (DECS Order No.30,s.1993) has been issued requiring all Grade VI pupils to take the National Elementary Achievement Test (NEAT).

The aforementioned DECS order states that taking the NEAT will be a requirement for graduation; however, a low NEAT score by itself will not necessarily prevent a pupil from graduating, since it will form part of the general average of each pupil "with the weight of one academic subject transmuted into percentage." The NEAT is in line with the thrust of the Department of Education, Culture, and Sports (DECS) to improve the quality of elementary education since it is intended to be the assessment that will provide a basis for measuring improvement in pupil achievement levels.

The NEAT consists of a battery of achievement tests of the multiple choice type which will cover the four areas of (1) English, (2) Mathematics, (3) Science, and (4) Heograpiya/Kasaysayan/Sibika. The construction of the tests will be similar to the national achievement tests that have been given to the sample regions in previous years in order to provide continuity over time. The tests were developed by the National Education Testing and Research Center (NETRC) in cooperation with the Bureau of Elementary Education (BEE) and were administered on August 31, 1993 to all Grade VI pupils in public and private elementary schools.

### **Assessment in The Elementary School System**

The role of assessment as one of the critical components of the goal-oriented instructional system for mastery learning has been emphasized in the implementation of the New Elementary School Curriculum (NESC) in the sectoral Program for the Decentralized



Educational Development (PRODED) which started in SY 1983. Alongside the grade by grade implementation of the NESC, testing programs to assess pupil performance were established at different sub-national levels of the school system as directed in MECS Order No.33,s.1983. So, evaluation committees at the regional, division, district, and school levels were formed by PRODED and all research, evaluation, and guidance supervisors attended a one-month intensive training at INNOTECH on different phases of assessment in the school system. They in turn trained the school heads in their respective schools at present as shown by the BEE monitoring teams' reports.

### **Assessment of Pupil Performance in the Curriculum**

The present curriculum (NESC) which was introduced in PRODED in SY 1983 was developed with the systems approach as the guiding principle. Basically, such a curriculum has in its design several components relative to a given set of objectives, i.e. the Minimum Learning Competencies (MLC), the relevant content, teaching processes, and appropriate assessment procedures that are all combined in an empirical manner to produce a viable and sufficient instructional system for mastery learning. This requires an expanded view of assessment where the emphasis is being shifted to gathering information about the learner's readiness for instruction as well as for his progress in the process of mastering what has to be learned. Furthermore, data on which decisions are based relative to the effectiveness of the instructional program in giving the student opportunity to mastery learning are sought.

Tests and other assessment tools which are to be used throughout the learning system are designed according to the goal-oriented instructional model for these purposes:

1. Measuring the input competence of the learner in relation to the learning task to be taken up.
2. Assessing the degree to which the learner has the competencies that are prerequisite to the mastery of learning tasks.
3. Diagnosing the learning style and the learning rate so as to best accommodate the individual learner.
4. Assessing the progress of the learner in order to introduce changes that will enable him to perform in the expected way.
5. Assessing the degree to which the output performance of the student meets predetermined objectives.
6. Determining the specific deficiencies in the quality of instruction.

Criterion-referenced tests have been used in determining the pupils' mastery of the curriculum's educational objectives so the development of these tests have been the one of the main topics of training workshops

and seminars among teachers and school heads since the 1980's. The Bureau of Elementary Education has developed an item pool for testing the important objectives in all subjects in the curriculum for all grades.

### Measures of Pupil Progress in Learning

To ensure continuous assessment of pupil progress, the Bureau of Elementary Education developed prototype tests which the supervisors of instruction in the field can a) use in the training of teachers on test development and b) use as samples when they prepare regional, division, district, and school tests. These tests are on: 1) readiness in language and numeracy for those entering school for the first time; 2) achievement in the different subjects in all grades (I-VI); 3) performance (psychomotor) skills; 4) listening comprehension and speaking skills; 5) mental ability; and 6) affective behavior, as attitudes.



## ASSESSMENT FOR ADMISSION IN HIGHER EDUCATION: THE PHILIPPINE PERSPECTIVE

Lucila F. Tibigar

*National Educational Testing and Research Center*

The Philippine Educational System follows a ladder-like sequence with defined number of steps in every stair case going to the topmost stairway, the first staircase composed of six steps in all public schools and most private schools termed as the elementary education level; the second case called the secondary or high school level is composed of four steps; but before proceeding to the last case of tertiary level there exists a set of steps analogically, the mezzanine or post secondary level. This post secondary level has its significance in terms of assessment for admission although not included within the purview of the regular education levels because its inclusion as one stair case in education defines what higher education in the Philippines perspective is all about.

Graduates of every level are given certificates of graduation called diploma. In the case of post secondary education, they are given technician certificates since what graduates from this level finished are technical-vocational courses which are highly specialized but maybe obtained in one or two or three years. As mentioned earlier, the inclusion of post secondary education among the staircases in the educational system of the Philippines has its significance in the assessment system in the country. First, it provides opportunity for the non-qualifiers to continue their schooling by taking technical- vocational courses; and second it serves as the preparatory ground for entrance to the degree courses or higher education after passing the college entrance examination.

Tertiary or higher education therefore, is defined as the third level in the Philippine Educational System offering four or more year degree-courses with a diploma awarded upon completion of the degree requirements with the corresponding Special Order. To date, there are a total of 1,038 schools offering tertiary or higher education degrees distributed in five school types namely: 1) the Public or Chartered Colleges and Universities with 150 schools, 2) Public or DECS Supervised with 92 schools, 3) Public Community Colleges with 29, 4) Private Sectarian Colleges and Universities or those managed by religious groups with 245 schools and 5) the Non-Sectarian Colleges and Universities run by private individuals or group of individuals other than the religious groups with 522 schools in operation.

Admission to degree courses in these schools requires passing the National College Entrance Examination or NCEE which qualification standard is expressed in terms of cut-off score. (6:np)

## **The National College Entrance Examination (NCEE): The State Assessment Tool For Admission to Higher Education**

The National College Entrance Examination or popularly known as NCEE has its beginning with the issuance of Presidential Decree 146 requiring a national entrance examination for high school graduates seeking admission to degree programs necessitating a minimum of four year study. As a result, entrance to college which required qualifying in this examination began in 1973. From then on, the NCEE has been used as a screening tool for entrance to a degree course. The NCEE is designed as a general scholastic aptitude test to assess abilities and skills that develop, generally and gradually over a long period of time, from what one has learn both in school and outside.

### **The Development of the NCEE**

The development of the first NCEE test started with a small scale pre-survey and documentary analysis zeroing in on the determination of the content of the test. A number of gainfully employed professionals in various occupational groups were asked about their opinion regarding the abilities needed for college success. Although the responses gathered varied widely, two groups of abilities were identified, namely: 1) Mental or Intellectual and 2) Personal-Social. It was decided that the mental component is more essential to college success hence, the identification of the following specific mental abilities to constitute the college entrance examination:

1. communicating one's ideas
2. expressing oneself in writing and speaking
3. comprehending what one reads
4. organizing logically one's ideas
5. doing mathematical exercises and problems
6. abstract reasoning

After identifying the vital components of the NCEE, the NCEE underwent several stages of development which are still being followed to date in the development of the tests being administered yearly to graduating high school students.

### **The NCEE Cut-Off Score**

The cut-off score is the criterion percentile rank of scores which would qualify an examinee to pursue at least a four-year degree in college. The first NCEE cut-off score was set at the 25th percentile and gradually increased to 50 and remained as such from 1986 to 1991. However, in 1992 it was reduced to the 19th percentile for the main reason of providing equal access to quality education to the great majority of graduating high school students and those who failed the higher cut-off score of 50th percentile in the previous years.

The cut-off score of 10th percentile was set per DECS Memorandum Order No. 113, series, 1993 however, the Honorable DECS Secretary enjoined all colleges and universities to set their respective NCEE cut-off scores higher than the prescribed cut-off.

### **Significant Findings of Studies Conducted on the NCEE**

Through the years of its administration as an assessment tool for admission in higher education, many evaluative studies have been conducted. More significant are those published in the Measurement, Evaluation, and Research Journal (MER Journal), Vol. 1, No. 2, April - June 1988 issue.

In most cases, changes came simply in the inclusion of the NCEE result as an additional requirement for admission. For a few schools, this meant more selective admission to some program (e.g. Education). At least 5 schools stopped administering their own tests to all applicants "to support scores provided by the NCEE for better evaluation".

In general, the NCEE appears to be predictive of academic performance in college. Its predictive power is increased, however, when combined with other independent variables. Tests like the OLMAT and DAT share a certain proportion of variance with NCEE.

The NCEE overall score correlated only moderately ( $r=0.40$ ) with the fourth year high school grade average in the study of over eight hundred students (Ibe, 1985a). The Mathematics test correlated with the fourth year high school grade at  $r=0.35$ . Of the NCEE subtests, Reading Comprehension almost consistently came out the most predictive of college performance. The NCEE subtests scores correlate significantly with grades in the tertiary level curriculum. As in the first and second year college, reading comprehension ability is the first predictor of third year college performance.

There is a positive correlation between the NCEE and the Philippine Board Examination for Teachers (PBET) ratings of teacher examinees in the 1989 and 1990 PBET indicating that the NCEE is a predictor PBET performance. (RED-NETRC, 1990 and Ibe & Siclon, 1991)

Prior to the administration of the NCEE, students, regardless of their academic aptitude for college work, were admitted to four-year degree courses. This resulted to a low college retention rate among those who enrolled. With the NCEE, only those who have the potential to do collegiate work have been allowed to pursue at least four-year degree courses. This, in effect, has resulted to substantial improvement of the retention rate by as much as 5.26 percent yearly.

This improvement in the retention rate brought by the NCEE similarly resulted to substantial financial gains, not only to the society in

general, but also to families in particular. The testing fee required for each student is just a token amount compared to the benefit (of saving college fees) that could be derived from knowing the student's scholastic aptitude prior to enrollment in a collegiate work. Thus, the NCEE provides the critical information for making decisions at this crucial level (i.e., the family) at a very small amount of money.

The National Educational Testing Center (NETC), now NETRC, conducted a study on the "Impact of the NCEE on the Quality of Education." The first phase of the study hypothesized that as a result of the requirement of qualifying in the NCEE instituted in 1973, the quality of tertiary education has improved. The indicator of quality adopted was the percentage of freshmen students who passed the course (4-year degree) they are pursuing. The percentages of the students passing during the first semester of the school years under study -- 1973-1974 (prior to the school years under study) -- 1976-1977 (three years after the institution of the NCEE) - were compared.

Eleven colleges and universities were randomly sampled throughout the country. The percentages of students passing each course were computed and the increase or decrease in percentage was determined, and likewise, computed by course. The results showed a higher percentage of freshmen who passed during the first semester of 1976-1977 as compared to that of the first semester of 1973-1974. An average increase of 5.26 percent in students passing a course was noted after the NCEE was instituted. This finding implies that as a result of the NCEE, those who would most likely not succeed were initially screened out leaving only those with greater likelihood of making good to pursue their respective course in college, thus improving quality of student input.

The second phase of the study was expanded. Using the same sample schools, the percentage of freshmen students who passed during both semesters of school year 1973-1974, that is, before the implementation of the NCEE, was compared with each of the enrollments during the four school years after its implementation, from 1974-1975 to 1977-1978.

Generally, the findings revealed that in all the courses pursued, there was an increase in the percentage of passing in 1974-1975 and every year thereafter. The increase in the percentage of freshmen students passing a course ranged from 0.86 percent in Social Work to 7.29 percent in Architecture. The biggest increase was observed in 1974-1975 in all the courses under study except Social Work.

The results of the second phase, therefore, confirmed the results of the first - that the NCEE screened students so that those who are more likely to succeed in college are identified. Consequently, there seems to be some degree of improvement in the quality of education.

## **Admission Practices**

Institutions are established with different visions and missions. One thing in common with these institutions for higher learning is their desire to contribute to academic excellence in their own community and in the national and international levels. Generally when they formulate their own policies, their immediate concern is the immediate environment. It follows that admission practices in assessing students who should be admitted to tertiary education vary. Some of the practices are:

1. Schools with more applicants than they can accommodate in each degree course offered, screen their own applicants by administering their own admission examination/s in addition to the NCEE. Before these applicants are registered, they are required to submit their school records as proof of academic excellence. The instrument used in screening applicants is a specific aptitude test. Graduating students take the entrance test in three or more schools to make sure that they are admitted in one of the excellent scores. Generally, students admitted in these schools have very much higher NCEE scores than what has been set by the DECS.
2. Other institutions administer their own entrance test in addition to the NCEE. Selection is not as rigid as the first group. Students admitted in this group have NCEE scores ranging from NCEE cut-off score for the year to the 70th percentile.
3. Some schools admit applicants who are NCEE qualifiers with a grade point average of 80 in the fourth year high school.
4. Majority of the schools admit all applicants to their school provided they qualify in the NCEE and they obtained at least a grade point average of 75 in the fourth year high school. There are negligible case when some students were admitted without qualifying in the NCEE and their graduation is delayed.

## **Recommendations**

In the light of what has been presented, I would like to offer these suggestions:

1. Use the results of the NCEE in the different areas (RA, VA, RC, MA) with the subject rating in the Student Form 137 (high school report card) in admitting students to a particular degree course.

Many psychologists believe that performance in tests of Mathematics Ability and Verbal Ability would give a gauge of one's general intelligence.

Likewise, the potential for success in certain college courses can be indicated by good performance in the following areas of the NCEE:

#### COLLEGE IN GENERAL

Verbal Ability  
Mathematical Ability/Reasoning

#### OFFICE WORK OR CLERICAL/STENOGRAPHY

Verbal Ability  
Reading Comprehension

#### JOURNALISM/ADVERTISING

Verbal Ability  
Reading Comprehension

#### ACCOUNTING/COMMERCE

Mathematical Ability  
Reasoning Ability

#### BOOKKEEPING, STATISTICS, CARPENTRY, TOOL MAKING, LAB. ASSISTANCE

Mathematical Ability  
Reasoning Ability

2. Make available to guidance counselors/admission officers statistics on manpower supply and demand which can be used in guiding students what course to take in relation with their aptitude test results. This will reduce the mismatch between graduates and the demand in the industry.
3. Administer a unified/integrated entrance examination in the different tertiary institutions. Each institution can set their own cut-off scores. This will free the students and their families from paying multiple entrance examination fees and from taking numerous admission tests to higher learning and at the same time they will be admitted in the school of their choice.



## ASSESSMENT IN THE SINGAPORE SCHOOL SYSTEM

Andrew Poh Sui Hoi  
*Nanyang Technological University*

The purpose of this study was to develop and adapt, as well as investigate, the validity of the Self-Estimate Aptitude Scale (SEAS) for secondary school guidance in Singapore. The SEAS is theorized to be multidimensional, having distinct subscales that the items were written to measure. The SEAS together with six objective aptitude tests and two career maturity variables, career planning and career exploration were administered to a stratified random sample of 1568 students from the Express (Fast track) and the Normal (Slow track) streams. There were 739 Secondary 3 (Grade 9) and 829 Secondary 4 (Grade 10) students.

Exploratory and confirmatory factor analyses showed that the SEAS has a well-ordered six-factor structure at the first-order level and a one-factor structure at the second-order level. This configuration is consistent across groups. All the items of the SEAS have high discriminating power and the reliability estimates of the subscales are also reasonably high.

The Campbell and Fiske's approach to multitrait-multimethod (MTMM) analysis showed that traits as measured by the SEAS and the aptitude tests possess convergent as well as discriminant validity to a limited extent. But there were more differences than similarities when the two types of measures were compared. Confirmatory factor analysis of the MTMM matrix revealed by the traits, perceived aptitudes as measured by the SEAS and demonstrated aptitudes as measured by aptitude tests, are quite different. Further examination of validity by canonical correlation analysis showed that the SEAS subscales are more highly associated with the career maturity variables of career planning and career exploration when compared to aptitude tests. Path analysis using LISTEL models, showed that the SEAS subscales have a significant positive path with the career maturity variables, indicating that students who score high on the SEAS are more mature in terms of career planning and career exploration. The aptitude tests have a weak negative path with the same career maturity variables, indicating that those who score high in aptitude tests need not be more mature in career planning and career exploration. Multivariate analysis of variance revealed that the numerical, spatial and mechanical scales of the SEAS were associated with significant differences in student ratings across gender. Significant differences were also observed for the numerical, spatial, and clerical scales of the SEAS across streams for the Secondary 4 sample. These results add credibility to the validity and usefulness of the SEAS in the context of career guidance programmes.

# THE DEVELOPMENT OF NATIONAL MONITORING SYSTEM OF EDUCATIONAL ACHIEVEMENT IN CHINA

Meng Hong-wei

*China National Institute for Educational Research*

## **The Demands For The National Monitoring System Of Educational Achievement In China**

Education now is a top priority in social and economic policy for modernization in China. The intellectual demands of work are growing. Economic competition around the world and co-operation between China and developed countries open new opportunities for China to enhance the effectiveness and productivity of its educational system.

In May 1977, in a talk to members of the Central Committee, Deng Xiao Ping set the tone for the interpretations of modernization with his remarks that "it is necessary to improve education at every level...." This means to "raise the standards of education at the same time as we make it available to more and more people." (Deng Xiao Ping, 1984, p.49)

At the fourth plenary of the Sixth National People's Congress, the report on the seventh five year plan specified that the administration of the educational institutions should be enhanced, gradually establishing a systematic educational evaluation and supervisory system.

Mr. Liu Bin, Vice-president of the State Education Commission stressed the need "to establish a set of scientific and reasonable standards for evaluating the quality of the basic education." (Liu Bin, 1989)

All these statements expressed by the top political and educational leaders stressed the necessity to increase the efficiency of the educational system in China. So the tasks faced by the educational researchers is to ask if the capacities of youth in the country are sufficient in both quantitative and qualitative respects, to know what degree of educational progress is being made and how its education system is compared with others. This information then can be used to guide school improvement, to help in priority setting, and to meet increased demands for accountability.

For an effective educational system, the quality of primary and secondary schools requires detailed, specific and reliable information on the performance of the educational system including student learning achievement. It is very desirable that the educational accomplishments of students are as high as possible with respect to the stated aims.

Between 1980 and 1984, a team of experts from the World Bank visited China twice, and reported that China's achievement in education since 1949 had been unmatched among developing countries of the same



income level. The experts also pointed out, however, that educational plans to move into the universal nine-year basic education entails a complete analysis of China's basic education system. This would require: a) a survey of physical facilities; b) a large-scale data collection survey using carefully designed sampling techniques leading towards an input-output analysis (IEA type study); and c) a provincial or even prefectural breakdown of school enrolments by grade, sex, ethnic group, and rural/urban residence as related to age group and the composition and distribution of the population (World Bank, 1985, p.17).

The World Conference "Education for All", held in Jomtien, Thailand in March 1990 highlighted that in order to achieve the goals of education for all, a country's information and knowledge base is vital in preparing and implementing a plan of action. One major implication of the focus on learning acquisition is that systems have to be developed and improved to assess the performance of individual learners and delivery mechanisms. Process and outcome assessment data should serve as the core of a management information system for basic education (Education for All: Appendix 2 Framework for Action). With regard to the enrolment of school age children, the net enrolment rate in 1990 has reached 97.83% across the country. In the rural area, this is 97.40%. With a national net enrolment rate of 98 percent, the great improvement of teaching conditions, one of the major challenges in the context of the efforts to universalize compulsory education is to ensure that all pupils at different stages of compulsory education attain the specified learning outcomes.

In February 1993, the Central Committee of Chinese Communist Party and State Council have issued a national program for educational reform and development in China. The program listed that education, i.e. basic education, needs to be expanded and the quality and effectiveness of schooling should be standardized and enhanced to a higher level.

The National Monitoring System will mainly focus on the learning acquisition as a significant factor in the Education for All (EFA) initiative:

*One key issue on which we must focus if we are seriously to pursue the goal of universal primary education - or, indeed, any of the EFA targets - is the minimum common level of learning that must be achieved by.....learners. (J.P. Grant)*

For a long time, the proportion of graduates who passed the tertiary entrance examination (i.e., the National Unified Entrance Examination for Institutions of Higher Education) and were admitted into universities or colleges, has always been considered as the most prestigious final evaluation of achievement of students and a measure of the quality of teaching in secondary schools by parents, teachers, and students themselves.

The introduction of the high school certificate examination system is trying to separate the selection examination from the quality monitoring examination. So far there is rarely a monitoring system on basic education, although there are some major monitoring programs.

### **The Major Monitoring Programs In China During 1985-1993: The Second IEA Science Study**

The SISS (Second IEA Science Study) is an international comparative educational research conducted by the International Association for Evaluation of Education Achievement (IEA) in the 1980's. It attempts to document both within each country and across the countries taking part, the status of science education programs in the schools in the 1980's (Keeves, 1984).

The China National Institute for Educational Research (CNIER) joined IEA in 1984 and took part in the SISS. The SISS in China was a first attempt at using a large-scale survey as a means of monitoring performance of the school system.

The SISS was conducted in China in 1985 with a target population consisting of students in Grade 3 in junior secondary schools in Beijing, Tianjing, and Taiyuan. A stratified two-stage probability sample of schools and students was drawn. (Meng Hong-wei, 1986).

Data on student performance of science achievement, student characteristics, home background, teacher and school characteristics were collected from five sets of tests and questionnaires, i.e., cognitive tests, student, teacher, school questionnaires and the opportunity to learn questionnaire.

The average mean score of Chinese students in the international core test (2M) can be compared with 26 other systems. Dr. Meng Hong-wei (1988, 1991, 1992) has presented the results of this study.

In 1987, the Division of Basic Education, State Education Commission launched a national survey of mathematics teaching at the end of junior secondary level, i.e. Grade 3 in junior secondary schools.

The State Education Commission conducted a survey in the Chinese language and the teaching of mathematics. Its purpose was to provide scientific and realistic suggestions for improving the quality of mathematics teaching in junior secondary schools, promoting the reforms of mathematics teaching and teacher training. The survey used a three stage random sampling design to draw 50 thousand students enrolled in 605 schools from 15 provinces. Data on student learning achievement, mathematics teaching, teaching materials and teacher characteristics were collected.

## UNESCO/UNICEF Joint Project

The Action Plan formulated to promote Education for All in China envisages the strengthening of supervision, inspection and evaluation of the quality of primary and junior secondary education. In order to develop an appropriate monitoring mechanism for primary education and to build up a national capacity, the Division of Basic Education, the State Education Commission was involved in the UNESCO/UNICEF Joint Project on "Monitoring Education for All Goals - focus on Learning Achievement." The project is in line with the strategies set out at the World Conference on Education for All which stressed the need to focus on actual learning achievement and outcome.

The target population in this project was specified as the Grades 4 & 6 students in the 6-year primary schools. Eight provinces were selected. They are Gallein, Habu, Shaanxi, Jiangsu, Hunnan, Guangdong, Sichuan and Yunnan. Twenty-four thousand students from 1,200 schools of each grade were involved. The tests of Chinese Language, Mathematics and selected life skills and a student questionnaire were taken by each student.

The general objective of the study is to develop and apply methods and indicators for monitoring of learning achievement of pupils at the primary level, on a regular and continuing basis, for the purpose of making available to decision makers relevant information which can form a basis for information of policies and programs for qualitative improvement of primary education.

There are some other large-scale surveys focusing on the monitoring of students achievement. A summary of these activities is presented in Table 1.

### The Proposals For a Nationwide Monitoring System

Now, when we think of a national monitoring system, the experiences and personnel trained from these previously projects are very important. A future nationwide monitoring system should take this factor into account and to involve as many as possible experts from all related areas.

As the Assistant Director General for Education in UNESCO, Collin N. Power in 1991 pointed out that "the basis of all monitoring is the establishment of efficient approaches to the collection, interpretation and use of all kinds of data. Thus, countries need to develop their own capacities for information-based management of education...". In China, under the authority of State Education Commission, there are various systems to collect all these information such as Statistics Department in Planning Division, Management Center of Educational Information, National Supervision Group, Pedagogical Research Departments in Basic Education Division. Now these different systems are working separately and from different views.

**Table 1**  
**The Major Monitoring Projects in China During 1985-1993**

Subject	Time	Target Population	Objectives	Sample	Sponsor
Moral education of youth and children in China	1991-1993	Grade 5 in Primary Grade 2 in Lower Secondary Grade 2 in Senior Secondary	The status quo of moral values and moral behavioural development	Non-probability typical sample (5,718 students)	Badi Foundation
National student Constitution and Survey	1985	7 - 22 years	The physical shape, function, quality of youth and children	Multi-stage random sample (864,872 students)	State Education Commission; State Physical & Sports Commission; Ministry of Public Health; State Nationality Commission
The Second International Science Study	1985	14 years Grade 3 Junior Secondary	Learning achievement in Physics, Chemistry, Biology and Geography	Two stage random sample (2,917 students)	The State Education Commission
Mathematics language Foreign language (English)	1985	Grade 3 Junior Secondary	Learning achievement in Math, Chinese Language Foreign Language (English)	The adjustment sample	The State Chinese Education Commission
Mathematics Chinese Language Life Skills	1993	Grades 4 and 6 Primary	Learning achievement in Math, Chinese Language and Life Skills	Multi-stage probability random sample (48,000 students 1,200 schools 8 provinces)	UNESCO/UNICEF The State Education Commission

It is very difficult to use these information directly for the purpose of a national monitoring of student achievement. But a common framework of national monitoring of student achievement worked out by these experts and close cooperation between them will maximize the utility of existing information from these systems. The national monitoring system should involve the experts from all these systems. A national committee can be set up to coordinate all these aspects.

## **The Proposed Organization Of The National Monitoring System Of Educational Achievement In China**

The National Educational Achievement Monitoring Committee will take overall responsibility for all activities of the educational achievement monitoring including communication with all-level Education Commission across the country, school districts, schools and other educational organizations; printing, distributing, collecting all materials for monitoring; scanning of most documents and entering other data to machine-readable format; initial cleaning of data; and providing the researchers with data files. It can also have the following sub-organizations such as:

1. **Assessment Review Committee.** The committee will be composed of teachers, administrators, teacher educators, members of the public scientific community, personnel from various branches in the State Education Commission, and representatives from educational researchers and curriculum developers.
2. **Item Review Committee and Provincial Interpretation Panels.** The committees meet periodically throughout the monitoring program to offer advice on and guidance for all aspects of the program and to review all materials at the various stages of editing and prior to their distributions to the field.

### **Target Population And Sampling Design**

In order to ascertain the extent to which the educational achievement have been made in the country, we need a representative sample of the target population in the country. So the sampling design is crucial for success of the national monitoring system. The sampling design should take into account the following factors: geography, population density, economic development and nationality. The sampling design also should take into account the homogeneity within a region or a school.

There is a lot of literature dealing with the random sampling designs. The problem is that in a country like China, the real random multi-stage sample is very difficult to implement because of very limited resources for educational research, huge population, and transportation to remote areas and so on.

The search for a statistically acceptable and realistic sampling design for a nationwide representative sample has been viewed as an important aspect in establishing a national assessment system for student achievement (Meng Hong-wei, 1991).

### **What should be monitored (Quality indicators)?**

The most important aspect of a national monitoring system will be what should be monitored. For a national monitoring system, the indicator should say something about the performance or behavior of an

education system and can be used to inform educational decision-making. Consequently, indicators are expected to reflect the condition of the system. To meet these goals, indicators must satisfy a number of substantive and technical criteria based on the experiences from the OECD international educational indicators project. For example, indicators should provide at least one of the following types of information:

1. information that describes the system's performance in achieving the desired educational conditions or results.
2. information about features of the system known to be linked with the desired conditions and outcomes.
3. information that describes enduring features of the system.
4. information that is relevant to educational policy.

Because of the complexity and diversity of the national education system, it is obvious that an individual indicator conveys limited information.

To compensate for the uni-dimensional nature of each indicator, a coherent set of indicators which provides a valid representation of the condition of a particular educational system is needed, not just a collection of readily available statistics. Ideally, a system of indicators should provide information about how individual indicator components work together to provide an overall effect. In other words, the policy and interpretative value of all the information to be gained from this system of indicators is greater than the sum of its parts.

To provide this overall picture, the selected indicators should be logically or empirically linked. The linkages should proceed from a model or framework that describes how the system works. The model by itself permits the broader assessment of an indicator's relevance. Drawing on the model, combinations of indicators can be used to diagnose current and future conditions. Further, particular combinations indicators can provide clues about what is wrong and what changes may lead to improvement.

In this aspect, there are a few educational researches dealing with a system of educational indicators or model on educational achievement in China. We should identify the important indicators which are meaningful for policy-makers and try to develop a model which can logically and operationally link them together. Meng Hong-wei (1991) has tried a structural model of factors related to student science achievement in China. This model has identified a set of important factors both for policy-makers and educational researchers.

Also these information cannot be best employed for policy purposes unless the educational system is systematically monitored over time and unless the information obtained about the system can be compared



internationally. In this sense the system of the indicators in the national monitoring system can be compared over time, and across nations.

### **Could We Have The National Standards For Student Learning Achievement?**

In 1992, a comparison among the provinces in terms of the graduate examination passing rate of junior secondary schools clearly showed the disparities among the provinces. Should we take the student learning achievement in developed areas as a national standards or other way? How can we set up national standards which will satisfy both the developed and less developed areas? How can we take care of the local needs in student learning and maintain the national standards as well? For a successful national monitoring system, we should take this factor into account and try to get some degree of the balance.

### **REFERENCES**

- Deng Xiao Ping. (1984). *Deng Xiao Ping's selected works*. Beijing: Foreign Language Press.
- Keeves, J. P. (1984). *The second IEA science study: A world perspective*. 25th IEA General Assembly Meeting, Singapore.
- Liu Bin. (1989). The certification of examination in senior secondary schools and reforms in basic education, "Adjustment of present syllabus of general senior schools and implement of certification examination of senior secondary schools" edited by The Division of Basic Education, State Education Commission. Chang Chun Press 1991.
- Hong-wei, M. (1986). "SISS Pilot Study in China", Inspection and Evaluation on General Education. Hubei Provincial Institute for Educational Research.
- Hong-wei, M. (1988). *The science achievement in secondary schools in China*. Paper presented in Annual Hong Kong Educational Research Association Conference, Hong Kong.
- Hong-wei, M. (1991). *The structural model of factors related to student science achievement in China*. Unpublished manuscript. University of Hong Kong.
- Hong-wei, M. (1992). The affective outcomes of schooling in China. *International Journal of Educational Research*, (17)3-4.
- The Central Committee of Chinese Communist Party and State Council. (1993). *The national program for educational reform and development in China*. China: People's Press.

- The State Education Commission. (1990). *Suggestions on the implementation of the certificate for examination system of general senior secondary schools* (document).
- The State Education Commission. (1991). *The guideline for the supervision and evaluation at primary and secondary schools*. Edited by Tao Xiping. Beijing: Education Science Press.
- World Bank. (1985). Long-Term Development Issues and Options. Annex 1: China. Issues and Prospects in Education, The World Bank, Washington, DC.
- Zhao Zi Yian. (1986). Report on the seventh five years planning. Office of Standing Committee of National People's Congress Conference. (Eds.) Collection of documents from the fourth conference of the sixth Chinese National People's Congress. People's Press.
- Final Report World Conference on Education for All: Meeting basic learning needs p 13, p 57, Jomtien, Thailand, 1990
- Nuttall D (1992). *The functions and limitations of international education indicators*. OECD International Education indicators: A framework for analysis. OECD Paris.



# EDUCATIONAL ASSESSMENT IN THE GUAM PUBLIC SCHOOL SYSTEM

Zenaida N. Natividad  
*Department of Education*

## Introduction

My task today is to give you a preview of educational assessment in Guam. I will do this in view with the factors and considerations that will come to play if Guam is to participate in an assessment at the international level.

## The Island of Guam

The island of Guam is a tiny patch of land in the vast Pacific Ocean. Guam is a territory of the United States of America. Residents of the island enjoy the same rights and privileges as any citizen of the United States of America. Historically, Guam's past is very similar to the Philippine experience: occupied by the Spaniards for centuries, followed by the Japanese, then finally the United States of America. If Magellan landed in the Philippines on March 16, 1521, this same explorer discovered Guam on March 6, 1521. This date is a public holiday on Guam - called the Discovery Day.

Guam is about 29 by 15 miles in land area with a very diverse population. There are about 39% local Chamorros (most of whom trace their ancestry to the Philippines), about 28% Filipino, 17% White non-hispanics, and the rest are other Asians such as Korean, Japanese, Vietnamese, Thais, Chinese, or settlers from the outlying islands such as Chuuk, Pohnpei, Saipan, Kosrae, Palau, Rota, Tinian, and other Micronesian settlers.

Though Guam is a U.S. Territory, it has its own Organic Act that embodies the local laws governing the land. The highest elective official is the Governor who is elected every four years. Guam has its own Legislature composed of 16 Senators and a Representative to the United States Congress. His name is Dr. Robert Underwood, former academic Vice President of the University of Guam.

There are three U.S. military installations/facilities: Andersen Air Force Base, Naval Air Station and Naval Station. I am mentioning this fact because the presence of U.S. military installations on the island have implications on public education:

1. The dependents of the military enter the public school system. Each military dependent enrolled in the public school means money to the Guam Department of Education. The Department of

Defense pays for the cost of educating every military dependent in Guam. The Guam Department of Education has a Federal Program office that supervises the implementation of federally funded education programs.

2. The presence of U.S. military bases in Guam makes the student population more diverse and somewhat transient.

In the late 1980's, the Compact of Free Association was signed between the United States and the Freely Associated States of the islands of Chuuk, Ponape, Kosrae, and others. The compact-impact agreement placed a variety of academic and administrative burden on Guam's educational system. For one thing, the people in these outlying islands do not have a sense about paper documentation. They came to Guam without passport, birth certificate, and immunization record. Most of them have very little or no schooling at all, and some do not even have permanent names. One day a kid enrolls in one school under the name "Sparkling New", the next day the same kid (with the kid's family) moves to another village, and enrolls in another school under the name "Sincerely Yours". Tomorrow, who knows where! For lack of a birth certificate needed to get themselves enrolled in the public school, some would even bring their mothers and ask the principal frankly and proudly, "Do I still need a birth certificate? Is my mother here not enough evidence that she gave birth to me?"

I will now brief you on public education in Guam.

### **The Guam Public School System**

The Guam education system follows the American education system and like the Philippines, its organizational structure is modeled after the military hierarchical structure.

The Board of Education is appointed by the Governor. The Director of Education is the Executive Secretary of the Board.

This organizational chart will soon be history as a result of the recent State of the Territory address by the Governor who announced that he will restructure the Department before his term expires in December this year.

The Guam Public School system has a current enrollment of 30,384 students in 21 elementary schools K-5, 6 middle schools 6-8, and 5 high schools 9-12.

Guam aligns itself to America 2000 which was initiated by former President George Bush and fully supported by President Bill Clinton. The Guam Department of Education is currently developing its local version of America 2000; we have given it a name: Guam Goals 2000.

While the Guam Goals 2000 is not yet in place, our educational mission and vision is specified in the blueprint for excellence developed

about 14 years ago. All curricular, instructional, and student evaluation activities of the Department are implicit or explicit implementation of the Blueprint for Excellence.

I will now turn to educational assessment in the Guam public school system.

### **Assessment in the Guam Public Schools**

Since Guam is a territory of the United States, we have access to a plethora of nationally normed achievement tests and a variety of personality, aptitude and attitude scales. Each school is given the liberty to utilize assessment instruments developed in the U.S. for various instructional purposes.

At the district level, however, the Guam Department of Education has a comprehensive testing program that consists of both criterion-referenced and norm-referenced testing. About 12 years ago, a public law was passed by the Guam Legislature that mandated the conduct of district-wide assessment. By law, every child enrolled in the public school system is supposed to be tested. Following the passage of this public law, the Guam Basic Skills Mastery Test or the BSMT was developed, pilot-tested, and now widely implemented throughout the public schools. The Guam Basic Skills Mastery Test is a locally constructed curriculum-referenced test that assesses mastery level in Language Arts and Mathematics in Grade Levels 2, 4, 6, 8, and 10. The Territorial Board of Education has identified 25 priority skills in Language Arts and Mathematics. The test items in the Guam BSMT were designed to test mastery in these priority skills. The Guam BSMT is administered by the Department twice each schoolyear: a pretest at the opening of school in September and a posttest at the end of school in May.

The comprehensive testing program also includes the administration of nationally normed achievement tests. About two years ago, the Department of Education administered the Science Research Associates test or the SRA, a nationally normed test. However, clamor for the need to assess not only basic skills but also higher order thinking skills, the Guam Department of Education pilot tested for the first time last year the GOALS test. This test assesses children's ability to reason, infer, make conclusions, or decisions based on complete or incomplete information. We have also just adopted the Stanford Achievement Test-8 or the SAT8, a nationally normed test used in some states in the mainland.

To compare our student's performance on a national scale, Guam participates in the administration of the National Assessment of Educational Progress or the NAEP. This year, our fourth grade students were tested in Math and Reading. An external examiner who is not a staff of the Department of Education conducts the NAEP test. The NAEP is not limited to student testing. Teachers and school principals also take part in the assessment. The teachers and principals respond to survey

questions that include teaching style and practices, availability of educational media within the school and the home, reading habits and preferences, and to some extent, educational philosophy as manifested by preference over one teaching practice or another.

At this point, you may be wondering, how do our kids perform on these tests? In spite of the fact that education enjoys the biggest slice of the budget pie, in spite of the presence of 28 or 56 station computer laboratories in the Guam public schools, in spite of the fact that our teachers are fully certificated, or of the fact that our Governor and our Legislature fully support education, in spite of all these....we have a sad story to tell as far as assessment results are concerned.

Like the rest of the mainland, we are a nation at risk. The only test we are really good at and that which we surely pass is called by the name "passing the buck".

If a checklist is passed today to survey the considerations that we need to address before we begin talking about national assessment on an international frame, Guam will most likely respond this way:

1. Do we have the blessings of the powers that will support the national assessment on an international frame?

Response: The likelihood that the powers will support this kind of endeavor may be determined by hierarchical logistic regression analysis based on the knowledge of the likelihood of how similar programs got in place in the Guam Public School System. Without the sophisticated statistical methods, however, perhaps the more succinct response is: Need we have to wait until election time? Because whoever sits as Governor decides what happens to public education.

If you cannot have direct access to the Governor, get a person of influence within the circle of the Governor. This person may or may not even be an educator.

2. Does Guam have an idea on how to use the outcome of assessment?

Response: Maybe. And even if it does have an idea, will Guam use the assessment outcome? With our comprehensive testing program in place for about twelve years now, educators are still grappling about how we must utilize the results.

3. Does Guam have professionally trained staff who can be involved in the endeavor and can defend the methodology?

Response: Yes. One of them is speaking to you right now.

## ASSESSMENT IN THE PRIVATE SCHOOL SECTOR: THE CEM EXPERIENCE\*

Esperanza C. Buen  
*Center for Educational Measurement*

In every society, education is an instrument for development. In the Philippines, the private education system is seen as uniquely Filipino contribution to the ways by which man fashions this instrument. While private education accounts for less than 10 percent of total enrollment in the elementary level, it does account for 36 and 80 percent of total enrollment in secondary and tertiary levels, respectively. With this in view, the problems of private education are not to be seen as problems of the sector alone. Government has a significant role in enhancing the status, welfare, and capacity of private schools to perform the part it has assigned to them (FAPE, 1986).

On this part, then, private education must be committed to the ideals of excellence and relevance to the imperatives of development. The practice of self-evaluation has been undertaken by many of these private schools in order to assess the quality of their services. Often, such efforts are couched within the system of voluntary accreditation where the schools can demonstrate the excellence of their programs, whether in terms of student achievement, the significance of their research efforts, or the value of the educational service that they provide.

It is within this context that the Center for Educational Measurement or CEM finds itself. CEM was established on September 1, 1978 in response to a widely acknowledged need for an independent testing and research agency that would help improve the system of admission and streaming in higher education, equalize educational opportunities through a system of scholarships and other forms of assistance, and upgrade academic standards through the use of measurement tools and data in institutional development and planning.

### Test Development

CEM's mission served as the impetus during the last fifteen years for the Center's test development activities which have resulted in the production and maintenance to date of, more or less, 50 educational test instruments. These are standardized tests classified into two general categories, *aptitude tests* and *achievement tests*. Aptitude tests are designed to predict success in future learning activities. Measures

\* The full text of this paper is published in the *Philippine Journal of Educational Measurement*, vol 5, September 1994, pp. 1-15.

developed under this category include the Philippine Aptitude Classification Test (PACT), the College Scholastic Aptitude Test (CSAT), the DECS- commissioned National Medical Admission Test (NMAT), and admission tests for allied health sciences, graduate courses, and scholarship programs. Achievement tests are designed as a measure of mastery, as a diagnostic instrument, and/or as a survey instrument. While aptitude tests are norm-referenced, achievement tests are criterion-based, although norms are also developed for the latter in order to maximize their utility.

Extensive measures are observed to safeguard the integrity of these tests. Various steps in test development are followed to ensure the qualitative and statistical validity of these measures. The items, especially, are continually evaluated and new forms are assembled to maintain parallelism in the distribution of item content, item difficulty, and item discrimination (Angoff & Dyer, 1971).

### **Centralized Testing Program**

The concept behind CEM's Centralized Testing Program is to offer all the services of a good testing program at the smallest budget possible (Asuzano, 1976). Through this program, schools that feel the need for assistance may enlist themselves as members, and in consultation with CEM, they further clarify their needs and identify the services available that would best fit their needs.

The Test Service office of the CEM undertakes the promotion activities and monitors test administration on the national level. It serves as the liaison between the Center's main office and a network of 23 test service centers located in all regions of the country. At present, the CTP of CEM services more than 400 schools all over the country.

### **Processing of Data**

After servicing the testing needs of schools through appointed examiners, the CEM retrieves test materials and does the centralized scoring of accomplished answer sheets. Through the years, it has acquired a modest degree of advancement in terms of its capacity to process test data through a computerized system. CEM also manages its own in-house printing which is tasked mainly to produce all test materials and to ensure the security and integrity of the contents of the test forms.

### **Instructional and Advisory Programs**

The CEM also provides professional assistance and services in data utilization and instruction to its members schools. Foremost among these



services is to help in the interpretation of test results and to assess the immediate uses of the data. It also conducts training programs to improve teaching methodology, curriculum planning, and in preparing plans of action for the improvement of programs for instruction, guidance, and admission.

Through its instructional and advisory programs, CEM adheres to its mission of assisting schools in upgrading their academic programs and in promoting quality education. The services range from continuing education through training and scholarship to consultative efforts in teaching strategies and information technology.

### **Research**

CEM regards research as an integral part of assessment. The thrust of CEM research is to contribute to the science and technology of measurement as well as its usefulness to the schools that it serves. Specifically, it generates new information for further refinement of assessment schemes, identify the factors that affect the output of programs, and evaluate the effectiveness of the interventions.

The Center also undertakes studies that monitor the psychometric integrity of its assessment instruments. This entails research studies on the validity of test results as used by the member schools.

In relation to this, CEM has undertaken evaluation studies that sought to determine effectiveness of programs and/or policies for purposes of making decisions on whether or not a program should be changed, continued, or terminated.

The prospective thrust of CEM is to undertake further studies on its existing instruments, do surveys on the school and classroom environments, and case studies on effectiveness of instructional interventions. The Center also wants to activate its role as a clearing house for educational research studies being done by its member schools. It can build up a research grid which, together with university-based studies, can serve as a valuable reference or information bank for researchers, students, faculty, and policy makers in regard to measurement and evaluation.

The Center perceives its role in the private education sector as an assessment body that lends support in the improvement of school programs and outputs. In some limited fashion, its sphere of influence has extended into the public sector, but always in the form of a cooperative venture with our counterpart in the government service.

## References

- Angoff, W.H. & Dyer, H.S. (1971). *The admissions testing program*. New York: CEEB.
- Asuzano, L.M. (1976). Concept of a centralized testing program. *FAPE Review*, Vol. 7, No. 2.
- Fund for Assistance to Private Education (1986). *FAPE's view of the role of private education in Philippine society*. (Pamphlet) Makati: FAPE.

## AN APPRAISAL OF TEACHING OF EVALUATION IN COLLEGES OF EDUCATION IN INDIA

G. Subramonia Pillay

*Department of Education*

*Madurai Kamaraj University, Madurai, India*

Examinations play an important role in the educational process. It would not be an exaggeration to say that the examinations have a determining influence not only on the system of education but on the standards of teaching and research and the careers of the millions of students. As such, the instructional system is ultimately based upon the method of evaluating results. The procedure of feedback and the revision on the results is again an important factor deciding the impact of the instructional system. In short evaluation is penetrating the entire educational system. The expert bodies have always been recommending various measures for strengthening the evaluative process because they find the existing system of examination and evaluation in this country not at all satisfactory to serve its purpose. No systematic study has been reported on the effects of the brief course on evaluation given to the B.Ed student teachers. Realizing the importance of such a study, the present attempt is made.

The objectives of the study are:

1. to find out the depth of content received by the student teachers of the B.Ed programme under different subjects
2. to find out the importance given to "evaluation" in the B.Ed programme.
3. to appraise the syllabus content of "evaluation" for the different subjects of the B.Ed course and
4. to suggest ways and means for effective teaching of evaluation in colleges of education.

To achieve these objectives, the following procedures were followed:

1. Development of an Achievement Test on Evaluation and administering it to the student teachers.
2. Content Analysis of the B.Ed Question papers
3. Content Analysis of the B.Ed syllabus meant for evaluation
4. Analysis of the Work Book prepared by the student teachers on Tests and Measurement

Structured interview was also conducted with Teacher Educators and Student Teachers.

### Major findings of the study

Among the nine colleges of education studied, St. Xaviers College of Education, Palayamkottai scored the highest mean achievement in evaluation and Dr. Alagappa Training College, Karaikudi scored the lowest mean score.

Compared with other groups, the achievement scores in evaluation of the student teachers who have opted for Test and Measurement as an elective were dispersed closer to the mean and therefore the groups were found to be homogeneous with respect to their achievement in evaluation.

Among the six optional subjects offered, the student teachers of English topped in scoring the achievement in evaluation.

Among the five aspects studied in evaluation, the student teachers achieved best in the two aspects such as, Types and Techniques of evaluation. They were poor in the construction of Tests, Characteristic of Testing and the Aims of Evaluation.

Student teachers of various colleges of education vary significantly in their achievement in evaluation.

No differential achievement was found between the Mathematics group and the groups of Natural Science and English student teachers.

The course on evaluation has been given significant importance in the University first semester examinations.

## MULTI-MEDIA EDUCATION: IMPLICATIONS FOR ASSESSMENT

Akira Ichikawa  
*Edogawa University*  
*Sociological Department, Japan*

In the future it is going to be called the "multi-media education" that is composed of various new media such as the satellite communications system, high definition TV, cable television, video disks, PCW recording, and CAI (Computer Assisted Instruction) system using microcomputer.

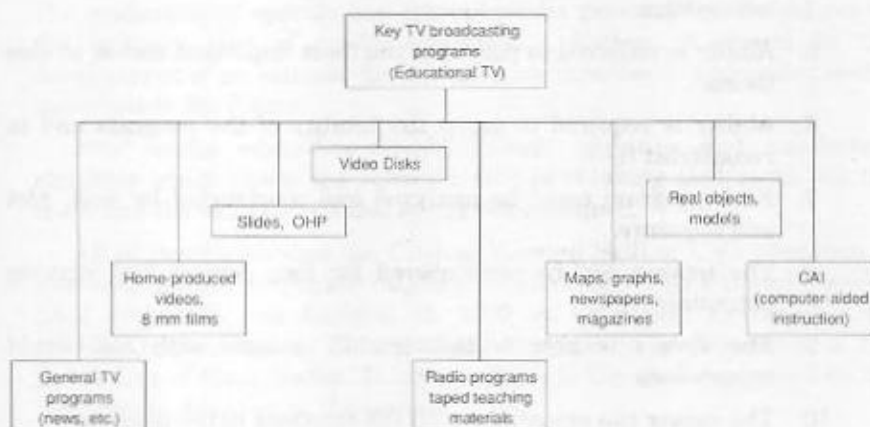
But even with growing competition with these media, conventional TV broadcasts still have obvious advantages in terms of interesting and emotional feeling in daily life. A research conducted by Toshiyuki Mizoguchi (Osaka University) in primary schools revealed that broadcasting education in the future is likely to combine multimedia learning system around the TV images (See Figure 1). He said that TV program is like a main dish among the routinary but indispensable foods for life existence while new media is like a delicious side dish for special effects.

TV school program still remains an important teaching material in motivating every child's scientific way of thinking. On the other hand, the new media with its multimedia use will assist in the concept formation of children in order to expand, fix, and access knowledge. This relationship between the TV program and new media is manifested in a case study made by Mazoguchi and Yosida about the broadcasting education in the primary schools of Kanazawa city. In this study, the respondents were shown tapes of "Green Earth", an environmental conservation program, and "We Are Global Family", a moral education program for motivation and emotional understanding. Then such learning was reinforced by showing graphics, maps, and statistics through microcomputer from educational database. The use of microcomputers in accessing the database is important for TV images in arousing pupils' motivation and emotional activity. TV images are not only daily and economic but also fresh and international. It is considered as the key of media system.

With the use of a video program, I made my own study about the multimedia model of instruction. The results of the study suggested a two-way recognition of TV images for the purpose of educational assessment: through the use of denotative and connotative meanings (See Figure 2).

Denotation refers to the common sense and the obvious meaning of sign in the first stage. This denotative recognition is the same for most children. The difference, however, lies in how the individual child perceived the connotative meaning of TV images.

Figure 1



Connotation refers to the meaning of the sign in the second order of signification. Image is also a language of communication and it expresses various form of signs reflected by educational technology. The sign meets the feelings and emotions of the users and the value system of their culture. John Fiske regarded this as meaning moving towards the subjective, or at least, the intersubjective. It is when the interpretant is influenced as much by the interpreter as by the object or the sign.

Roland Barths, a French philologist of semiotics who coined the word "connotation", pointed out that the difference between connotation and denotation can be seen in any images of photography. Denotation is the mechanical and objective reproduction of the image in the world at which the camera or video is taken while connotation is the human part of the process. It comes from the artistic sense of the producer and the director, technical decisions such as the camera focus, camera angle, etc. It also reflects the images of the inner words.

In other words, denotation is *what* is taken shot and connotation is *how* it is imaged. This is a different function.

These two meanings of sign are sometimes mixed by the educational process which only complicate the assessment and evaluation of multimedia effects. On the other hand, it is important to include the vivid informations and emotional expressions of the artist in evaluating Film and TV education programs.

The educational effects of TV images can be classified as follows:

1. The TV images can be clearly recognized and easily memorized by strong impressions.
2. The TV images provide an active and interesting learning process.
3. The viewer can distinguish changes in time, total images and partial image, montage editing by means of image language and vivid news of modern world.



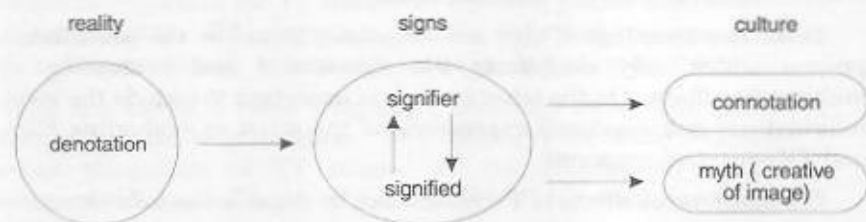
4. Ability is required to follow the structure of educational information.
5. Ability is required to point out the most important scenes in this thesis.
6. Ability is required to grasp the totality of the program and to reconstruct it.
7. Each program could be analyzed and constructed by shot, plot and sequence.
8. The images can be remembered for long periods and making potentiality.
9. The viewer is able to link the TV images with real world experiences.
10. The viewer can empathize with the emotions in the program.
11. Viewing stimulates and motivates psychological learning.

These areas of ability have been mentioned several times in broadcast education but now that various media come to compete, these skills assume more importance. Moreover, it is necessary for the children of our time to learn the high quality images to make it easy for them to understand the fascinating artistic images and vivid imagination of the more modern world in the future.

Systematic curriculum construction should also focus on areas of reading and writing as well as the appreciation of good emotionalized literature. Artistic works are encouraged especially among the young people for the enrichment of the humanistic personality.

Thus it seems that aside from the old media way of evaluating film and broadcast education, it is also important to rediscover the function of images through its connotation meaning of expression media and vivid understanding of the changing real world.

Figure 2



Fiske noted that the cultural influences of television goes beyond the "popular leisure activity" that is full of advertisement, mass communication, and style of life. Its more immediate purpose is to develop sensitive viewers who can appreciate qualitative programming and a critical sense of form.

S.M. Masterson insisted learning practical skills for media education. The evaluation of specific and current media presentation should not be the primary goal of media education. Rather, it should be the development of an attitude for critical consciousness in addressing media materials in the future.

New media education should include cognitive and intellectual elements which stress the scientific way of thinking that seeks out the truth and the valid facts about media circumstance.

All of these embodies the Critical Viewing Skill or CVS education of Professor James A. Brown (Alabama University). The Critical Viewing Skill education was founded in 1982 at Grunwald, Germany when UNESCO sponsored an International Symposium on Education of the Public Use of Mass Media. It is in response to the challenge posed by the multimedia situation at the present.

## PROFESSIONAL OCCUPATIONS IN THAILAND: A CASELOOK OF REGISTRATION AND LICENSURESHIP

Wilars Singhawisai

*Secretary General*

*The Thai Office of the Civil Service Commission*

In order to minimize the likelihood of damage or danger affecting the health and safety of others as a consequence of insufficient knowledge of relevant principles, facts, or procedures, all Thai graduates of professional areas need to exhibit reasonable minimum job competencies. Like many other countries, professional practitioners in Thailand, particularly those of the traditional professions such as law and medicine, beyond their formal education specific "proficiency bars" of professional requirements have been used before a registration permit or a license is granted prior to their entry to the occupation in practice. The bars in considering the qualifications evidence to justify the graduates' suitability are determined through a post-degree education and experience, examination, and training specifically demanded by a particular occupation which will be elaborated more shortly. For example, a degree in law should acquire either a 1-year internship with an attorney office under a close supervision of an experienced attorney then take a written test or first passing a written test then a 6-month practicum before being granted an attorney work permit.

The rationale of requiring a certain kind of efficiency bars for the professional occupations emerges from three essential characteristics of education, experience, training, and examination: (1) they are believed to be the best indirect predictors of those prospective practitioners, (2) they are readily verified and virtually guaranteed by an authorized, either public or non-government organization, and (3) they are widely presumed to be practically associated with professional job performance success (Ash, Johnson, Levine, McDaniel, 1989).

This paper is intended to describe and differentiate various approaches of considering eligible professionals in Thailand. First, we brief you on areas of professional occupations under study and responsible organizations. Second, we spell out minimum qualifications and proficiency bars required by each profession in considering the issuance of a registration or a license. Subsequently, we reveal statistics of professionals in each occupation. Then finally, we conclude with related problems and plausible solutions. As an overview, Figure 1 describes the flowchart of professional paths elaborated in this paper.

### **Professions and overseeing agencies**

Professional practices in Thailand are under a close supervision of mostly state-run offices attached to the ministries as shown in Table 1,

Table 1  
Thai Professions Under Study and Responsible Areas

Profession	Under the Supervision of	Year State Act Initiated (Year Revised)
1. Attorneys	The Law Society of Thailand* c/o Ministry of Justice	1957 (1985)
2. Auditors	The Board of the Supervision of Auditing Practices, Ministry of Commerce	1962**
3. Barristers at Law	The Thai Bar*, c/o Ministry of Justice	1965
4. Engineers and Architects	The Board of Control of the Engineering and Architectural Professions	1962*** and 1965
5. Medical Doctors	Medical Council*	1923 (1982)
6. Nurses and Midwives	Nursing Council* c/o Ministry of Public Health	1961 (1985)
7. Pharmacists, Dentists Asso. Med.	The Board of Medical-Related Registration	1936 (1987)
8. Veterinarians	The Board of Veterinary Practitioners Ministry of Agriculture & Cooperatives	1962

\* As a non-government organization

\*\* The latest revision of some board regulations in 1992

\*\*\* Some ministry regulations updated in 1969 and 1977

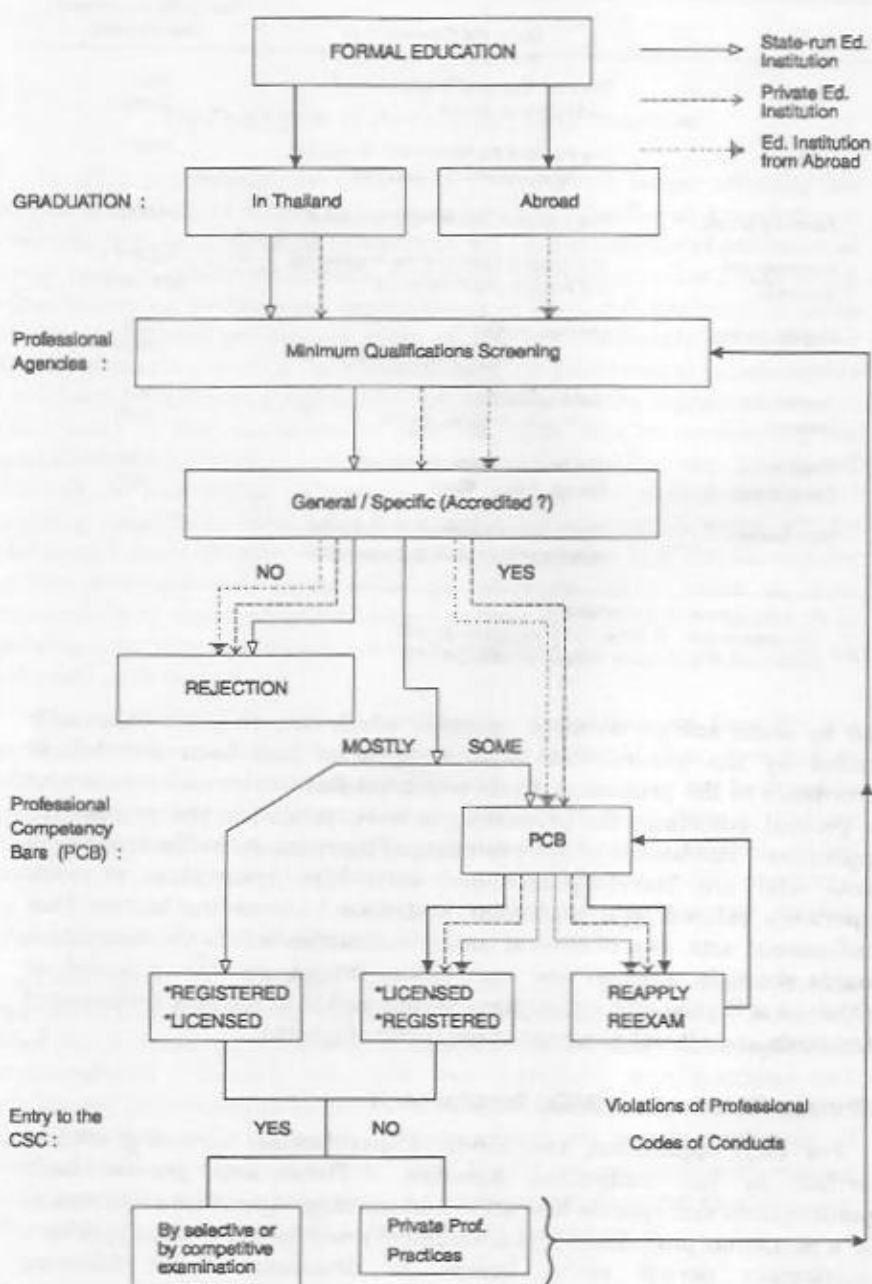
and by some non-government agencies which are, in part, financially funded by the government. These agencies had been founded in accordance to the professional acts which set forth rules and regulations as general guidelines for permitting a work permit to the prospective applicants. The boards of the professional agencies generally consists of those who are knowledgeable and earn high recognition of work experience related to a particular profession. According to the Thai professional acts, two of several major authorities which the occupation boards share in common are: a) to issue, extend, provoke, suspend, or withdraw a license or a registration permit and b) to monitor professional standards and ethical practices of the member groups.

#### Minimum Qualifications (MQs) Requirements

For each application, two kinds of qualification screening will be verified by the authorized agencies. These are: general basic qualifications and specific education and training experience requirement for a particular profession. It is a burden of proof for those who apply for a registration permit or a license to demonstrate the following qualifications:

1. Thai nationality
2. Twenty years old or over on the application date

Figure 1  
A Flowchart Representation of Professional Paths in Thailand



3. Psychological and physical sanity
4. No arrest or sentence record(s) by a finalized verdict
5. No indecent or unethical behavioral records which can be considered as violations of codes of ethic of a particular profession, and
6. No ailment symptoms of contagious diseases which may interrupt regular professional functioning

In addition, it should be noted that the Thai Law Society further requires that the applicants have no bankruptcy records and must not be a government (local) official. In the case of obtaining a medical degree outside the country, the Medical Council states that an applicant needs to verify that he/she lives in Thailand. The Nursing Council prohibits from applying for a registration permit those applicants whose license were rejected twice and those who do not practice nursing career continuously for 2 years.

Table 2  
Competency Requirements for Profession

Profession	Requirements					
	Prior Education (Years)	Written Exam	Hands on Test	Practicum/ Internship	Board's Review	Interview
1. Medicine	6	1 <sup>d</sup>	1	1*, 2	1	-
2. Veterinary	4 - 6	1 <sup>a,d,e</sup>	1 <sup>a</sup>	-	1	1 <sup>a,d,e</sup>
3. Dentistry	4 - 6	1 <sup>c,d</sup>	1	-	-	-
4. Pharmacy	4 - 5	1 <sup>c,d</sup>	1 <sup>c,d</sup>	-	-	-
5. Architecture	5	1 <sup>e</sup>	-	-	1 <sup>e</sup>	1
6. Barristers	4	1 <sup>b,c,d</sup>	-	-	-	1
7. Asso. Med.Sc.	4	1 <sup>c,d</sup>	1	-	-	1
8. Physiotherapy	4	1 <sup>c,d</sup>	1	-	-	1
9. Nursing	4	1 <sup>b,c,d</sup>	-	2	-	-
10. Engineering	4	1 <sup>e</sup>	-	-	1 <sup>e</sup>	1
11. Auditing	4	1 <sup>b,c,d</sup>	1	1	-	-
12. Attorney	4	1 <sup>b,c,d</sup>	-	1	-	-

Note I. 1 = required  
2 = encouraged  
- = none

a = Diploma or a vocational certificate  
b = Bachelor's degree or higher (State University)  
c = Graduates as b but from a private unit  
d = Graduates as b but from abroad  
e = Degree equivalent to b.

II. Applicants for professions # 2, 3, 5, and 9 need to pass a law examination test first before proceeding to taking the subject- and the performance tests.

\* required for the applicants who finished the study from India



As a specific basic requirement of education and experience, the applicants have to earn a relevant professional degree—normally at a baccalaureate level as appeared in the second column of Table 2 from an educational institution, either in Thailand or abroad, accredited by the Office of the Thai Civil Service Commission. Thus, a transcript of records in conjunction with a study program is generally required. It is interesting to note that medical doctors who obtained their degrees from abroad are encouraged to get an internship or an externship at least one year with a medical office in Thailand. With an exception for those who received their degrees from India, they are also required to obtain the internship experience before being eligible for the proficiency tests. It is clear that the different requirements imposed are up to the professional acts and the overseeing boards to set up or update the regulations. The processes of updating the professional requirements of dentists, doctors, pharmacists, and nurses are now underway.

### **Professional Proficiency Bars**

The recommendations concerning use of specific education or training requirements beyond formal education, demanded by a particular profession are logically analogous to the assessment of specific job-related knowledge or skills necessary to perform a job successfully. As a matter of fact, it is assumed that there is one best education and experience requirement appropriate for a profession. Basically, Thai applicants who received relevant professional education and training from a state-run university in Thailand are eligible to obtain an affiliate membership status or the registration permits. More stringent procedures apply for those who received professional education/training from abroad and from private education institutions in Thailand.

After screening the applicants' preliminary qualifications, different requirements have been imposed for different professions. As shown in Table 2, it is obvious that the acquisition procedures do vary. However, across all professions, a written examination is the top favorite choice used in determining the optimal qualifications. At a lesser extent, hands-on performance tests and interviews are the next popular predicting tools to be used especially by the medical-related sciences.

*Non-compensatory Hurdle Approach.* Essentially, the applicants need to exhibit a minimal level of proficiency on the requirements sequentially. Better proficiency on one or more requirements cannot compensate for deficiency on another. As an example, applicants for dentistry, pharmacy, and associated medical sciences need to pass a law examination with a minimum of 60% first before being eligible for taking several subject tests. They are allowed to take each subject no more than three times within a 2-year period. To be specific, 13 subject tests are required for dentistry, 7 for the pharmacy, and 5 for the associated medical science. Then, they are finally qualified to take 3, 1, and 4 performance tests for

each profession, respectively. It should be noted that for some professions like architecture, engineering, medicine, and veterinary science, the professional boards still have a final say in reviewing the applicants' qualification suitability.

*Passing Points.* Generally, the cut-off scores of the grand total scores - as the passing or acceptable indicants of professional competencies - are set at 60% except for those of the architects, the engineers, and the lawyers which required minimally at 50% of each hurdle. Thus far, there is no carefully researched evidence to support the linkage between the passing point and career success.

The applicants for auditing, law, nursing, and pharmacy are encouraged to have tutorial classes offered by university departments or the professional society in order to prepare them for the upcoming examination.

### **Vital Statistics of Professional Groups**

The data presented in the following sections are from the records (some since 1962 up to late 1993) of the overseeing organizations previously mentioned in Table 1. Among twelve professions under study as revealed in Table 3, a large number of professional clusters with nurses, engineers, and attorneys, respectively with a smaller number of those in the health-related areas. The evidence may reflect two plausible explanations. First, the data tell us the limitation of medical-related science education provided by the government or private sectors due to resource constraints. Second, the hardship in obtaining the medical-related degrees and stringent competency requirements for the licenses may be another reason. Current statistics show that only 10% of Thai graduates finished their higher education in the health-related sciences and 22% in the science and technology areas (Thai Ministry of University Affairs, 1993).

*Entry To The Civil Service System.* A public office is authorized by the Thai civil service commission to recruit the professionals of health-related and sciences and technology areas since the Thai bureaucracy system is now facing a shortage of supply of these professionals. However, for the social science professionals like attorneys, barristers at law, and auditors, they need to pass a competitive examination, not by selection.

### **Conclusions: Problems, Plausible Solutions, and Trends**

Compared to the total population, professional occupations are increasingly in a high demand for Thailand amidst her rapidly social and economic development towards a newly industrialized country in Asia. In principle, some major concerns worth mentioning here are (1) how to maintain our professional standards with ethical practices and (2) how to attract the new professionals and retain the incumbents to work for the government. In practice, the corrective as well as the preventive

Table 3  
Accumulated Number of Professionals (up to 1993)

Profession	Total Number with Licenses or Registered (numbers issued in 1993)	
Architects <sup>a</sup>	4,316	(262)
Asso. Med. Scientists	2,848	(246)
Attorneys	31,660	(1,484)
Auditors	3,080	(113)
Barristers at Law	53,203	(n/a)
Dentists	4,614	(290)
Engineers <sup>b</sup>	55,835	(1,360)
Medical Doctors	19,295	(886)
Nurses & Midwives <sup>c</sup> - Level I	65,453	(4,006) <sup>d</sup>
- Level II	53,280	(4,540) <sup>d</sup>
Pharmacists	9,292	(484)
Physiotherapists	992 <sup>e</sup>	(100)
Veterinarians	6,359	(n/a)

Note: a - Three Membership categories as b but with work output at least one project per year.

b - The membership status for engineers is classified into three categories, namely:

Level I - Affiliates (with bachelor's degree)

Level II - Professional members (as level I + 3 years professional work control experience)

Level III - Senior professionals (as level II + 7 years professional work control experience)

c - Level I means nurses with a 4-year bachelor's degree.

Level II means nurses with additional 2-year degree to acquire the bachelor's degree.

d - 1992 data record

e - the number is the available data since 1988.

measures are essentially related to the direct compensation of professional careers, a shortage of supply, revisions of work stature of the Thai bureaucracy system, and their career advancement in the public services.

Due to more attractive compensation of the private sectors, the Thai civil service commission is now in the process of updating the compensation systems equitably about 80% of the pay grades paid by private sectors. Moreover, monthly position allowances are given for professionals. We try to fairly establish a rapid and flexible career

advancement for the professionals with respect to their major education and training experience.

With regard to a number of health-related and technology-related professionals, there is an optimistic trend that the Thai government does allow private educational institutions to share the responsibility on condition that the educational standards of excellence have to be continuously maintained. It is also hopeful that in the future, written professional tests in *English* may be able to help Thai professionals who have been living abroad and not much familiar with the written tests in Thai. Lastly, it is essential that the professional affiliations, whether it is a government or non-government agency, apply current measurement and evaluation principles to their practices in screening for the qualified professionals.

### References

- Ash, R.A., Johnson, J.C., Levine, E.L., & McDaniel, M.A. (1989). Job applicant training and work experience evaluation in personnel selection. *Research in Personnel and Human Resources Management*, 7, 183-226.
- Law Society of Thailand (1993). *History of the law society of Thailand and the lawyers act of 1985*. Bangkok: The Law Society of Thailand.
- Sompong, P. (1992). *Medical-related registrations in Thailand*. Bangkok: The Office of the Permanent Secretary for Public Health.
- Thai Ministry of Agriculture & Cooperatives (1992). *The veterinary surgeons act of 1962, its revision, and the board's regulations of 1992*. Bangkok: The Board of Veterinary Practitioners.
- Thai Ministry of Commerce (1993). *Laws related to auditing and the auditor act of 1962*. Bangkok: Division of Commercial Registration.
- Thai Ministry of Interior Affairs (1991). *The engineering act of 1962, its revisions, and related rules and regulations*. Bangkok: The Office of the Board of Control of the Engineering and Architectural Professions.
- Thai Ministry of Public Health (1982). *The medicine act of 1982*. Bangkok: The Medical Council.
- Thai Ministry of Public Health (1985). *The nursing and midwifery profession act of 1985*. Bangkok: The Nursing Council.
- Thai Ministry of Public Health (1993). *Handbook of registration permit examination for the first class pharmacists*. Bangkok: The Division of Medical-Related Registration.
- Thai Ministry of University Affairs (1993). *Annual Report*. Bangkok: Division of Policy and Planning.

## THE SECONDARY EDUCATION DEVELOPMENT PROGRAM (SEDP) EVALUATION

Eugenia M. Moraleda  
*Curriculum Development Division*  
*Department of Education, Culture, and Sports*

The Bureau of Secondary Education is presently conducting an evaluation study designed to investigate the extent to which objectives of the Secondary Education Development Program have been achieved. The objectives include:

- a) the improvement of quality secondary education graduate
- b) the improvement of the internal efficiency of the secondary education system
- c) the expansion of access to secondary education; and
- d) the promotion of equity in the allocation of resources especially at the local level.

What is SEDP? SEDP is the response to

- a) the need to continue pupil development started by PRODED
- b) research findings indicating a need to improve student performance in science, math, and communication arts
- c) findings that ineffective teaching, inadequate facilities and instructional materials contribute to unsatisfactory student performance
- d) the need to improve policy making and increase the internal efficiency of the secondary education system

SEDP components include curriculum development, staff development, physical facilities development, technical assistance (local and foreign fellowship), and special studies/researches. Furthermore, SEDP focuses on (1) curriculum reform, (2) quality textbooks/teachers' manual on 1:1 ratio, (3) science and work education equipment, (4) staff development (short/long term), and (5) research studies on NCEE, barangay high schools, career guidance, teacher recruitment, etc.

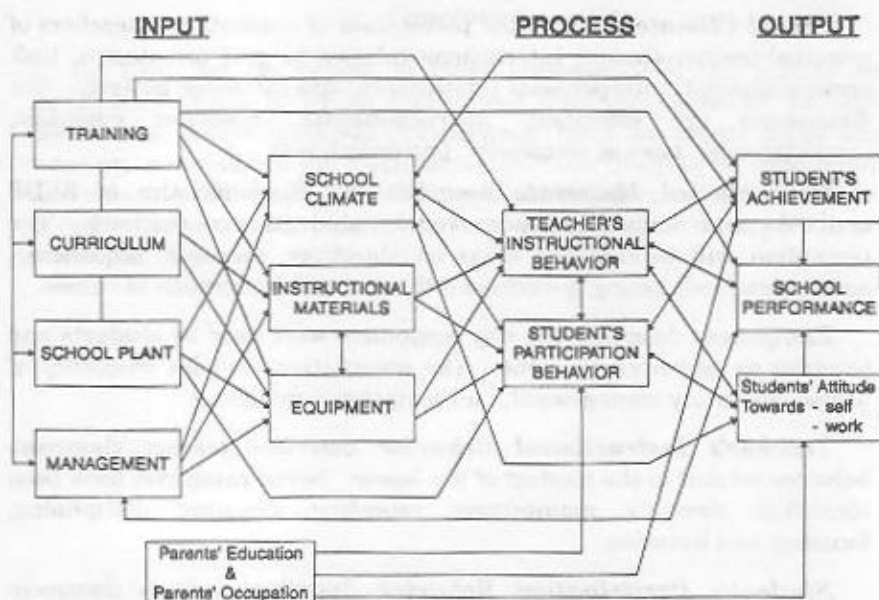
The evaluation study will adopt a causal model (which recognizes the interplay of school and non-school factors in producing educational outputs). The study will focus on the educational outputs (student achievement and attitudes, and school performance), the effect of SEDP inputs on the educational outputs, and how the inputs were processed - thus producing the outputs. This is shown in the evaluation framework (See Figure 1).

Student achievement and attitudes, and school performance are said to be influenced by both school and non-school factors. Since the purpose

of the study is to determine the impact of SEDP inputs, we have chosen variables representative of these inputs, namely:

- a) training (teachers, supervisors, and principals)
- b) curriculum (The New Secondary Education Curriculum and SEDP textbooks and teacher's manual)
- c) school plant and equipment
- d) management

Figure 1  
SEDP Evaluation Framework



The process variables include equipment (utilization of equipment), school climate, teacher's instructional behavior, and student's participation.

For the non-school variables, we have parent's occupation and parent's education. These measures capture the kind of intellectual environment in the home, the motivation and support the student is getting from his family.

### The Variables

**Training** describes the SEDP training program for trainers, teachers, and administrators as perceived by the trainee and observed by process observers. Six areas will be the points for observation: trainer characteristic, training design, context mastery, conduct of the training, and application of learning.



**Curriculum** describes the characteristics of the NSEC as to relevance, contemporaneity, balance, vertical coherence, and continuity as perceived by experts, teachers, and supervisors.

**School Plant** details the physical characteristics of the school as perceived by teachers, school personnel, and administrators. The observation will focus on the three areas: site characteristics, building characteristics, and furniture/equipment characteristics.

**Management** describes the manner regular office (regional, division) interactions were carried out relative to the attainment of SEDP objectives. Six characteristics were identified: leadership, motivation, communication, decision making, goal-setting, and control.

**School Climate** refers to the perceptions of students and teachers of principal-teacher-student interactions relative to goal orientation, task accomplishment, interpersonal relationship, and influence patterns. Six dimensions are identified: instructionalism, academic emphasis, supportiveness, morale, reciprocity, and orderliness.

**Instructional Materials** describes the characteristics of SEDP textbooks and manuals as perceived by students and teachers. The perception will include the areas on objectives, concepts, sequencing, appropriateness, pacing, government thrusts, and integration of values.

**Equipment** describes the way equipment were used by students and teachers as perceived by some. The observation includes frequency of usage, laboratory management, and competency matching.

**Teacher's Instructional Behavior** describes teacher classroom behavior related to the conduct of the lesson. Seven categories have been identified: directing, maintaining, repeating, grouping, disciplining, focusing, and lecturing.

**Student's Participation Behavior** describes students' classroom behavior during class time. Six categories have been identified: responding, initiating, consulting, obeying, not attending, and performing.

**Student Achievement** refers to the scores of students in the NCEE and the BSE achievement tests.

**School Performance** describes the school in terms of performance indicators such as (participation rate, cohort survival rate, etc.)

Student's self-concept describes the ideas and attitudes of the student, of himself derived from his own experiences with the self. It reflects student's awareness of "who am I" and "how I feel" about "who am I." Seven dimensions have been identified: physical self, emotional self, intellectual self, spiritual self, social self, political self, and economic self.

Students' concept of work describes students' ideas and attitudes towards work and career.

## **Methodology**

The study will be conducted in 15 regions. Students, teachers, and principals of 120 schools will comprise the respondents. The data will be obtained during school year 1994-1995 through documents, reports, checklists, questionnaires, observation forms, and cognitive tests.

The evaluation process consists of the activities delineating, obtaining, providing, and utilizing.

The political analysis will include descriptive, stepwise multiple regression, and path analysis.

## **Conclusion**

By 1995, we are hoping to make statements on the quality and significance of the SEDP as synthesized from the stories of the many people who shared their SEDP experience, their judgments about what went on, especially in the classrooms, and their insights on what is true and valuable. The hope, is for these professional discussions to illuminate decision makers and practitioners to action for a better secondary education.

# ASSESSMENT IN VOCATIONAL EDUCATION: AN AUSTRALIAN PERSPECTIVE

Peter Thomson

*National Centre for Vocational Education Research  
Adelaide, South Australia*

## Introduction

Australian technical and vocational education is currently undergoing substantial changes. In common with a number of other countries, there has been a shift from the more traditional norm-referenced methods of assessment towards criterion-referenced approaches. The change has been driven by a determination on the part of the government to make educational outcomes more closely related to the requirements of the workplace.

In this paper I will be describing the background of this new approach, identifying some of the assessment challenges it presents and some of the traps to be avoided by any country about to go down a similar path. To help with an understanding of what follows, it is important to define how "competency" and "competency-based assessment" are being used in Australia.

Competency is the ability to perform to the standards expected in the workplace. The standards cover:

1. knowledge
2. skills
3. attitudes

Competency-based assessment (CBA) is the process of gathering and judging evidence against the standards that are expected in the workplace. (CBA is therefore a form of criterion-referenced assessment.)

## Key Features of the System

When the new vocational education and training system is fully in place it will have the following features:

1. training will be based on identified competencies derived from a system of national competency standards;
2. assessment and certification will be competency-based;
3. a quality assurance system covering the registration of training providers and accreditation of courses will facilitate the recognition of training and competence wherever required;
4. a greater emphasis on meeting industry needs with a more active involvement of small and medium sized businesses in training reform;

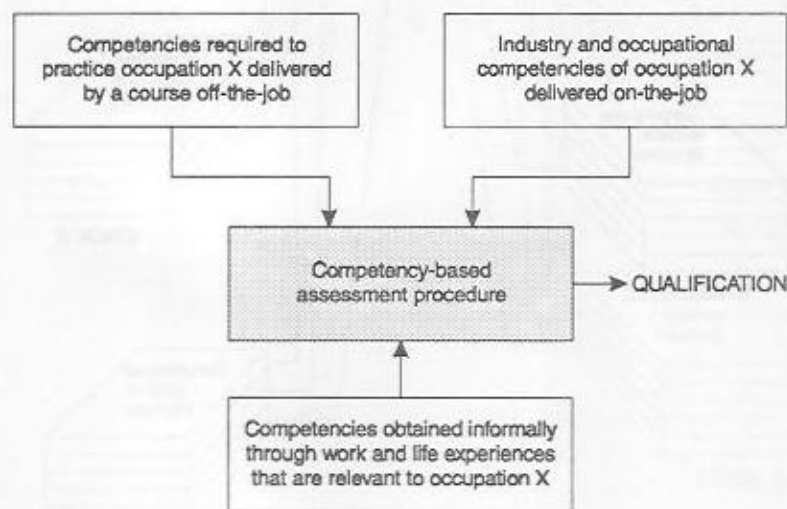
5. recognition of prior learning and credit transfer arrangements will facilitate access to training and recognition;
6. the system will link with the work of the schools and will provide a range of pathways to training and employment which will include achievement in key areas of competence (DEET, 1992).

It is also important to place these features in the larger context of Australia's technical and vocational education system.

### The Relationship Between Competencies and Their Assessment

One of the features of the new Australian system is that it allows for the recognition of competency *regardless of how it has been acquired*. In practice this means that not only competencies acquired as a consequence of on- and off-the-job training courses, but also those competencies acquired at work or informally, such as through a hobby, can be used as credit towards a qualification. The way this works is shown in Figure 1.

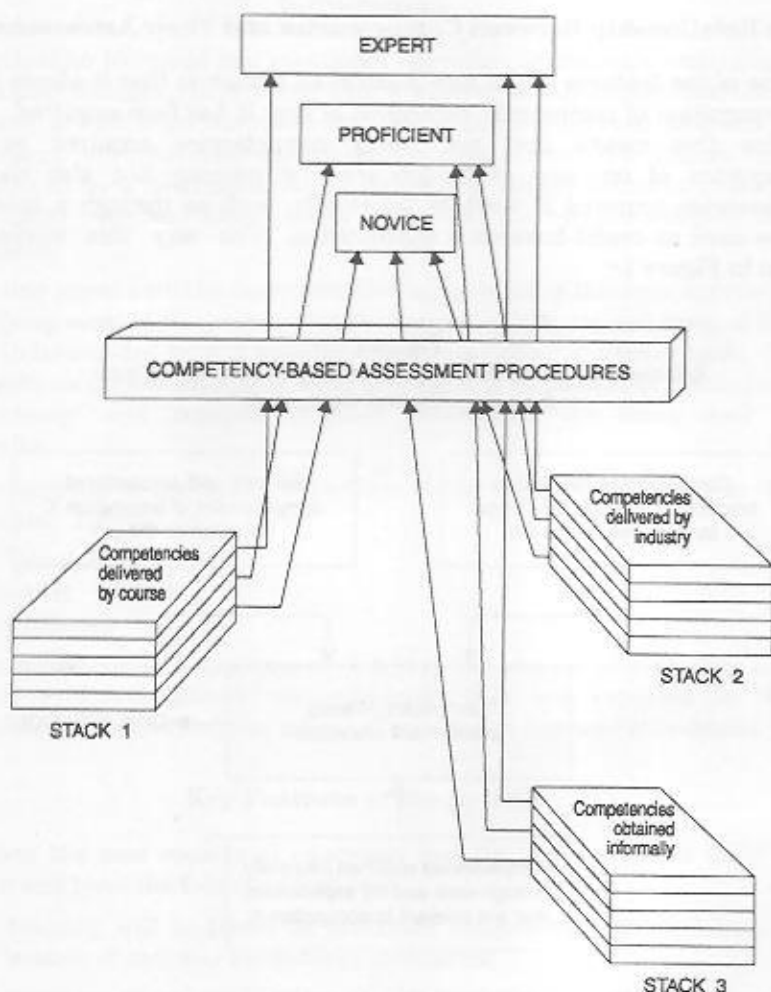
Figure 1  
Relationship Between Competencies and Their Assessment  
for a Hypothetical Occupation



However, the two-dimensional picture given by this figure hides a complexity. The facts are that people practice occupation X with varying degrees of proficiency, that requirements vary within industries, and that the set of competencies obtained informally are likely to vary substantially from individual to individual.

Figure 2 is a better representation, as it shows the relationship between the possible levels of an individual's competency (novice - proficient - expert) and the source of competency (off-job course work - on-job industry work - informal activities) that are used.

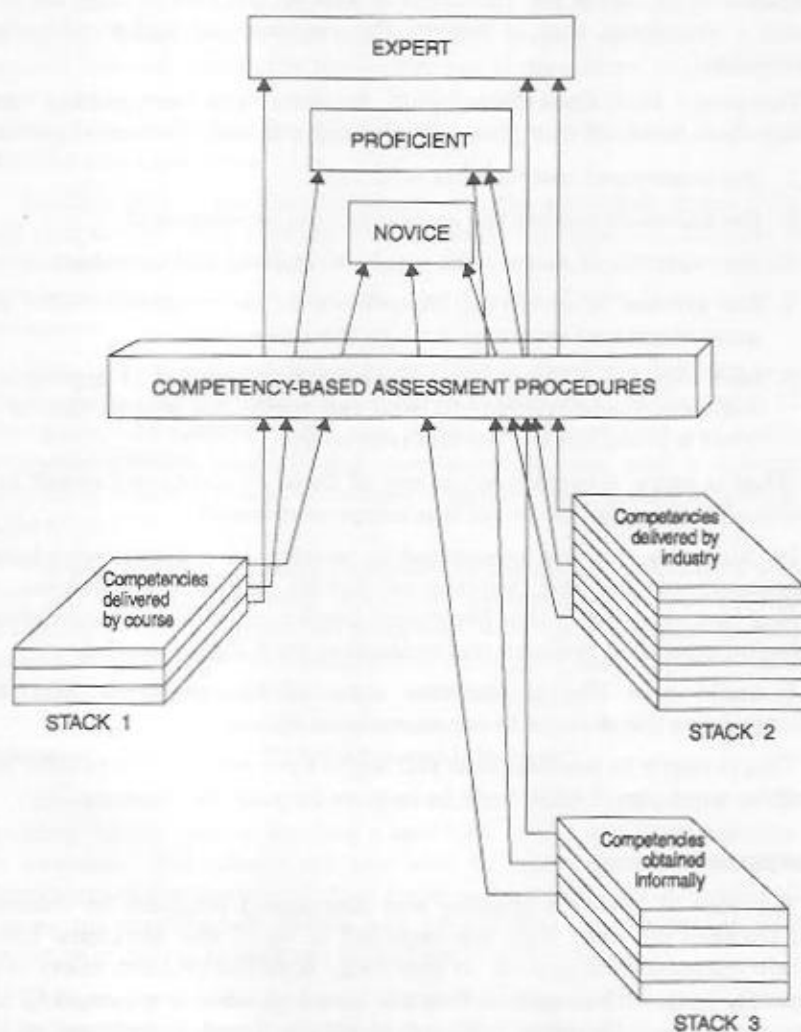
Figure 2  
Relationship Between Competence Requirements  
And Levels of Occupational Proficiency



However, even this figure is a simplification because different levels of performance require different mixes of competencies. For example, in the case of the novice not all the competencies required will necessarily come from the lowest levels of the "stack", nor are the stacks necessarily of equal size.

Figure 3 is perhaps more representative of the direction in which we are headed.

Figure 3  
Relationship Between Competence Requirements  
And Levels of Occupational Proficiency



#### Assessment in a Competency-based System

Vocational education people say CBA is good because it measures the skills needed to do jobs and it is based on the standards expected in those jobs. Therefore it provides a relevance that has previously been missing from our education and training systems.

The counter argument coming from some university people is that CBA is bad because its job market is so diverse. Therefore, it is impossible to achieve a close relationship between the higher education job market and competency standards. They say that for competency-based assessment to work you would need hundreds of assessments to match the hundreds of jobs at the end of, say, an arts degree - something that is beyond the resources of higher education institutions.

However, the assessment gulf between the antagonists and protagonists is not all that great. Best practice in both systems requires:

1. the assessment instruments to be valid
2. the assessors making the assessments to be competent
3. the reporting of assessment results to be clear and unambiguous
4. the process of certifying competence to be compatible with the assessment and reporting done by the assessors
5. each element of the system to be subject to scrutiny by appropriate authorities and individuals who can verify the overall quality of what is being done at the different levels.

What is more, a breakdown in any of these elements will result in a defective system whether or not it is competency-based.

In Australia, we are committed to moving to a competency-based assessment system and, because of certain political imperatives, to moving fast. Inevitably this has meant implementation has raced ahead of the underpinning research and evaluation work that is needed.

I would now like to describe some of the problems that are accompanying the changes to our assessment system.

This is partly in the hope that you might have some advice to offer and partly to warn you of what could be in store for your own systems.

### **Workplace Standards**

The idea of basing a training and assessment program for students and trainees on what they are expected to do in the workforce has a certain commonsense appeal. It gets away from the problem many of us currently have with programs that are based on what is expected by the universities and therefore relevant to only a small percentage of the student population.

Competency-based training and assessment has a beguiling simplicity. All we need to do is describe what we expect of successful workers or employees and then train them, assess them and certify them as appropriate. Unfortunately, it is not that easy. The task of describing what we expect people to do in work (the "competency standards") and devising test to measure or judge whether they have met the standards is proving to be a great challenge.



In Australia we are finding a general dissatisfaction with the competency standard statements so far written. This is partly because they are based on what is required across an industry and their relevance is sometimes lost in individual workplaces. The fact is that individual workplaces within the same industry have their own way of doing things. They have their own work culture and the language used to describe things can be unique to that setting. This doesn't necessarily make the general industry standards irrelevant, but it does mean that assessment instruments need to be tailored to suit each workplace which poses a substantial challenge for workplace assessors who have had limited training and experience.

Another part of the dissatisfaction with the standards stems from the fact that we are often working from untested assumptions. Because there have been few field trials we are still not sure if the competency standards statements are adequate descriptions of what is expected in the workplace.

More importantly, we do not know if the tests we are devising to judge the students and trainees are predictive of successful performance in the workplace. As already noted, one reason for this is that our system of competency-based training and assessment is new and a substantial amount of research needs to be done before we can be sure we are on the right track.

We are, however, suitably depressed by the picture painted by workers who tell us that assessing workplace performance is extraordinarily challenging. For example, we must deal with the persistent inability of success in training to be a predictor of success on the job (Landy & Farr, 1983, Wignor & Green, 1991).

### **Assessing Competency That Is Acquired Informally**

One element of Australian government policy associated with its training reform agenda involves a new look at the way qualifications can be awarded. Individuals are now able to receive certification of their competencies *irrespective of how those competencies are obtained*. This means an experienced person can be certified as competent without necessarily having to undergo formal study and training.

This process is already a part of the education and training systems in a number of countries. It goes by a variety of names like:

1. Assessment of Prior Learning (APL)
2. Accreditation of Prior Experiential Learning (APEL)
3. Current Competency Assessment (CCA)

In Australia, we call it RPL - the recognition of prior learning and it is much vaunted as a significant advance over the way things were previously done. However, the process is the subject of limited research

and may well need to undergo substantial changes if it is to avoid finishing up in the graveyard of assessment innovations.

Mullin (1992) who also noted the lack of research in this area, goes on to demonstrate that assessors of prior learning need special skills. In particular, he shows that if the incidence of competence in a population representing the RPL assessment is low, then the assessment methods used will need to be very powerful in detecting non-competency. Yet, virtually all the Australian RPL assessment is directed at detecting competence, with the importance of detecting non-competence receiving little attention.

It almost goes without saying that assessors charged with the responsibility of verifying that a trainee is competent must be sure (beyond reasonable doubt) that the necessary standard of performance and underpinning knowledge is present. But when people are applying for the right to practice in a vocation through the assessment of their prior learning, the demands placed on the assessors are formidable. In these cases the individuals are not seeking the novice status of new graduate (with its expectation that learning will continue as they gain experience in the job), but rather are seeking the right of full or expert status. The task is not impossible, but it requires a higher level of assessment skills than those associated with the assessment of the normal trainee intake.

### **Industrial Relations Issues**

If these challenges about how to deal with competencies acquired on-the-job or informally are not enough, to them we must add the problem of industrial relations issues intruding into the procedures. Despite the recent revision in the National Training Board's policy (NTB 1992) which breaks the link between competency standards and industrial award classifications, the principle is already firmly established in a number of industries. What was policy is now used in a *de facto* way to argue over terms and conditions of employment.

Nowhere are industrial relations pressures more apparent than in the assessment of prior learning. Whole workforces are restructuring and existing employees must be slotted into job classifications which are being based, at least in part, on the skills that individuals possess.

The inevitable consequence of this situation is that we are seeing terms and conditions of employment being used as bargaining chips to help determine movement from one level of qualification to another. This is probably the ultimate contaminant of any educational assessment process. All of us involved in the educational side of the process need to seriously consider whether there is any point in continuing our efforts when industrial relation negotiations are introduced into the procedures.

## Coping With Street-wise Individuals

Another option we have in the new system is the ready availability of assessment. (The term "on-demand assessment" is regularly heard in the rhetoric of the more evangelical advocates). Apart from the substantial cost implications of providing tests to individuals when they believe they are ready to be tested, there is also a danger that the practice might invalidate the assessment.

The fact is that the proportion of students or trainees will always try to "beat the system". Some years back Hively, et. al (1968) related the story of a group of candidates who:

*...discover that a good way to make progress in the curriculum is to skip through the programs of instruction and take the tests as quickly as possible. They then ask to have their errors explained, and come back immediately to take the test again. This turns out to be embarrassing for the instructor, who has trouble "protecting" the tests under these circumstances, and confusing to anyone who attempts to use the tests to evaluate the effectiveness of the curriculum. (p.280)*

On-demand competency-based assessment has every chance of revisiting this problem. If the rules allow repeat assessments without a prescribed period of remedial education and training then individuals who are not competent will escape detection.

Assessors will need to insist that the system in which they work has rules that protect against abuses of this sort.

## None Of The Above

The problems that have been identified are not the only ones we are facing. Time does not permit me to give you the full story. I could, for instance, have said a lot about assessor training. We are trying to train people in competency-based assessment skills in short courses which last two to four days. And I could have said even more about the cost of implementing competency-based approaches. Cost is proving to be a great worry to the business community. There is an expectation on the part of government that at least some of the cost of assessing in the workplace should be borne by the employers. In our country, like most others, this is not a good time to suggest employers bear more costs for anything!

## On A Brighter Note

Although the problems of competency-based assessment are numerous we need to remind ourselves of the reason the system has been embraced with such enthusiasm. Foremost amongst these is the dissatisfaction with the alternative -- the traditional norm-referenced approach -- which emphasizes knowledge acquisition rather than performance.

Furthermore, it can be argued that the problems identified here are a function of the demands to put a new system in place as speedily as possible. As a consequence the necessary research and development work is either incomplete or not yet started.

While this is obviously a dangerous situation to be engaged in, there are some hopeful signs that resources will be allocated to enable us to address the problems.

### References

- DEET (1992) Department of Employment Education and Training. Australian Country Paper. OECD Conference paper, Oporto, Portugal.
- Landy, F. & Farr, J. *The measurement of work performance, methods, theory and applications*. Academic Press, New York.
- Mullin, R. (1992). *Decisions and judgments in NVQ-based assessment*. NCVQ R&D Report No.14, London.
- NTB (1992) National Training Board. *Policy and guidelines* (2nd edition), NTB, Canberra.
- Hively, W., Patterson, H.L., & Page, S.H. (1968). A "universe-defined" system of arithmetic achievement tests. *Journal of Educational Measurement*, (5)4, pp. 275-290.
- Wigdor, A.K. & Green, B.F. (1991). *Performance assessment for the workplace*, vol. 1.

# QUALITY ASSURANCE SYSTEM IN THAI HIGHER EDUCATION: CURRENT DEVELOPMENTS AND ISSUES

Amornwich Nakornthap  
*Ministry of University Affairs, Thailand*

Quality assurance in higher education has been a popular topic of discussion within the higher education community lately. A quick review of quality assurance movements across the continents revealed a strikingly similar pressure placed on higher education institutions to demonstrate "accountability" to the government and the public. Various approaches and methods, ranging from self-regulated type to centralized control are being employed to guarantee the wise and effective use of public money and to secure public confidence in higher education. As the movement towards higher education accountability grows, however, there is a concern and a need for a quality assurance system that is most responsive to the nature and necessities of higher education institutions.

As for the higher education system in developing countries, the same phenomenon is quite evident as outputs of higher education are becoming more crucial to building national competitiveness. In an effort to find the most acceptable model of quality assurance for higher education, particularly for the developing world, the author will use the development of higher education quality assurance system in Thailand as a case. Some patterns of this model might be identifiable to the situations in other developing countries sharing the same experiences of growing public demand on higher education. Special attention will be given to the essential role of the government in fostering the system of quality assurance while preserving the principle of "autonomy" and "self-governance" in higher education.

## Background of Thai Higher Education

Throughout the history of Thai higher education, the government actively involved in the development and provision of higher learning. It was started from the founding of the first university, Chulalongkorn, in 1916 to meet the government's manpower needs, the establishment of Thammasat University in 1933 to fulfill the requirements of the post-revolution democratic society and the more recent setting up of regional universities throughout the kingdom during 1960-1968 to accelerate the pace of economic and social development of the country. Most of the major changes and expansion have been initiated by the government trying to make the higher education system responsive to national priorities of the time.

Administratively, therefore, since the early period of higher education development in Thailand, all universities were attached to some sort of government agencies, i.e., Chulalongkorn and Thammasat is under the Ministry of Education, the medical university of Mahidol is under the Ministry of Public Health, Kasetsart (agricultural university) is under the Ministry of Agriculture, etc. In 1959, they were all transferred to the Prime Minister's office to improve the standard and coordination among universities and to facilitate cohesiveness in government support. In 1972, all public universities were separated from the Prime Minister's Office to be independently supervised by the newly-established Ministry of University Affairs (MUA). Later on in 1975, all private degree granting institutions, appearing on the scene following the Private Higher Education Act of 1969, was also put under the supervision of the Ministry of University Affairs.

The role of the government in higher education was strengthened in 1961 when the first Five-Year National Economic and Social Development Plan was introduced identifying higher education institution as an integral part of the national development scheme. When the MUA was established in 1972, the Plan was also introduced as part of the National Economic and Social Development Plan to be carried out and monitored by MUA. At present, the higher education system is under the seventh plan covering the period from 1992 to 1996.

After the founding of MUA, with support from the national development planning, the growth and expansion of the system continued. Among the notable developments were the establishment of two open universities, Ramkhamhaeng in 1971 and Sukothaithammarat in 1978. At present, nearly one million students are enrolled in 50 public and private institutions under MUA with over 20,000 instructional staff and approximately US\$50 million are allocated for the annual budget.

### **Government's Roles in Quality Assurance of Higher Education**

The new role of MUA as an independent unit solely responsible for the supervision of higher education started the new era of higher education administration. Furthermore, the role of government is evident through the MUA's implementation of bureaucratic control over the operation of universities including policy and planning, personnel administration, budget allocation, and academic standards. With regard to quality assurance mechanism for higher education, the following activities of MUA come into focus:

**Macro-planning:** The opening of new academic program is subject to pre-approval through the planning process. In other words, any prospective programs need to be justified and pre-registered for each 5-year National Higher Education Development Plan as to be eligible for approval by MUA. Unregistered programs will be automatically rejected from the approval process. The planning mechanism thus provides some



guarantee of the justification and appropriateness of new academic programs in the universities.

**Academic Standards Setting:** The Bureau of Higher Education standards has developed sets of standard for various curricular levels to be observed by both public and private institutions in developing new programs. Such standards include only minimum requirements such as minimum requirements for general education courses, minimum credits for major courses, and minimum credits for total program, etc. Some specifications regarding credit-counting system or academic calendar system are also included. For public institutions that have been granted the self-approval authority, these standards serve as guidelines for program development and the institution still has to submit the approved program to the MUA for information.

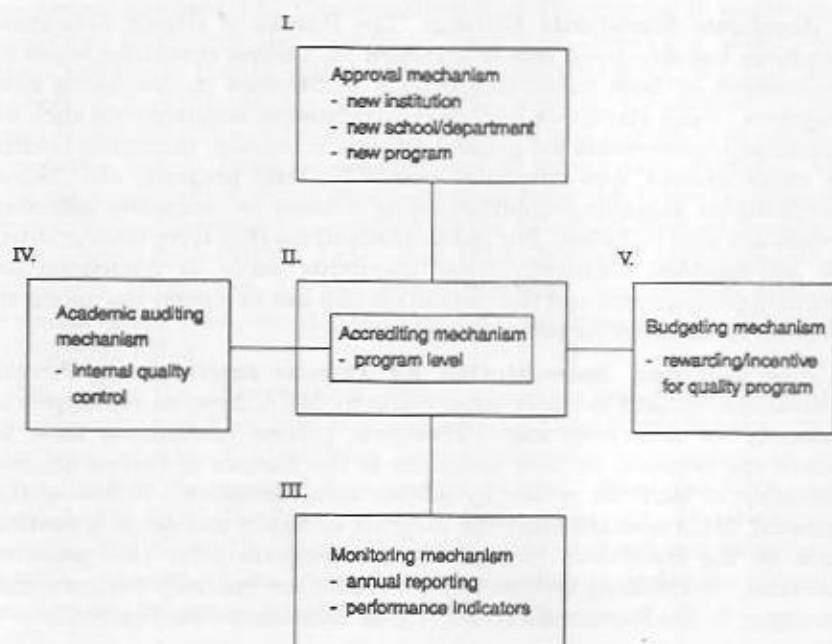
**Approval and Accreditation for Private Institutions:** Private institutions, subject to closer supervision by MUA, have no self-approval authority for new programs. Therefore, private institutions have to submit the proposal for new programs to the Bureau of Private Higher Education of MUA for review by subject area committee. Following the approval, MUA also monitors the program annually and sends a visiting team to the institution to accredit the program after two years of operation. Accrediting criteria and procedure are specially designed and developed by the Bureau of Private Higher Education (See Figure 2).

The above mentioned mechanisms for quality assurance currently in use are, however, ineffective. All three mechanisms only reflects a traditional bureaucratic control over the operation of universities and yet does not provide sufficient means of assuring quality in higher education. The planning process only provides some broad screening for new programs while the academic standards set forth by the Bureau of Higher Education Standards only serve as a general framework for curriculum development. Exception is made for private institutions where the approval and the accreditation processes of the Bureau of Private Higher Education seem able to assure certain level of quality in private institutions. However, the current system employed in private institutions might be too rigid and obtrusive if one considers it from the viewpoint of institutional autonomy. On the contrary, caution is raised over the self-approval authority granted to public institutions because insufficient standards and poor monitoring mechanism by the MUA might yield minimal accountability to the public.

### Current Developments

As the issue of accountability has been continuously raised in recent years among the higher education communities, efforts have been made by the MUA to strengthen the quality assurance mechanism for higher education institutions. Currently, five complementary mechanism have been identified as potential means of assuring and enhancing quality in higher education. This is shown in Figure 1.

Figure 1  
Quality Assurance System in Thai Higher Education



1. *Approval Mechanism:* The approval process for new institutions or programs is maintained as an appropriate means for ensuring quality in higher education. However, the criteria employed in the process need to be reviewed and improved to provide sufficient ground for evaluating the merit of each new proposal.
2. *Accrediting Mechanism:* The proposal is made to utilize the accreditation activities, currently in use with private institutions with similar criteria and procedure.
3. *Academic Auditing Mechanism:* This new mechanism modeled after the Academic Audit Unit (AAU) of the Committee of Vice Chancellors and Principals (CVCP) in England reflects the effort of MUA to provide a parallel and complementary means to the quality assurance system. The emphasis is placed on the internal quality control at the institutional level as a foundation of a continuing and effective quality assurance. The MUA formed and appointed the Academic Auditing Sub-committee to develop commonly agreed guidelines for the internal quality control mechanisms. The committee is also in-charge of making arrangements with the institution for an on-site visit by a team of auditors to review the effectiveness of the mechanisms. Such process is voluntary by nature and the Academic Auditing Sub-committee only plays a facilitative role in disseminating

examples of good practice to the wider audience of institutions. However, the auditing information could complement other quality assurance activities such as accreditation and budget allocation.

4. *Monitoring mechanism:* The new proposal also aims at strengthening the monitoring mechanism through the use of performance indicators covering areas such as general performance, financial performance, and outcome performance. Effective reporting and information systems are needed to facilitate the proper use of indicators.
5. *Budgeting Mechanism:* Budgeting mechanism, if wisely and cautiously utilized, could become another useful means of quality assurance in higher education. More specifically, if budgeting process can be used to provide rewards and incentives for quality programs or institutions, the mechanism should be welcomed by both the policy makers as well as the higher education community. Accreditation information, auditing information, and various performance indicators can be used to provide qualitative aspects in the budget allocation process.

Table 1  
Criteria Employed in Various Quality Assurance Mechanisms

Approval Criteria	Accrediting Criteria	Auditing Criteria	Follow-up Criteria	Budgeting Criteria
<b>Institution :</b> - Need - Mission - Organization & administration - Curriculum - Faculty/staff - Library - Physical facility - Student services - Finance - Monitoring & evaluation mechanisms	<b>Curricular process :</b> - Administration - Faculty/staff - Physical facility - Library - Student services  <b>Curricular outcome :</b> - Student achievement - Job-seeking ability - Employers' view  <b>Curricular Self-monitoring :</b> - information system - faculty evaluation - external examiners - graduate/alumni follow-up - periodic evaluation	<b>Institutional research :</b> - organization - staff - resource - information system  <b>Instructional development :</b> - faculty evaluation - graduate/alumni follow-up - external examiners  <b>Staff development :</b> - faculty development - administrative personnel development - support staff development	<b>General Performance :</b> - student information - curriculum information - faculty information - Physical facility information - Library information - Special programs information  <b>Financial Performance :</b> - expenditures by programs/items - capital vs concurrent expenditures - income-generating activities  <b>Outcome Performance :</b> - Student achievement - Job-placement records - employer's survey	- Accreditation information - Auditing information - Performance information
<b>Program :</b> - Administration - Faculty/staff - Physical facility - Library/learning resources - Student services - Monitoring & evaluation mechanisms				

Figure 2  
Accreditation Process

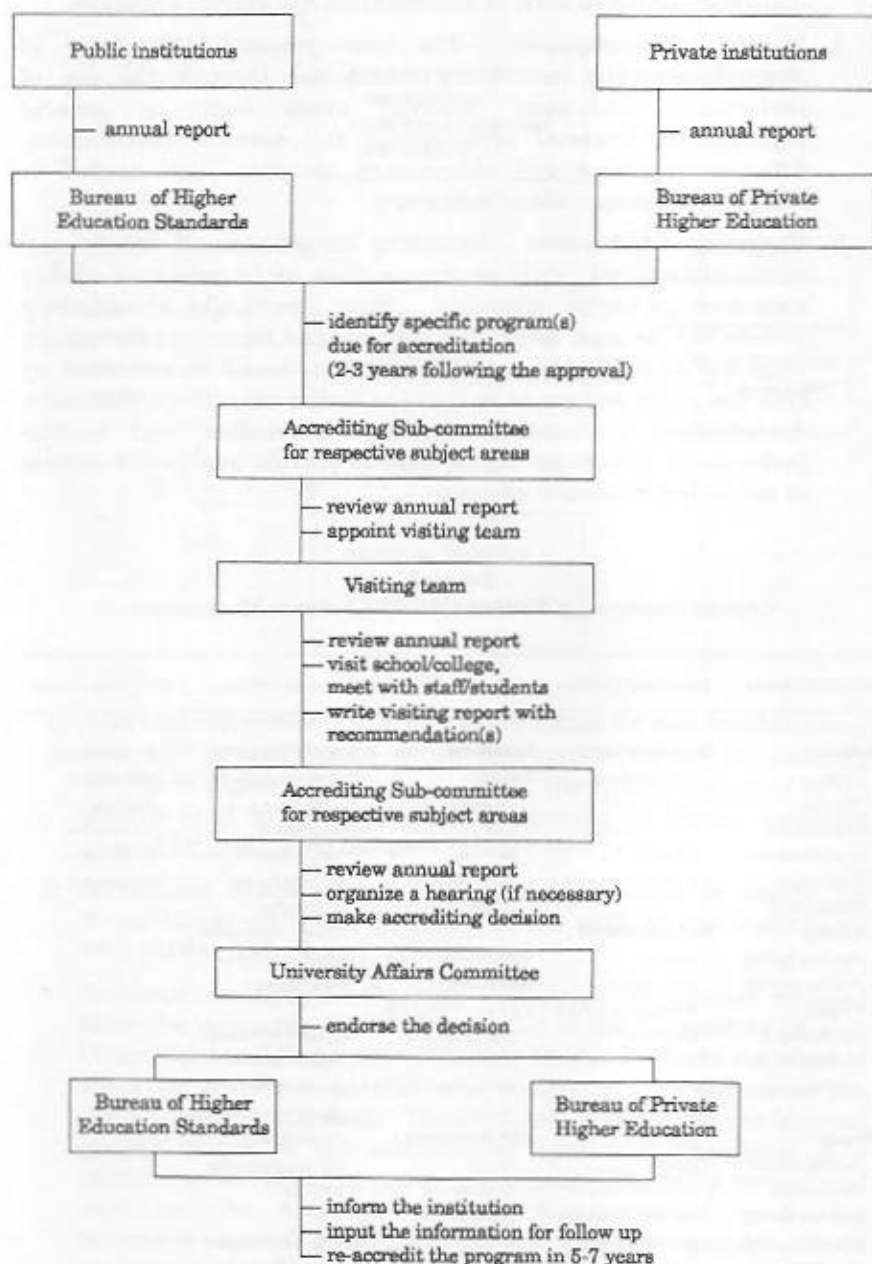
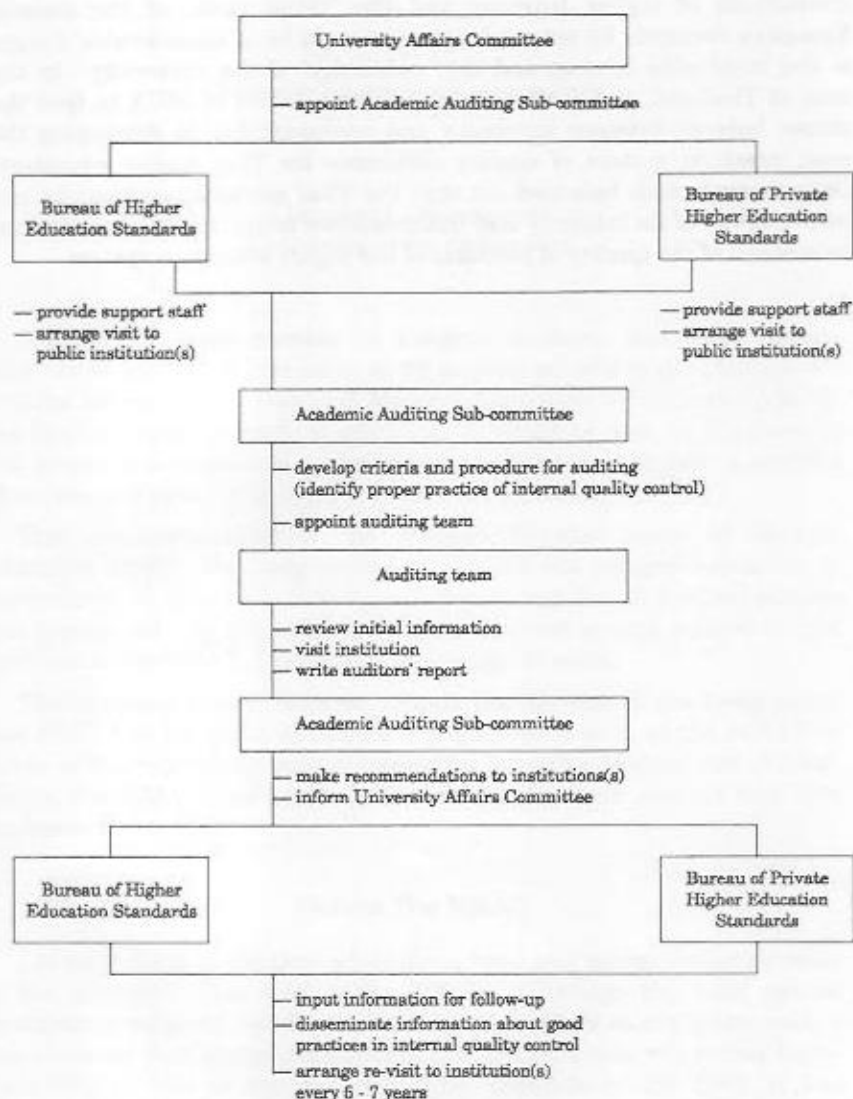


Figure 3  
Academic Auditing Process



### Conclusion

The notion of accountability in higher education is gaining momentum among various governments, especially in the developing countries where higher education is placed side by side to the national development. It is certain that more and more proposals regarding quality assurance in higher education will appear on the scene in these countries as "accountability" becomes the key word not only to the higher education

community, but to the politicians and public at large. On the other hand, one must be aware of the nature and need of the universities as the institutions of higher learning and the "think tank" of the society. Excessive demands by external agencies could be of considerable danger to the invaluable mission and the "autonomy" of the university. In the case of Thailand, the challenge lies with the effort of MUA to find the proper balance between autonomy and accountability in developing the most practical system of quality assurance for Thai higher education. Only through such balanced act that the Thai academic community can be preserved of its integrity and independence while the Thai society can be assured of the quality of products of the higher education system.



# ASSESSMENT FOR ADMISSION TO MEDICAL COLLEGES IN THE PHILIPPINES

Basilio R. Iledan

*Center for Educational Measurement*

*and*

Lenore Ll. Decenteceo

*University of the Philippines*

Annually, a large number of Filipino students, and some foreign students as well, seek admission to 26 medical schools in the Philippines. But the advent of the National Medical Admission Test, or the NMAT, has limited access to medical education to students who, on the basis of test scores, are presumed to be better qualified to complete a medical education and eventually become successful doctors.

The constitutionality of the decision by the Board of Medical Education (BME), the body tasked to regulate the medical education in the country, to impose a common admission test for all medical schools was questioned. At the extreme, various interest groups pushed for the abolition of the NMAT. The issue was brought to court.

The Supreme Court, however, upheld the decision of the lower court: that BME had the right to impose a requirement such as the NMAT by virtue of the regulatory powers vested on it by the Medical Act of 1959. Hence, the NMAT has been administered every year since it was first implemented in 1985.

## **Before The NMAT**

Let us go back to the time when there were only seven medical schools in the country. This was in the 1950's. Although the total annual freshman enrollment was from 3,000- 4,000 students at any given year, it was observed that during those years, the drop-out rate was rather high - from 30% to 50% of the total freshman population. By 1960, it was evident that the medical schools were admitting too many students. To encourage a reduction in freshman enrollment, the government, through the BME, initiated two things: (1) two-year fellowship programs to the United States for medical faculty, and (2) financial aid to the school for the purchase of medical supplies and equipment.

It took more than four years to bring the figures down to approximately 1,600 annually. Admission policies seemed to have tightened. Accordingly, the passing rate for medical licensure increased

in the late 1960's. The number of successful examinees was between 85% and 90%.

From 1972 until about 1984, the scenario changed. Nineteen additional schools were established in different regions of the country. By schoolyear 1986-1987, a quota was determined for each school. At this time, the actual freshman enrollment was 3,331. In 1985 and 1986, the failure rate increased - 52% and 30%, respectively. Among others, the admission procedures and practices of some schools were questioned.

To arrest this decline in performance in the Board, the following measures were proposed: (1) that the use of a reliable and valid screening instrument be administered prior to entry into medical school, and (2) that this must be developed, administered, and maintained by an entity independent of any of the schools associated with medical education.

The Center for Educational Measurement (CEM) met the requirements of the Board and was therefore commissioned to undertake the development, administration, and maintenance of the testing aspect of the program. Since 1985, the CEM has performed this task. For almost a decade now, the role that the NMAT plays in the selection of the students for admission in Philippine medical schools continues to be a debated issue.

## **Establishing The Predictive Efficacy Of The Test**

### **Problem**

The general purpose of the study was to examine the relationship between scores in the eight subtests of the NMAT, namely, Verbal, Quantitative, Inductive Reasoning, Perceptual Acuity, Biology, Physics, Chemistry and Social Science, and the average percentage grades (APGs) obtained by students in the first and second year of medical school. The specific problem can be expressed as follows: To what extent and in what manner would grades in the first two years of medical school can be predicted from scores in the NMAT subtests?

### **Sample**

One thousand four hundred ninety (1,490) medical students constitute the sample of this study. They belonged to the first batch of examinees who took the NMAT in 1985. At the time of this study, they were in their fourth year of medical school.

### **Variables**

*Predictor Variable or Independent Variable.* The scores in the eight NMAT subtests are the predictor variables of the study.

*Criterion Variable or Dependent Variable.* The APGs of final grades of each year level constitute the criterion variables of the study.

## Data Gathering

Average percentage grades (APGs) for the first two years of medical school were retrieved from records filed by the schools with the Association of Philippine Medical Colleges (APMC). The NMAT scores were retrieved from the files of the CEM.

## Analysis

Academic grades of the students were converted into a single scale because each of the 14 medical schools where they came from had its own system of grading. The extent of the relationship between academic grades and the NMAT scores was then examined by generating an intercorrelation matrix and multiple correlations relative to the prediction of first and second year final grades. Analyses of data were done at the CEM.

## Results and Discussion

Table 1 presents the correlations between NMAT scores and APGs for the first year and second year.

Table 1  
Correlation Coefficients of NMAT scores and APGs

NMAT Subtest	Year Level	
	I	II
1. Verbal	.28	.34
2. Quantitative	.41	.36
3. Inductive Reasoning	.14	.08
4. Perceptual Acuity	.14	.15
5. Biology	.33	.34
6. Physics	.47	.44
7. Chemistry	.46	.43
8. Social Science	.31	.38

The range of correlations is from 0.14 to 0.47 and from 0.08 to 0.44 for NMAT scores and first and second year APGs, respectively. From the magnitude of the correlations, it can be seen that the Physics subtest, followed by Chemistry, is the most important predictor of academic grades at both levels. The lowest correlations were obtained in Inductive Reasoning and Perceptual Acuity.

Tables 2 and 3 present the results of a multiple regression analysis. It also includes the relative contribution of specific subtests to the prediction of academic grades.

Table 2  
Multiple Regression Statistics and Percentage  
Contribution of Predictors to First Year Grades

PREDICTORS	Percentage Contribution
Physics	34
Chemistry	31
Quantitative	26
Social Science	10
Multiple R = 0.544	
$R^2 = 0.2962$	
Index of Forecasting Efficiency = 15.98 %	

Table 3  
Multiple Regression Statistics and Percentage  
Contribution of Predictors to Second Year Grades

PREDICTORS	Percentage Contribution
Physics	27
Chemistry	23
Social Science	22
Quantitative	13
Verbal	9
Perceptual Acuity	6
Multiple R = 0.525	
$R^2 = 0.2758$	
Index of Forecasting Efficiency = 14.72 %	

The multiple Rs with their corresponding  $R^2$  are 0.544 and 0.525 for the first year and second year, respectively. The  $R^2$ s indicate that: (1) scores in the Physics, Chemistry, Quantitative and Social Science subtests explain 29.62% of the variance in the first year APGs, and (2) scores in the Physics, Chemistry, Social, Quantitative, Verbal, and Perceptual Acuity subtests explain 27.58% of the variance in the APGs of second year students.

The question now is whether or not the NMAT is a good predictor of academic performance in the first two years of medical school. Obviously, the higher the value of the Multiple Rs between test scores (the predictor) and APGs (the criterion), the better. That is, the basis for evaluating any predictor is in relation to other prediction measures. Unfortunately, there are no officially reported studies on instruments being used in the admissions procedures of medical schools in the Philippines. Hence, at this time, no comparisons can be made.

The question then should be: How high must the correlation be for the test to be useful? The range of correlations from 0.30 to 0.80 represents the level of the validities usually found for useful test in psychological and educational practice. Tests that correlate with a coefficient less than 0.30 are considered of limited value, while correlation coefficients greater than 0.80 are seldom achieved by these instruments (Guilford & Fruchter, 1978).

It is clear from the data that Multiple R values for both year levels fall within the range of significant values ( $F = 136.61, p = .001$ ;  $F = 69.07, p = .001$ ).

The Physics subtest was the best predictor of academic performance in the first year. This is borne out by its percentage contribution to prediction (see Table 1). Next in importance is the Chemistry subtest followed by the Quantitative subtest, which appears to be the most pervasive.

Similarly, in terms of percentage contribution to prediction, the Physics, Chemistry, Quantitative, and Social Science subtests came out to be the best predictors of academic performance in the first year of medical school (see Table 2).

Verbal and Perceptual Acuity subtests did not figure as significant predictors in the first year of medical school.

The results of the multiple regression procedure applied to the test scores and average percentage grades supported the prediction of achievement in the first two years of medical schools. That scores in the NMAT subtests are good predictors of grades in medical courses is established in this study.

The findings indicate that although the NMAT explains only about 30% of the variance in the academic performance of students in each of the first two years of medical school, the use of the NMAT improves the prediction of academic performance in the medical course by as much as 15%. This is indicated by the *Indices of Forecasting Efficiency* at each year level involving specific subtests of the NMAT.

## ISSUES AND PROBLEMS IN EDUCATIONAL RESEARCH IN THE PHILIPPINES

Jesus A. Ochave  
*Philippine Normal University*

PRO  
*Philippine Society for Educational Research and Evaluation*

Some problems in educational research in the Philippines can be classified into two clusters. The first cluster is structurally external, i.e., these are problems that stem out of the socio-political climate in the Philippines. Samplers are as follows: 1) culture and practices of practitioners in education including that of management are not research oriented; 2) minimal support system for research in education, and for research in general; 3) politics, i.e., major educational decisions do not always consider research findings; 4) educational research not generating a critical mass of knowledge for practitioners, and research methodology seem to lag behind the demands of practice; 5) non-visible research agenda in education at the national level; 6) unknown data banking system on research findings in education; 7) irrelevant educational research problems vis-a-vis social and community development needs; and 8) minimal presence and visibility of educational researches as a group vis-a-vis the political body and society at large.

The second cluster is internal or inherent to the current state of the art of educational research and the competence of educational researchers. Among the problems in this category are as follows: 1) lack of familiarity with new developments and technology of knowledge/methodology in educational research, evaluation, measurement, and testing, e.g., g-theory, MANOVA, latent trait models, metaphors in evaluation; 2) apparent fragmentation of research findings, absence of discernible coherence (the need for instructional and learning theories for Filipinos, medley of ad hoc hypotheses of thesis projects; 3) absence of evidence to blend earlier findings, i.e., the pool of accumulated knowledge, with new findings; 4) no visible clearing house in educational research, absence of a system of dissemination and sharing of findings; 5) low level of statistical analysis (tally), variance explanation is seldom appreciated; and 6) the need to see beyond the  $t$ 's and the  $F$ 's and recognize the researcher's role in the educational process and policy development.

Issues in educational research are controversies inherent in educational research. Samplers are as follows: 1) allocation of the dearth of resources in educational research, i.e., intellectualization or policy oriented; 2) concern over statistical assumptions for parametric tests; 3) research methodology, quantification or "thick description", univariate or multivariate; 4) generalizations of research findings given shifting



conditions, and the issue of "fittingness" of results; 5) test item analysis: empirical results vs. judgment of experts; 6) scoring of objective multiple choice tests, 0/1 or weighted scoring?

This list of problems and issues in educational research is by no means exhaustive. Nonetheless, the identification was cross validated by the responses of some professors, graduate students, and graduate student-practitioners in education. What seems to be important now are the implications suggested by these problems and issues for determining the future agenda of educational research in the Philippines and within the context of education research and assessment in the Asia-Pacific region.

## PANUKAT NG PAGKATAONG PILIPINO (PPP)

Annadaisy J. Carlota  
*University of the Philippines*

This morning I would like to describe our research on the development of a measure of personality which we now call the Panukat ng Pagkataong Pilipino (PPP). My presentation is divided into four parts as shown in the outline which has been distributed to you. First, I will talk about the specifications of the present form of the test; second, the procedure that we followed in developing the test; third, a short summary of data on reliability and validity; and fourth, suggestions for its use in educational research and assessment.

### The Specifications of the PPP

The PPP is a 220-item inventory which assesses the 19 personality dimensions listed in Table 1. The table shows the original Pilipino labels of the dimensions and their corresponding English translations. Some of them are not really exact translations. Most of us who work on translating foreign measures know the problems associated with translations. I will not go into those now. However, the translations should give you an idea of the content of the different dimensions. For further clarification, please refer to Table 2 which gives the definitions of these dimensions.

Table 1  
The Personality Dimensions of the Panukat ng Pagkataong Pilipino (PPP)

Dimension	
Pagkamaalalahanin	(Thoughtfulness)
Pagkamaayos	(Orderliness)
Pagkamadaldal	(Social Curiosity)
Pagkamagalang	(Respectfulness)
Pagkamahinahon	(Emotional Stability)
Pagkamalikhain *	(Creativity)
Pagkamapagkumbaba	(Humility)
Pagkamapagsapalaran *	(Risk Taking)
Pagkamaramdamin	(Sensitiveness)
Pagkamasayahin	(Cheerfulness)
Pagkamasikap *	(Achievement Orientation)
Pagkamasunurin	(Obedience)
Pagkamatalino	(Intelligence)
Pagkamatapat	(Honesty)
Pagkamatiyaga	(Patience)
Pagkamatulungin	(Helpfulness)
Pagkamaunawain	(Capacity to be Understanding)
Pagkapalakaibigan	(Sociability)
Pagkaresponsable	(Responsibility)

**Table 2**  
**Definitions of the Personality Dimensions of the PPP**

<b>Dimension</b>	<b>Brief Definition</b>
Pagkamaalalahanin (Thoughtfulness)	Consideration for the welfare of others
Pagkamsayos (Orderliness)	Orderliness in appearance as well as work
Pagkamadaldal (Social Curiosity)	Curiosity or inquisitiveness about people's lives
Pagkamagalang (Respectfulness)	Recognition and acceptance of peoples's beliefs and right to privacy, respect for elders
Pagkamahinahon (Emotional Stability)	Control of emotional expressiveness
Pagkamalikhain (Creativity)	Capacity to be innovative and flexible
Pagkamapagkumbaba (Humility)	Modesty in the face of accomplishment
Pagkamapagsapalaran (Risk-Taking)	Tendency to seek new challenges, to take risks
Pagkamaramdamin (Sensitiveness)	Tendency to react readily to emotionally arousing situations of the negative type
Pagkamasayahin (Cheerfulness)	Disposition to be cheerful, capacity to percieve humorous aspects of situations
Pagkamasikap (Achievement Orientation)	Tendency to strive for excellence at work
Pagkamasunurin (Obedience)	Comformity to other people's demands, tendency to be generally agreeable
Pagkamatalino (Intelligence)	Self-perceived level of intelligence
Pagkamatapat (Honesty)	Sincerity and truthfulness
Pagkamatiyaga (Patience)	Capacity to tolerate repetitive/routine activities
Pagkamatulungin (Helpfulness)	Readiness in offering assistance to others
Pagkamaunawain (Capacity to be understanding)	Tolerance for the shortcomings of others
Pagkapalakaibigan (Sociability)	Capacity to enjoy the company of others, to make friends easily
Pagkaresponsible (Responsibility)	Carrying out one's tasks on one's own initiative

Each personality dimension corresponds to a subtest comprising a homogeneous collection of items. In effect, therefore, the PPP is a battery of 19 personality tests. A consequence of viewing the PPP in this way is that a researcher or practitioner who may be interested in assessing only one specific dimension need not administer the entire scale. What should be done, instead, is to select and administer only the items pertinent to that particular dimension.

The PPP was originally constructed in Pilipino which used to be the national language; now, of course, the national language has been changed to Filipino, although this is still largely Pilipino-based. At present, the PPP has translations in English and in a number of Philippine languages which include Cebuano, Ilocano, and Ilonggo.

The items of the PPP are in the form of statements; half of them are positively stated and the other half are negatively stated. Responses to the items are scored on a 5-point agree-disagree scale. Following classical test theory, the total score for a personality dimension is obtained by summing the scores on the relevant items. The higher is the total score, the higher is the level of that particular dimension.

The PPP is intended for use with individuals aged 13 and above. It usually takes about 45 minutes to an hour to administer. There are two forms of norms for the test, namely: (1) percentiles and (2) normalized scores with a mean of 50 and a standard deviation of 10.

### **The Development of the PPP**

Let me now go to the development of the PPP. The first step in developing the instrument was to identify the personality dimensions that would be salient to describing the Filipino personality. In order to do so, first, the related literature was reviewed. Second, a survey was conducted which I thought would provide additional information that would be helpful for identifying the relevant dimensions. The survey data were intended to supplement the theoretical-conceptual literature with an empirical data base. The respondents comprised over 267 Metro Manila residents varying in age from 13 to 68 years. Each respondent was asked to describe: (1) a person he/she knew and liked, (2) a person he/she knew and disliked, and (3) himself/herself. At the same time, a respondent was requested to define the personality dimensions that they mentioned or to give the behavioral manifestations of those dimensions. Based on the frequencies with which respondents mentioned them, a ranking of those dimensions was drawn up. This information was combined with that obtained from the literature and the list was finalized. Sixteen dimensions were identified in this manner; three more were added because of the researcher's personal interest in them at that time. These are the three which are marked with asterisks in the list of Table 1.

Four hundred and twenty five (425) items were constructed to tap the 19 personality dimensions. These items were administered once more to a sample of 245 respondents with characteristics comparable to the survey sample. Analysis of the items made use of item-total correlations. Two hundred and twenty (220) items survived item analysis.

### The Psychometric Properties of the PPP

In this section, I would like to describe the psychometric features of the test.

#### The Reliability of the PPP

In the course of developing the test, internal consistency reliability for each subtest was computed three times: (1) for the best 10 items, (2) for the best 12 items, and (3) for the best 14 items. This was done for the different dimensions with the exception of the subtest on *Pagkamasikap* (Achievement Orientation) for which only 6 items were retained after item analysis. The various computations of internal consistency were obtained to in order to arrive at the final number of items which would comprise a subtest of a dimension. The number of items yielding the highest reliability was selected for final inclusion.

Table 3  
Number of Items Retained After Item Analysis and Internal  
Consistency Reliabilities (Coefficient Alpha) of the PPP

Dimension	Number of Items	Reliability
Pagkamaalalahanin (Thoughtfulness)	14	.84
Pagkamaayos (Orderliness)	10	.76
Pagkamadaldal (Social Curiosity)	12	.44
Pagkamagalang (Respectfulness)	10	.75
Pagkamahinahon (Emotional Stability)	14	.70
Pagkamalikhain (Creativity)	14	.78
Pagkamapagkumbaba (Humility)	12	.73
Pagkamapagsapalaran (Risk Taking)	10	.51
Pagkamaramdamin (Sensitiveness)	12	.61
Pagkamasayahin (Cheerfulness)	14	.83
Pagkamasikap (Achievement Orientation)	6	.63
Pagkamasunurin (Obedience)	14	.56
Pagkamatalino (Intelligence)	12	.94
Pagkamatapad (Honesty)	10	.89
Pagkamatiyaga (Patience)	12	.89
Pagkamatulungin (Helpfulness)	10	.67
Pagkamaunawain (Capacity to be Understanding)	10	.74
Pagkapalakaibigan (Sciability)	14	.68
Pagkaresponsible (Responsibility)	10	.83
Total Number of Items:	220	
Mean Coefficient Alpha:	.72	

From Table 3, we can see that, on the whole, the reliability of the PPP is comparable to most published personality tests. Its average internal consistency reliability is 0.72. Across the subtests, the range of reliability coefficients is from 0.44 for *Pagkamadaldal* (Social Curiosity) to 0.94 for *Pagkamatalino* (Intelligence).

A group of researchers, responding to criticism about the length of the PPP, shortened each subtest (except for *Pagkamasikap*) and examined the effects on reliability. Of course, as expected, the abbreviated PPP, consisting of 150 items, showed a decrease in average reliability to 0.54. For 15 subtests, internal consistency reliabilities declined; 12 were unchanged; and 2 showed an increase. The final ideal length of the PPP is the subject of continuing research at present.

### The Interscale Correlations of the PPP

Interscale correlations of the PPP were also computed to determine how and to what extent the different subtests relate to one another. Table 4 presents the results of this analysis.

Table 4  
Summary of Interscale Correlation Matrix of the PPP

Dimension		Average Inter- correlation	Range of Inter- correlation	Number of Intercorrelations	
				Positive	Negative
Pagkamaalalahanin	(Thoughtfulness)	.26	-.10 to .51	17	1
Pagkamaayos	(Orderliness)	.19	-.23 to .61	14	4
Pagkamadaldal	(Social Curiosity)	-.10	-.05 to .25	4	14
Pagkamagalang	(Respectfulness)	.32	-.26 to .52	16	2
Pagkamahinahon	(Emotional Stability)	.24	-.36 to .47	16	2
Pagkamalikhain	(Creativity)	.24	-.15 to .42	16	2
Pagkamapagkumbaba	(Humility)	.22	-.39 to .47	16	2
Pagkamapagsapalaran	(Risk-Taking)	.29	-.05 to .43	17	1
Pagkamaramdamin	(Sensitiveness)	-.09	-.37 to .37	6	12
Pagkamasayahin	(Cheerfulness)	.18	-.15 to .40	17	1
Pagkamasikap	(Achievement Orientation)	.27	-.18 to .52	16	2
Pagkamasunurin	(Obedience)	.06	-.15 to .20	12	6
Pagkamatalino	(Intelligence)	.19	-.30 to .46	16	2
Pagkamatapat	(Honesty)	.24	-.26 to .53	16	2
Pagkamatiyaga	(Patience)	.24	-.37 to .54	15	3
Pagkamatulungin	(Helpfulness)	.24	-.25 to .41	16	2
Pagkamaunawain	(Capacity to be understanding)	.33	-.09 to .54	16	2
Pagkapalakaibigan	(Sociability)	.17	.03 to .37	18	0
Pagkaresponsible	(Responsibility)	.24	-.32 to .61	16	2



The following observations emerge from an examination of the data shown. First, the correlations are generally low to moderate in magnitude. Second, there are more positive than negative correlations overall as well as for each individual subtest, with the exception of two dimensions, namely *Pagkamaramdamin* (Sensitiveness) and *Pagkamadaldal* (Social Curiosity). I interpreted these findings to be supportive of certain hypotheses that had been posed earlier. First, the generality of positive correlations was interpreted to mean that the subtests are measuring something in common and that this commonality is the construct of personality. Second, the low to moderate magnitude of the correlations implied that the different subtests are measuring dimensions of personality which are relatively distinctive of one another. Third, the reversal of the directional trend in correlations for the two dimensions of *Pagkamaramdamin* and *Pagkamadaldal* warranted an explanation. I think that the basis for this finding may be found in the scoring direction that is used for the PPP. As had been intended, the subtests are scored in the direction of the socially desirable end of the dimensions. Thus, for instance, the higher is the score for *Pagkamaalalahanin* (Thoughtfulness), the more thoughtful/considerate that person is. Thoughtfulness is the socially-valued end of this dimension. It may be surmised, at this point, that perhaps being too *maramdamin* and *madaldal* (too sensitive or socially curios, respectively) are not positively valued in our culture. If the scoring of these subtests were reversed which would as a consequence, make them consistent with the direction of the rest, it is expected that these two dimensions will then correlate positively with the other personality dimensions. Again, this is another aspect of the PPP that is presently being examined.

### The Validity of the PPP

Let me now turn to some studies on the validity of the PPP. The usual design of these studies is correlational. Typically, the PPP is correlated with the *Panukat ng Ugaling Pagkatao* (PUP) which was developed by Dr. Virgilio Enriquez. There are a number of dimensions in the two tests which appear to be parallel in nature. Thus, studies have correlated the subtests of the hypothesized parallel dimensions. To date, all of the available studies using this design have been conducted with high school students. The students have come from three different types of schools. The first study by Bautista, et al. had participants from a public school; that of Abaya, et al., from a private high school; and Agana, et al., from a science high school. Their findings are found in Table 5. In general, the correlations are low to moderate in magnitude but they are statistically significant in a number of cases, as indicated by the asterisks. What is especially encouraging in the results is that, with few exceptions, the correlations, those that are statistically significant as well as those that are not, are in the predicted direction.



Table 5  
Parallel Scales of the PPP and the PUP

Author/Year	PPP	PUP
Bautista, et al. (1984)	Pagkamasunurin	Sunod-sunoran
	Pagkamagalang	Tigas ng Ulo *
	Pagkamapagkumbaba	Pagkamagalang **
	Pagkamahinahon	Pagkamapagkumbaba *
		Pagkamapagtimpi *
Abaya, et al. (1984)	Pagkamatapat	Pagkapalaaway **
	Pagkamatulungin	Salawahan **
		Pagkamatulungin
	Pagkamalikhain	Pagkamalikhain
	Pagkamasunurin	Tigas ng Ulo
Agana, et al. (1984)	Pagkamatulungin	Pagkamatulungin
	Pagkamagalang	Pagkamagalang
	Pagkamaramdamin	Pagkamaramdamin
	Pagkamapagsapalaran	Lakas ng Loob
	Pagkamasikap	Ambisyon
	Pagkamaramdamin	Pikon **
	Pagkamapagkumbaba	Pagkamapagkumbaba **
	Pagkamapagsapalaran	Pagkasigurista **
	Pagkaresponsible	Pagkaresponsible **
	Pagkamaalalahanin	Pagkamaalalahanin
	Pagkamalikhain	Pagkamalikhain
	Pagkamasunurin	Pagkamasunurin
	Pagkamatulungin	Pagkamatulungin
	Pagkamagalang	Pagkamagalang
	Pagkamaramdamin	Pagkamaramdamin
	Pagkamapagkumbaba	Pagkamapagkumbaba
	Pagkaresponsible	Pagkaresponsible
	Pagkamatiyaga	Katiyagaan
	Pagkamapagsapalaran	Lakas ng Loob
	Pagkamahinahon	Pagkamapagtimpi
	Pagkamaalalahanin	Pagkamaalalahanin

\* Correlation significant at  $p=.05$

\*\* Correlation significant at  $p=.01$

Other studies were more innovative in design, as far as the local literature in measurement is concerned. One such study which was conducted by Bernardo, et al. used a multi-trait multi-method design for convergent and discriminant validation of *Pagkamaramdamin* (Sensitiveness), *Pagkamatiyaga* (Patience) and *Pagkamalikhain* (Creativity) against parallel subtests of the PUP. Findings were positive for the first two dimensions but the *Pagkamalikhain* subtests of the PPP and PUP showed little convergence.

Three studies were conducted which correlated selected PPP subtests with the behavioral criteria, which are a welcome departure from the typical criteria that are employed in validation studies. There were two studies on *Paghamapagsapalaran* (Risk-Taking) conducted by Musni, et al. and Feliciano and Mercado. Both of these obtained correlations are in the predicted direction but they, unfortunately, were not statistically significant. A third study by de Guzman, et al. correlated scores based on behavioral manifestations of friendliness in a contrived setting and scores on the *Pagkapalakaibigan* (Sociability) subtest of the PPP. The correlation was 0.33 and was statistically significant.

Finally, I would like to mention that Alikpala and de los Reyes examined the translation equivalence of the original Pilipino and English versions of the PPP. Their study provided evidence in support of the equivalence of the translation of 17 of the 19 subtests.

### **The Use of the PPP in Educational Assessment and Research**

I would like to state, at the outset, that the PPP was not developed specifically for use in the educational assessment. It was designed and constructed to assess personality in varied settings, including, of course, the educational setting. In actual practice, the PPP is not as widely used in the educational setting as I would like it to be. As a matter of fact, it is perhaps more widely employed in industry for the purpose of pre-employment screening, as implied by the observation that most purchases of the test have been by industrial companies.

Let me now mention some studies in the educational setting which have employed the PPP. Most of the researches that have been conducted have attempted to describe selected students by drawing up their group personality profile. One example is the dissertation of Bartolome. Her study describes the personality profile of students who have been nominated by their professors as being more creative than the rest. The participants came from different fields and professions.

Let me end my presentation by describing what we have been doing in the Department of Psychology as far as the PPP is concerned. In the department, a colleague and I have administered the PPP to selected classes for several semesters. We are continuing to do so in order to build a data bank for the test. We have correlated PPP subtest scores with some achievement indicators, namely, overall GPA, GPA for non-psychology courses, GPA for psychology courses and average psychology course grades in two categories: the methods courses and the content courses. Data-gathering is ongoing. Although we have done small groups analyses, we have not really looked at the entire picture. In that connection, we consider what we have been doing in the larger context of research on the identification of non-cognitive influences on learning, including those factors that may facilitate or hinder learning. In terms of application, one direction that we can see such researches to be taking is

the use of the PPP dimensions, as well as other measures of non-cognitive characteristics, as indicators or predictors of achievement. I see the potential for this application to be more salient in higher education, possibly at the tertiary level or in graduate education. Since the aptitude/ability-based screening for these levels tends to be rather strict, students in these levels are generally at the higher end of the ability continuum and are relatively more homogeneous in ability than they are at the lower levels. Despite these facilitating features, however, a wide range of variability in actual academic performance is still observable. This observation seems to indicate that the predictability of academic achievement may be improved by supplementing information about cognitive abilities and related aspects such as cognitive styles with information about the noncognitive factors that affect learning. Some of those which we are presently exploring are achievement orientation, sense of commitment, perseverance, and resourcefulness.

# SHORT FORMS OF THE MARLOWE-CROWNE SOCIAL DESIRABILITY SCALE (MCSDS): COMPARATIVE PERSPECTIVES OF THE CLASSICAL AND MODERN TEST THEORIES

Werawat Punnitamai  
*Thai Psychological Association*

For test developers, it is ideally desirable to construct a test which consisted of a small number of scaling items but provides adequate measuring qualities and information of a trait under interest. Psychometric properties of the short forms of Marlow-Crowne Social Desirability Scale (MCSDS) derived from two measurement approaches -- the classical test theory (CTT) and the item response theory (IRT) are examined. Six item selection methods for achieving the MCSDS short forms have been investigated. MCSDS short forms previously developed via the CTT approach were compared to four proposed IRT item selection methods in terms of measurement precision, item overlap, relative efficiency, and item-trait reliability under three fixed test lengths.

Preliminary IRT assumption assessments revealed that MCSDS data were relatively unifactorial, considerably independent of inter-item responses of American and Canadian examinees, and were adequately fit by the 2-parameter logistic IRT model. The property of item parameter invariance was virtually supported across different subgroups of students.

Across the 7-, 10-, and 13-item test lengths, items chosen by the IRT selection techniques displayed impressive psychometric advantages over the random selection and the classical item selection methods at least six aspects: (a) comprising of higher discriminating items, (b) providing a greater value of (average) test information, (c) exhibiting higher degrees of item overlap and relative test efficiency, (d) showing a better reflection on the distributional form of the original MCSDS, (e) yielding considerably lower standard errors of trait estimates, and (f) associating with higher item-trait reliability of the short forms. Additional research avenues of applying the IRT concepts to personnel selection are also suggested.

# AN INTRODUCTION TO ITEM RESPONSE THEORY USING QUEST

Malcolm Rosier

*Survey Design and Analysis Services Pty. Ltd.*

An important technique in social science research is the combination of a series of variables into a single composite variable. The composite variable has the advantages of *parsimony* and *stability*. *Parsimony*, because a single variable can be used for description or in multivariate analyses instead of a series of separate variables. *Stability*, because the composite variable can give a more reliable estimate of the quality being measured than any of the separate constituent variables.

## Tests and Scales

A basic example of a composite variable is a dichotomous test score. The items in the test are the constituent variables in the composite, scored 1 for a correct answer and 0 for an incorrect answer. The test score is constructed by summing the values for the constituent variables.

Another example is a polychotomous Likert scale, where each of the items has several ordered categories. For example, a descriptive scale may list activities to which the person responds "often", "sometimes" or "never". An attitude scale may use the response pattern "agree", "neutral" or "disagree". A scale to measure performance may list a series of tasks for which the response pattern is "good", "fair" or "poor".

The responses to the items in a Likert scale are usually coded 1/2/3/4 and so on. For example: "often" = 3, "sometimes" = 2, and "never" = 1. The scale score is then constructed by summing the scores for the constituent items, or by more complex methods such as factor weights derived from a factor analysis.

## Traditional Measurement Theory

The raw data for calculating test scores may be presented as a person-item matrix. In Table 1, the persons are arranged by scores (ability) and the items are arranged by difficulty.

The quality of the individual items making up the test score may be evaluated in terms of several item characteristics:

- The p-value or facility of an item is the percentage of persons who have the correct answer.
- The discriminating power of an item is given by the item-total correlation. That is, the correlation across all persons between the score for that item and the total scores (or the total scores

excluding that particular item). This is the point-biserial correlation coefficient.

- The quality of the test as a whole may be evaluated psychometrically by the Cronbach alpha coefficient and by measures of reliability.

Table 1  
Person-Item Matrix

Person	Item						Total
	1	2	3	4	5	6	
1	1	1	1	1	1	1	6
2	1	1	1	1	1	0	5
3	1	1	1	1	0	1	5
4	1	1	1	0	1	0	4
5	1	1	1	0	0	0	3
6	1	1	0	0	0	1	3
7	1	1	0	0	0	0	2
8	1	1	0	0	0	0	2
9	1	0	0	1	0	0	2
10	1	0	0	0	0	0	1
Total	10	8	5	4	3	3	33

Many of the assumptions underlying the above traditional test measurement are being challenged. A key problem is that the evaluation of the quality of the items and the test as a whole depends on the *particular group of persons* used for the evaluation. In addition, reliability measures such as split halves or test-retest are essentially artificial.

One problem in traditional test theory is that the total test scores do not form a scale at the interval level of measurement (for example, the measurement of height, weight, temperature). For example, in a 20-item mathematics test, an increase in test score from 1 to 2 does not represent the same change in mathematics ability as an increase from 9 to 10.

### Item Response Theory

The Item Response Theory of Measurement (also known as latent trait theory or Rasch analysis) is designed to avoid many problems with traditional measurement theory. It reflects more precisely the relationship between the measurement process and the underlying dimensions of trait being measured.

The basic aim of item response theory (IRT) is to produce a calibrated scale of the underlying dimension which is at the interval level of measurement. The same scale is used to obtain a measure of the difficulty of an item (together with an estimate of the measurement error associated with the item) and to obtain a measure of the ability of a

person (together with an estimate of the measurement error associated with the person).

The basic equation for item response theory states that the probability of a person obtaining the correct answer to an item depends on the ability of the person and the difficulty of the item, as given by the formula:

$$P_{ni} = \frac{e^{b-d}}{1 + e^{b-d}}$$

$P_{ni}$  = the probability of a correct response by person  $n$  to item  $i$

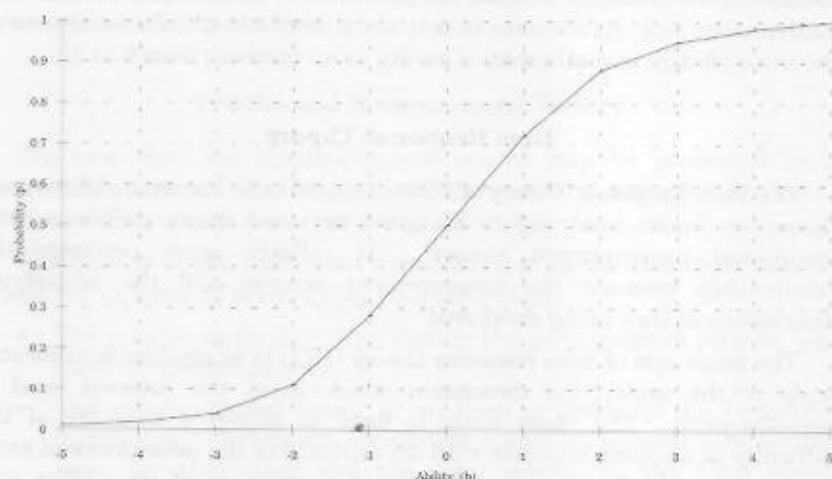
$b$  = the ability of person  $n$

$d$  = the difficulty of item  $i$

The relationship may be shown graphically in Figure 1. The graph shows the probability associated with obtaining a correct response on an item of difficulty  $d_i$  for persons across the range of ability. Where  $b_n$  is greater than  $d_i$ , the probability of a correct response is greater than 0.50. Where  $b_n$  is less than  $d_i$ , the probability of a correct response is less than 0.50. Where  $b_n$  is equal to  $d_i$ , the probability of a correct response equals 0.50.

The focus of IRT is on the individual items, and not on the test as a whole. In the estimation process, the influence of the particular group used to calibrate the scale is minimized by adjusting the mean ability of the calibration group and the spread of ability within the group.

Figure 1  
Item Characteristic Curve  
Difficulty = 0.0





The score for a given person does not depend on a particular set of items. The ability of a person is given by the location of that person on the calibrated scale. The accuracy of the person's ability is best estimated by using a series of items whose difficulty is close to that ability. The magnitude of the accuracy is measured by the measurement error, estimated for each person.

It is not necessary to give the same items to all persons. There is no need to give easy items to persons of high ability since they contain no information that is useful in deciding the location of that person on the scale. Similarly, there is no need to give hard items to persons of low ability.

Items calibrated by means of IRT can be used to construct item banks available for the selection of sets of items for groups of persons with defined ranges of ability. Tailored testing can then be employed, especially when administered by computers. The person is initially given a small set of items covering a range of difficulty. The person's responses are then used to select items closer in difficulty to the person's ability, until an accurate measure of that ability is determined. There is no need to give easy items to persons of high ability (and vice versa) so that the total number of items needed is less than for a conventional test.

### **Rating Scales and Partial Credit**

The above discussion dealt with dichotomous data such as test scores. A more advanced IRT model deals with polychotomous items (including descriptive scales, attitude scales, performance measures) where there are two or more response categories for each item. In this case the model estimates the probability associated with each category of each response.

### **Features of *Quest***

Further insight into the advantages of IRT can be demonstrated by outlining some of the features of *Quest* that are not usually available in other IRT programs.

- *Quest* can operate in interactive mode while exploring the data or in batch mode to handle a sequence of commands.
- *Quest* can handle dichotomous and polychotomous items.

*Dichotomous items* (such as test items) are scored as correct or incorrect. *Quest* estimates the scale value associated with the correct score for each item, together with the measurement error. The actual response may be coded numerically (1/2/3/4 etc.) or alphabetically (A/B/C/D/ etc.). You provide the key to the correct responses (distractors) for each item. *Quest* can handle up to 10 responses. With *Quest* you can assign a different value to the correct response for each item.

*Polychotomous items* have two or more response categories for each item. It is assumed that the categories are ordered (for example, from low to high) with respect to the underlying dimension. *Quest* estimates the scale value associated with each category for each item, together with the measurement error for each category. The actual response may be coded numerically or alphabetically. Since there are no right or wrong answers, you do not provide a key. You do not have to have the same number of categories for each item. *Quest* can handle up to 10 categories.

- *Quest* can estimate the polychotomous responses according to the rating scale model or the partial credit model. In the rating scale model, the magnitude of the scale difference between the categories is constrained to be the same across all items. In the partial credit model, this constraint does not apply and *Quest* estimates the scale values for the categories separately for each item. That is, "partial credit" is given to each of the response categories.
- *Quest* can handle up to 2000 cases (persons), 100 items, and a combination of 10 groups of persons or subscales of items. There can be up to 10 response categories for each item.
- The test data for analysis by *Quest* can be readily imported from the text files using standard FORTRAN statements. *Quest* lets you export results of the analyses to text files with a choice of field delimiters to facilitate incorporation into spreadsheet, database or other programs.
- *Quest* can handle recoding of items, including the deletion of specific items or persons from an analysis.
- *Quest* has flexible procedures to handle missing data. This means that a person's score which may be interpreted as his location on the calibrated scale, is based on the information available for that person. In contrast, in traditional measurement theory it is usually necessary to assume that a missing response is an incorrect response. This facility also enables different forms of tests to be calibrated at the same time, as long as there are common link items and/or persons.
- *Quest* permits you to anchor item or case estimates to specified values, usually obtained from a prior analysis. This facilitates the equating of tests and item banking. It also enables the estimation of parameters excluding items or persons with poor fit characteristics.
- *Quest* lets you define subgroups of persons (for example, boys and girls) or subscales of items. You can obtain parameter estimates for the subsets in order to compare them. *Quest* can also calculate correlations between the subgroups, with correlations based on raw scores and on the IRT scale estimates.

## Quest Output

The basic output for *each item* consists of the scale estimate (in logits) for each item, the estimated measurement error for each item, and a measure of the fit of each item to the underlying measurement model (a standardized t-value). These data are useful in developing a total test or scale. The t-values may also be displayed graphically.

The basic output for *each person* consists of the scale estimate (in logits) for each person, the estimated measurement error for each person, and a measure of the fit of each person to the underlying measurement model (a standardized t-value). The t-values may also be displayed graphically. For persons, these data give a measure of the consistency with which each person has responded to the set of items. This provides an opportunity to enhance the accuracy of the estimates by deleting mis-fitting persons from the estimation procedure. Their scores can still be estimated later but the removal of their "poorer" data enhances the quality of the basic estimation of the model.

The summary statistics show the overall fit of the model and the reliability. There is also a graph which shows the calibrated scale (in logits) on a central vertical axis, with the distribution of person scores (on the left side of the graph) and the location of item difficulties for dichotomous scales or the location of each response category for polychotomous scales (on the right side of the graph).

*Quest* can produce a diagnostic chart (kidmap) for each person. It shows the person's score and measurement error, together with the correct and incorrect items. The diagnostic value lies in identifying items or groups of items which were incorrect but which the person should probably have got correct. For dichotomous data, the selected responses for the incorrect items (distractors) are also shown.

*Quest* can produce all the traditional item analysis statistics, with the percentage correct (p-value) for each item and the point-biserial correlation for each response. However, it also incorporates the IRT estimates. For each item it shows the mean ability score (in logits) for all of the persons in the dataset selecting each response (including those with missing data for that item). *Quest* also calculates the traditional Cronbach alpha reliability coefficient. However the *Quest* value for a reliability coefficient is more accurate, especially where there are missing data.

## Quest Specifications

*Quest* is available on several computer platforms.

- The *Standard DOS version* will run on 8086 and 80286 machines. It requires 580 Kb of free RAM. It does not require a maths coprocessor but will run faster if there is one.

- The *Enhanced DOS version* requires an 80386 machine with a math coprocessor, or an 80486 machine. It makes use of extended memory, and runs much faster than the standard version.
- The *Macintosh version* requires 1 Mb of free RAM, but runs faster with an SE/30 or better, 2 Mb of free RAM and a math coprocessor.

Custom versions can be prepared to cater for larger datasets or other special requirements. A VAX/VMS version is also available.

### Summary

The ideas of item response theory have been steadily developing over the last 20 years. It has been used in many large-scale programs of educational assessment including:

- The Second International Science Study of the International Association for the Evaluation of Educational Achievement (IEA)
- The National Assessment of Educational Progress in the United States, and the associated International Assessment of International Progress.
- The New South Wales Basic Skills Testing Programs and other major educational assessment programs in Australia

Until recently, comprehensive computer programs have not been readily available to enable social science researchers to explore the technique. *Quest* now makes this possible with its rich range of features.

### References

- Andrich, D. (1988). *Rasch models for measurement*. Newbury Park, California: Sage Publications.
- Hambleton, RK, Swaminathan, H & Rogers, HJ. (1991). *Fundamentals of item response theory*. Newbury Park, California: Sage Publications.
- Masters, GN. (1988). Partial credit model. In: Keeves, JP (ed). *Educational research, methodology, and measurement: An international handbook*. Oxford: Pergamon Press.

# THE APPLICATION OF IRT AND OPTIMAL TEST DESIGNS IN THE ASSESSMENT OF MATHEMATICAL ABILITIES

Dr. Peter Lam Tit Loong

*Centre for Applied Research in Education (CARE)*

*National Institute of Education, Singapore*

Optimal test design involves the selection of items based on the assumption of the additive property of item information in Item Response Theory (IRT) from which the precision of a test can be controlled at different regions of the ability continuum. The choice and the level of competency of the items selected by the particular item selection heuristic depends on the anticipated ability distribution of the group of examinees to be tested and the test specification table used.

A standard procedure for test design based on the IRT model is described by Birnbaum (1968) which involves setting up a target information curve from which the test is to be built and selecting items with information function that will fill the area under the target information curve. The item information function is added cumulatively with back-tracking if necessary in order to fill the whole target information curve. In Birnbaum's (1968) and Lord's (1980) description of the heuristics involved in item selection, it is assumed that the selection process is done by hand.

A linear programming approach applied to Birnbaum's method of test design was developed and implemented in a number of studies (e.g. Van der Linden, 1987; Theunissen, 1985 & 1986). Theunissen's (1986) and Adema's (1990) use of binary programming enables the test developer to build a test by first, setting the target information function of the test and then proceeding to select items based on specific binary programming algorithms within the computer environment. Based on Theunissen's (1985) binary programming model, the computer program, Optimal Test Design (OTD) (Verschoor, 1992) was developed.

The study involved a real data simulation procedure as a first phase to the actual implementation of OTD in computer-administered testing in the schools.

## Theoretical Framework

Central to the application of IRT to testing is the use of item information function as the basis for item selection. For any fixed value of ability  $\theta$ , the parameter  $P_i(\theta)$  is the probability in which the examinee gets the answer correct for item  $i$ .

The item information function (Lord and Novick, 1968) for item  $i$  is given as:

$$I(u_i, \theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta) [1 - P_i(\theta)]} \quad (1)$$

The information function is derived from the maximum likelihood function for  $\theta$  based on the observed item responses,  $u_i$ . This function is the expected value of the inverse of the error:  $SE = \sqrt{[1/I(\theta)]}$ .

On the condition that local independence of item responses is kept, the item information is additive so that a test comprising a set of items will have the test information given by the summation of the item information:

$$I(u_i, \dots, u_n, \theta) = \sum \frac{[P'_i(\theta)]^2}{P_i(\theta) [1 - P_i(\theta)]} \quad (2)$$

Using a set of ability values and the corresponding set of items in the item pool, an item information matrix or information table (Thissen & Mislevy, 1990) can be created and stored in the computer. The information table is used for test designs in which a target information function for the test is specified and the items are selected.

In the implementation of IRT based test designs, Lord (1980) outlined Birnbaum's (1968) procedure for the use of item information function in test designs as follows:

1. Describe the shape of the target information curve in which the test is to be built.
2. Select the items with item information that will fill up the hard-to-fill areas under the target information curve.
3. After each item is added to the test, cumulatively add the item information for the presently selected item to the test information of the previously selected set of items.
4. Continue selecting the items until the test information approximates the target information curve satisfactorily.

A linear programming model formulated to solve a test construction problem attempts to optimally select a number of items in the test subject to the constraints that at least a certain amount of information is obtained at some pre-specified ability levels.

This model is stated as follows:

$$\begin{aligned} \text{Maximize:} \quad & \sum_{i=1}^I \sum_{k=1}^K I_i(\theta_k) x_i \\ \text{Subject to:} \quad & \sum_{i=1}^I I_i(\theta_k) x_i \geq T(\theta_k) \end{aligned} \quad (3)$$

$$\sum_{i=1}^I I_i(\theta_k) x_i \geq T(\theta_k) \quad (4)$$

$$\text{so that:} \quad \sum_{i=1}^I x_{i1} = \sum_{i=1}^I x_{i2} = \sum_{i=1}^I x_{i3} = \dots = \sum_{i=1}^I \sum_{t=1}^T x_{it} = N \quad (5)$$



In the above model,  $x_i$  is the decision variable for the  $i$ th item in the bank where  $i = 1, 2, \dots, I$ . If  $x_i = 1$ , the item is included in the test. If  $x_i = 0$ , the item is not included in the test.  $T(\theta_k)$  is the target information value at the ability level  $\theta_k$ .

The main purpose of this problem is therefore, to load a test with a number of items from a bank so that at a number of  $\theta$  points, the information  $[I_i(\theta_k)]$  in the test is above the target. The additional constraint is that the test must be balanced between the three content areas and all three subtests make up a total test length of  $N$ . Solution of this problem is done by an algorithm called the simplex algorithm, implemented in most computer programs and in particular, OTD (Verschoor, 1992).

### Method

Item-response data obtained from a paper-and-pencil administration of a 75-item math achievement test on 500 Grade 3 children from 5 elementary schools was used in the computer simulation study. The items from this test was previously calibrated using the two-parameter logistic model. The rationale for using this model instead of the 3PL model was that traditionally, the Grade 3 pupils were not fully exposed to guessing in multiple choice tests. Qualitative examination of the response sheets revealed that most pupils chose to omit items which they did not know. This could be seen by the large number of response blanks scattered everywhere in the item response matrix. In order to account for item discrimination and to conform towards the concept of classical discrimination, the 1-P (or Rasch) model was not used. The computer program ASCAL (Assessment Systems Corporation, 1987) was used to calibrate the items.

The purpose of this study is to redesign the test using the item pool from the original 75-item test and with a length of 40 items for the purpose of selecting a group of high ability level pupils for scholarship programs. The actual response strings of another group of 148 pupils of high ability level and who had previously taken the 75-item math test were used to score the redesign tests. To avoid capitalizing on chance, the response strings from the 148 pupils were excluded from the item calibration process.

The maths achievement test covers three almost equally balanced content areas: Number, Measurement, and Geometry. The redesigned tests bear the same content-balancing but items were selected at the high criterion cutoff region.

A fixed length test of 40-item with maximum information peaked at the proficiency range from 1.0 to 2.5 was developed using OTD procedure. The number of items reflecting content-balancing in Number, Measurement, and Geometry are 13, 14, and 13 respectively. The



development of this test was initiated by setting a target information of 0, 5, 5 at the proficiency levels: -2.0, 1.0 and 2.5. This would mean setting the precision of the test between  $\theta = 1.0$  to 2.5 at a SE of .045.

The computer program, OTD (Verschoor, 1992) was used to enumerate the set of design problem. The item bank file was created for input into the OTD environment. The target information was set for the specified ability levels in the development of the test within the OTD environment. Final adjustment by hand was made to obtain a content ratio of 13:14:13 for a test of 40 items. Another 40-item test with the same content balancing was developed using items with maximum classical difficulties and high ( $>0.2$ ) classical discrimination in the three content areas. For each of the two tests that was designed, the response strings obtained from a sample of 148 scholarship students were used to score the test based on the parameters obtained from the calibration sample of 500 pupils. The Bayesian modal procedure which is an extension of the maximum likelihood estimate procedure designed to eliminate infinite estimates was used in all scoring procedures. The conventional test scoring subsystems of MicroCAT (Assessment Systems Corporation, 1989) was used for computation.

The 148 scholarship pupils were originally classified at three criterion cut-off levels of 1.5, 2.0, and 2.5 based on the 75-item achievement test.

The effectiveness of the two item selection procedures were examined by the decision accuracy of reclassifying the pupils and comparing with the original classification. Decision accuracy was estimated by two indices:

- a) The percentage of the 148 pupils misclassified as failures when they were originally classified as successful by the original 75-item test.
- b) The percentage of pupils misclassified as successful when they actually failed in the original 75-item test.

The decision accuracy was determined at three proficiency levels: 1.5, 2.0, and 2.5.

## Results

### Descriptive Statistics

In the calibration process, there was no extreme lack of fit of the items based on *Chi square* statistics. The items generally converged after the tenth iteration.

Descriptive statistics of the item pool (see Table 1) obtained from the calibration sample of 500 Grade 3 pupils on the 75-item test showed that the distribution of IRT difficulty levels were uneven across content areas with Geometry items being the most difficult compared to the other two content areas. However, items in this content area were not as

discriminating compared to the items in the Number area, which were the most discriminating. This trend was also reflected in the classical estimates. The uneven distribution parameter estimates of the item pool has implications for test designs. Without any content balancing rule imposed in the item selection heuristic it is likely that more Number items will be selected since they tend to be more discriminating and hence, possess more information.

Table 1  
Descriptive Statistics of Item Pool

	Mean	SD	Min	Max
Item Pool (75)				
a	0.55	0.30	0.40	1.95
b	1.37	1.00	-0.78	3.00
Bis	0.35	0.16	0.04	0.74
Diff	0.31	0.14	0.07	0.72
Number (22)				
a	0.65	0.44	0.40	1.95
b	1.29	1.10	-0.42	3.00
Bis	0.40	0.20	0.12	0.74
Diff	0.33	0.15	0.11	0.63
Measurement (28)				
a	0.55	0.28	0.40	1.76
b	1.15	1.03	-0.78	3.00
Bis	0.36	0.16	0.04	0.66
Diff	0.34	0.16	0.07	0.72
Geometry (25)				
a	0.45	0.08	0.40	0.72
b	1.69	0.81	-0.44	2.96
Bis	0.29	0.08	0.17	0.44
Diff	0.25	0.10	0.14	0.54

Figure 1 shows the test information curve of the 75- item test. As expected, the curve reflected a typically peaked test giving rise to the common "Bandwidth-Fidelity" problem. That is, this test would be measuring quite precisely (Information = 15; SEM = 0.26) at the proficiency band between -0.5 and 0.0. The steep drop in test information below the proficiency level of -0.5 showed that the test would not be effective in measuring low proficiency level pupils especially those below -2.0. On the other hand, the skewed shape of the test information curve indicated that the test could be used for measuring higher ability level pupils. The criterion cut-off was set at criterion levels: 1.5, 2.0 and 2.5. Obviously, in the redesigning of the tests, items with high information at the lower proficiency levels had to be eliminated.

Figure 1  
Test Information of 75 - Item Math Tests

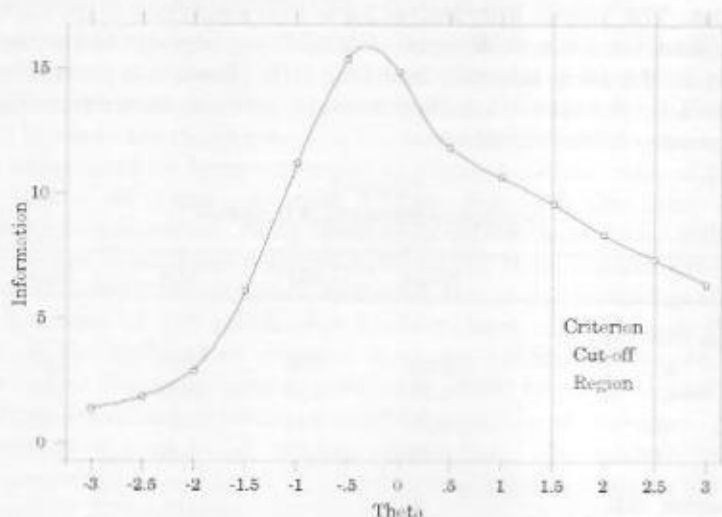
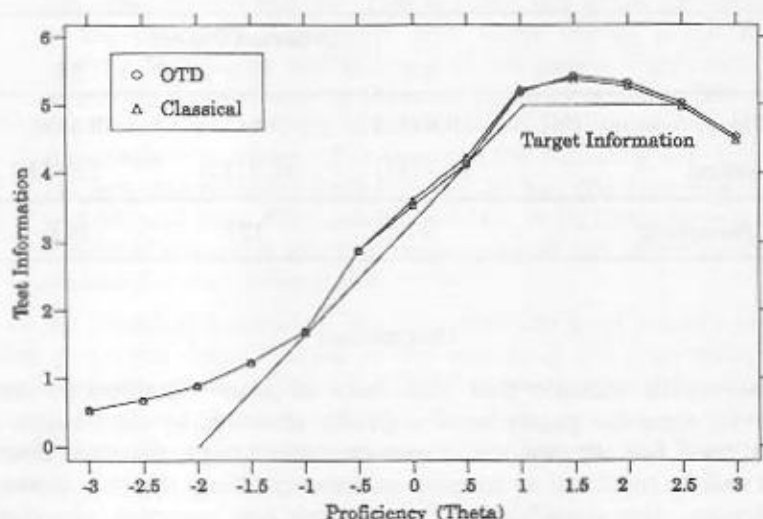


Table 2 shows the summary of item parameter statistics of the two test designs. The test design using the Classical selection procedure showed significant comparability with its IRT design counterpart. The item overlap between the two tests was 29 or 72.5 percent. Figure 2 showed the obtained test information curves with the target set at three proficiency levels: -2.0, 1.0 and 2.5 with target information: 0, 5, 5 respectively. Although the target information was never used in the classically designed test, nevertheless, the obtained information curve fitted the target almost as well as that of the IRT designed test. The redesigned tests of 40 item length would measure with a precision of at least a SEM of 0.45 (information = 5) for the criterion cut-off region between the proficiency levels of 1.5 and 2.5.

Table 2  
Descriptive Statistics of Classical and IRT Test Designs

		a		b		Bis		Diff	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
OTD (40)		0.44	0.07	1.75	0.43	0.29	0.10	0.25	0.05
Num	(13)	0.43	0.05	1.80	0.53	0.31	0.12	0.25	0.05
Meas	(14)	0.42	0.05	1.76	0.43	0.28	0.12	0.26	0.05
Geo	(13)	0.46	0.10	1.69	0.33	0.29	0.07	0.24	0.04
Class. (40)		0.44	0.08	2.01	0.56	0.29	0.12	0.22	0.05
Num	(13)	0.46	0.12	1.90	0.66	0.32	0.16	0.23	0.05
Meas	(14)	0.43	0.05	1.89	0.51	0.27	0.11	0.23	0.06
Geo	(13)	0.44	0.07	2.23	0.48	0.27	0.08	19.00	0.04

Figure 2  
Obtained Test Information of 40 - Item Tests



Analysis of the decision accuracy of the two test designs (see Tables 3 and 4) shows the following:

- Both redesigned tests showed a higher percentage of pupils misclassified as successful when they were originally classified as failures in the three criterion cut-off points.
- There was a lower percentage of pupils misclassified as failures when they should have been successful.
- The test designed by the IRT method showed greater decision accuracy especially at the criterion cut-off points of 2.0 and 2.5 compared to that of its Classical counterpart. In fact there was no successful pupils misclassified as failures at these two criterion cut-off points.

Table 3  
Percentage Successful Misclassified as Failures

	Criterion Cut-off		
	1.5	2.0	2.5
OTD	9.3 (5)	0.0 (0)	0.0 (0)
Classical	9.3 (5)	19.0 (4)	28.6 (2)
N Successful	54	21	7

Table 4  
Percentage Failures Misclassified as Successful

	Criterion Cut-off		
	1.5	2.0	2.5
OTD	10.6 (10)	9.4 (12)	3.5 (5)
Classical	11.7 (11)	10.2 (13)	7.8 (11)
N Successful	94	127	141

### Discussion

The results indicate that both tests do show classification errors. However, since the pupils were originally classified by the 75-item test which itself has its own set of measurement errors, decision accuracy could well be redefined as decision consistency. Since the true scores are not known, the classification comparisons are regarded as relative. Although the redesigned tests showed classification inconsistencies, the study showed that:

- a) A content-balanced test eliminating items that do not really cater to the criterion levels under consideration showed a reliable measure of classification consistency with that of the original 75-item test.
- b) The redesigned tests with a test length reduced by nearly 50% would result in savings in administration time. Aberrant responses of pupils due to test fatigue would be reduced. The original 75-item test took about two hours to administer in an unspeeded mode. The redesigned tests of 40 item length would probably take about 1 hour to administer.
- c) The IRT-designed test appeared to be more consistent in classification compared to its Classical counterpart. However, the differences may not be that significant at the lower criterion cutoff points. Bearing in mind the closeness of both test information curves, the study showed the relevance of incorporating Classical methods within the IRT frame work.
- d) One of the assumptions of IRT is the fungibility (interchangeability) of test items. Based on this assumption, since both tests have almost the same information curves (and hence the same measuring precision), it would be correct to conclude that both decision accuracies would be almost the same. The results showed that this is not so. One possible explanation is that many of the Geometry items have low discrimination values but high difficulty values. The forged choice of the Geometry items over that of the higher discriminating Number and Measurement items

for the purpose of content-balancing may result in incorporating items that drew in aberrant and other inconsistent responses from the experimental sample. This is partly due to the fact that some of the Geometry items were trial items testing skills that are slightly beyond the ability levels of the pupils. Furthermore, the Geometry items chosen by Classical criteria but not by IRT criteria are items that are very difficult in the IRT sense but with low discrimination values. For example, the following are items that fall into this category: Item #63 ( $a=0.40$ ,  $b=2.68$ ); Item #66 ( $a=0.40$ ,  $b=2.54$ ) and Item #75 ( $a=0.40$ ,  $b=2.14$ ). Such items have a higher chance of attracting aberrant responses and may affect the decision accuracy of the Classical test.

In the Classical framework, the item difficulty level actually pertains to the proportion correct based on the sample of the population. It is sample dependent compared to the IRT framework where the parameter estimates are sample independent. However, as long as subsequent samples come from the same population, the Classical test design is still relevant.

In the Classical framework, items with difficulty (or proportion correct) levels between 0.4 and 0.6 and high ( $>0.2$ ) discrimination levels are generally chosen in order to maximize variance. Choosing items with high difficulty levels (low proportion correct), would mean that item variance be compromised resulting in lower reliability estimates. However, this can be compensated by only selecting items with reasonably high discrimination levels. The study showed that even though the general rule for item selection in the Classical sense is not followed, selecting items with high discrimination tends to exhibit a compensatory effect as can be seen by the almost comparable test information curve compared to that of the IRT designed test.

### Conclusion

Where schools are equipped with the necessary computer hardware, the use of automated test development within the computer framework is a viable option. Given an item bank possibly shared between a number of schools, it is possible for teachers to improve the test development process by making use of automated methods of test designs. The latest version of OTD allows for test designs based on the Classical estimates. This would be advantageous to school practitioners who may be unfamiliar with the complexities of IRT.

The study provides for extension in the following areas:

- a) The present item pool of 75 items is insufficient to build a credible item bank. Hence the need to extend the item pool with subsequent development of test items and appropriate test equating in the calibration process.

- b) The test can be extended to include curriculum areas pertaining to higher grade levels.
- c) With the credible size item bank, more innovative procedures, including adaptive testing is possible.

Van der Linden (1987) reported that the binary programming model used by Theunissen (1985) and implemented in OTD resulted in a characteristic hump in the obtained test information function even though the target was set to be uniform across ability levels. This is because of the way in which the algorithm will select more items located in the middle of the interval specified by the target, resulting in a high test information in this region. This characteristic hump was apparent in this particular study and actually conform to the original intention of creating a peaked test for the purpose of criterion-referenced testing. As a possible further study in IRT test design, the minimax and maximin models developed by Van der Linden (1987) specify the minimization of the largest deviation between the test information and the target information and result in a closer approximation to the target. The prototype software, CONTEST (Boekkooi-Timminga, 1992) was recently developed to handle this model. An area for further study will be a comparison of the efficiencies of Van der Linden's (1987) minimax/maximin models and Theunissen's (1985) model in optimally selecting items given a uniform, peaked and bimodal target and the implications of these models in test construction.

### Bibliography

- Adema, J.J. (1990). The construction of customized two-stage tests. *Journal of Educational Measurement*, 27, 241-253.
- Assessmet Systems Corporation (1987). *MicroCAT Testing System*. St Paul MN: Author.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick, *Statistical Theories of Mental Test Scores*. Reading MA: Addison-Wesley, 1968.
- Boekkooi-Timminga, E. (1992). *Models for Computerized Test Construction*. De Lier, The Netherlands: Academisch Boeken Centrum.
- Computing Resource Centre (1992). *STRATA: Statistics/Graphics/Data management*. Santa Monica, CA: Author.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory: Principles and Applications*. Boston: Kluwer Nijhoff.



- Lam, T.L. & Foong, Y.Y. (1991). Development and validation of hierarchical testlets in two-stage tests using integer linear programming. Paper presented at the Annual meeting of the American Educational Research Association, Chicago, 1991 (ERIC: TM 016-919).
- Lam, T.L., Swaminathan, H & Hambleton, R. (1992). Use of binary programming in test designs to address content balancing in adaptive tests. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.
- Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, Mass: Addison-Wesley.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Erlbaum, Hillsdale, New Jersey.
- Theunissen, T.J.J.M. (1985). Binary programming and test design. *Psychometrika*, 50, 411-420.
- Theunissen, T.J.J.M. (1986). Some applications of optimization algorithms in test design and adaptive testing. *Applied Psychological Measurement*, 10, 381-389.
- Thissen, D. & Mislevy, R.J. (1990). Testing algorithms. In H. Wainer (Ed.) *Computer Adaptive Testing: A Primer*. Erlbaum, Hillsdale, New Jersey.
- Van der Linden, W.J. (1987). Automatic test construction using minimax programming in Wim J. van der Linden (Ed.) *IRT- based Test Construction*. (Research Report No. 87-2). Enschede, The Netherlands: University of Twente.
- Verschoor, A. (1992). *Optimal Test Design*. Arnhem: CITO.

## DIRECTORY OF PARTICIPANTS

### 1ST ASIA-PACIFIC INVITATIONAL CONFERENCE ON EDUCATIONAL ASSESSMENT & RESEARCH

Hotel Inter Continental Manila, Philippines

February 20-22, 1994

## A U S T R A L I A

### **DR. MALCOLM ROSIER**

Managing Director  
Survey Design and Analysis Services Pty. Ltd.  
249 Eramosa Rd. West, Moorooduc, Victoria 3933  
AUSTRALIA

### **DR. PETER P. THOMSON**

Deputy Director  
National Centre for Vocational Education Research  
252 Kensington Road, Leabrook, Adelaide 5068  
SOUTH AUSTRALIA

## G U A M , U . S . A .

### **DR. ZENaida NAPA-NATIVIDAD**

Territorial Program Evaluator  
Research, Planning and Evaluation Division  
Department of Education  
P.O. Box DE Agana, GUAM 96910  
U.S.A.

## I N D I A

### **DR. G. SUBRAMONIA PILLAY**

Professor & Head  
Department of Education  
Kamaraj University  
Madurai 625021  
INDIA

## J A P A N

### **PROF. AKIRA ICHIKAWA**

Professor  
Edogawa University  
Sociological Department  
474 Komaki, Nagareyama City, Chiba  
JAPAN

### **MR. NOBUYUKI KOBAYASHI**

Assistant Resident Representative  
Japan International Cooperation Agency  
12/f Pacific Star Bldg. Cor. Gil Puyat & Makati Avenue  
1200 Makati City  
PHILIPPINES

## MALAYSIA

### **MOHD ROSLI MOHD GHAZALI**

Head, Research & Policy Unit  
Assistant Director  
Examinations Syndicate  
Jalan Duta, 50605 Kuala Lumpur  
MALAYSIA

## SINGAPORE

### **DR. ANDREW POH SUI HOI**

Lecturer  
National Institute of Education  
Nanyang Technological University  
469 Bukit Timah Rd. S1025  
SINGAPORE

### **DR. PETER LAM TIT LOONG**

Senior Lecturer  
National Institute of Education  
Nanyang Technological University  
469 Bukit Timah Rd. S1025  
SINGAPORE

## THAILAND

### **DR. SANGOB LAKSANA**

Director  
Educational Assessment and Testing Services  
Dept. of Curriculum & Instruction Development  
Ministry of Education  
Bangkok 10300  
THAILAND

### **DR. AMORNWICH NAKORNTHAP**

Chief  
Cooperative Programs  
International Cooperation Division  
Ministry of University Affairs  
328 Sri Ayudhya Rd., Bangkok 10400  
THAILAND

### **MR. SOMKOAN POTHONG**

Instructor  
S. W. U.  
THAILAND

**DR. WERAWAT PUNNITAMAI**

Examiner  
Office of the Civil Service Commission  
(Recruitment & Selection Center)  
Dusit, Bangkok 10300  
THAILAND

**DR. WILARS SINGHAWISAI**

Secretary-General  
Office of the Civil Service Commission  
Phitsanulok Rd., Bangkok 10300  
THAILAND

**DR. RUTCHANA TAINSRI**

Consultant  
Human Potential Development  
61 Boonchusi Dindang, Bangkok 10400  
THAILAND

**PHILIPPINES**

**DR. THELMA G. ABIVA**

Professor  
Dept. of Educational Leadership and Professional Services (DELPS)  
College of Education  
University of the Philippines  
1101 Diliman, Quezon City  
PHILIPPINES

**DR. ADRIAN A. ARCELO**

President  
Fund for Assistance to Private Education  
7/f Concorde Condominium, Cor. Salcedo & Benavidez Sts.  
1299 Legaspi Village, Makati City  
PHILIPPINES

**DR. LETICIA M. ASUZANO**

President  
Center for Educational Measurement  
6/f Concorde Condominium, Cor. Salcedo & Benavidez Sts.  
1299 Legaspi Village, Makati City  
PHILIPPINES

**MS. DOLORES B. BONIFACIO**

Assistant Director  
Civil Service Commission  
Batasan Hills, Batasang Pambansa, 1126 Quezon City  
PHILIPPINES

**MS. ESPERANZA C. BUEN**

Director  
Center for Educational Measurement  
6/f Concorde Condominium, Cor. Salcedo & Benavidez Sts.  
Legaspi Village, 1229 Makati City  
PHILIPPINES

**DR. AIDA C. CALUAG**

Chairperson  
Education Department  
Ateneo de Manila University  
Loyola Heights, 1108 Quezon City  
PHILIPPINES

**DR. ANNADAISY J. CARLOTA**

Professor  
Department of Psychology  
College of Social Sciences & Philosophy  
University of the Philippines  
Diliman, 1101 Quezon City  
PHILIPPINES

**DR. CHONA D. CASTAÑO**

Senior Education Program Specialist  
Bureau of Elementary Education  
Dept. of Education, Culture & Sports  
Pasig, 1601 Metro Manila  
PHILIPPINES

**DR. ROSE MARIE S. CLEMENA**

Dean  
College of Education  
De La Salle University  
2401 Taft Avenue, 0922 Manila  
PHILIPPINES

**DR. MELECIO DEAUNA**

Professor  
Science Education  
De La Salle University  
2401 Taft Avenue, 0922 Manila  
PHILIPPINES

**DR. LENORE LL. DECENTECEO**

Associate Professor  
Department of Psychology  
University of the Philippines  
Diliman, 1101 Quezon City  
PHILIPPINES



**DR. ABRAHAM I. FELIPE**

President  
EDUVISION 2000  
2nd floor DAP Building, San Miguel Avenue  
Pasig, 1601 Metro Manila  
PHILIPPINES

**MS. ANA T. GUERRERO**

Supervising Psychologist  
Central Guidance Bureau  
Ateneo de Manila University  
Loyola Heights, 1108 Quezon City  
PHILIPPINES

**DR. BASILIO R. ILEDAN**

Director  
Center for Educational Measurement  
6/f Concorde Condominium, Cor. Salcedo & Benavidez Sts.  
Legaspi Village, 1229 Makati City  
PHILIPPINES

**DR. MIREN G. INTAL**

Executive Director  
Human Resources Center  
Ateneo de Manila University  
Loyola Heights, 1108 Quezon City  
PHILIPPINES

**MS. TERESITA G. JUNTILLA**

Administrative Officer  
Center for Educational Measurement  
6/f Concorde Condominium, Cor. Salcedo & Benavidez Sts.  
Legaspi Village, 1229 Makati City  
PHILIPPINES

**DR. MARCELINA M. MIGUEL**

Director IV  
Bureau of Elementary Education  
Dept. of Education, Culture & Sports  
Pasig, 1601 Metro Manila  
PHILIPPINES

**MS. AURORA P. MINA**

Senior Research Officer  
Center for Educational Measurement  
6/f Concorde Condominium, Cor. Salcedo & Benavidez Sts.  
Legaspi Village, 1229 Makati City  
PHILIPPINES

**DR. EUGENIA M. MORALEDA**

Chief  
Curriculum Development Division  
Bureau of Secondary Education  
Dept. of Education, Culture & Sports  
Pasig, 1601 Metro Manila  
PHILIPPINES

**DR. JESUS A. OCHAVE**

Dean  
Graduate College  
Philippine Normal University  
Taft Avenue, Manila  
PHILIPPINES

**MR. HERMOGENES P. POBRE**

Chairman  
Philippine Regulation Commission  
P. Paredes St., Sampaloc, Manila  
PHILIPPINES

**MS. GILDA LOIDA R. REYES**

Senior Research Officer  
Center for Educational Measurement  
6/f Concorde Condominium, Cor. Salcedo & Benavidez Sts.  
Legaspi Village, 1229 Makati City  
PHILIPPINES

**MS. MA. ANGELES A. SAMPANG**

Senior Research Officer  
Center for Educational Measurement  
6/f Concorde Condominium, Cor. Salcedo & Benavidez Sts.  
Legaspi Village, 1229 Makati City  
PHILIPPINES

**DR. LIDINILA M. LUIS-SANTOS**

Chief  
Curriculum Development Division  
Bureau of Elementary Education  
Dept. of Education, Culture & Sports  
Pasig, 1601 Metro Manila  
PHILIPPINES

**DR. LUCILA F. TIBIGAR**

Director III  
National Educational Testing & Research Center  
Dept. of Education, Culture & Sports  
Pasig, 1601 Metro Manila  
PHILIPPINES

**DR. GUNDELINA A. VELAZCO**

Director  
Institute of Testing and Evaluation Office  
De La Salle University  
2401 Taft Avenue, 0922 Manila  
PHILIPPINES

**MS. MIRAFLOR M. VILLEGAS**

Program Officer  
Center for Educational Measurement  
6/f Concorde Condominium, Cor. Salcedo & Benavidez Sts.  
Legaspi Village, 1229 Makati City  
PHILIPPINES

