



**USAID** | **INDONESIA**  
FROM THE AMERICAN PEOPLE



# AN EXPLORATORY STUDY OF THE UJIAN NASIONAL

A STUDY PREPARED FOR DECENTRALIZED BASIC EDUCATION – 3

A BASIC EDUCATION DEVELOPMENT PROJECT SUPPORTED BY  
THE UNITED STATES AGENCY FOR INTERNATIONAL DEVELOPMENT (USAID)

**ROBERT CANNON** with **RINA ARLIANTI**





# **AN EXPLORATORY STUDY OF THE UJIAN NASIONAL (SMP/MTs)**

A STUDY PREPARED FOR DECENTRALIZED BASIC EDUCATION – 3

A BASIC EDUCATION DEVELOPMENT PROJECT SUPPORTED BY  
THE UNITED STATES AGENCY FOR INTERNATIONAL DEVELOPMENT (USAID)

**ROBERT CANNON with RINA ARLIANTI**

JAKARTA, INDONESIA

27 NOVEMBER 2009

**COVER PHOTOGRAPH**

Students undertaking UN, Banjarmasin, 2008





# CONTENTS

<b>1 EXECUTIVE SUMMARY AND RECOMMENDATIONS.....</b>	<b>1</b>
1.1 Background .....	1
1.2 Methodology .....	1
1.3 Issues .....	1
1.4 Recommendations for Reform .....	2
<b>2 BACKGROUND.....</b>	<b>5</b>
2.1 The Purpose And Scope of the Study.....	5
2.2 Methodology .....	5
<b>3 EXAMINATIONS IN INDONESIAN BASIC EDUCATION.....</b>	<b>7</b>
3.1 Rationale .....	7
3.2 The Content and Construction of the UN.....	8
3.3 Grading.....	11
3.4 Weighting.....	12
3.5 UN subjects – Specific Issues .....	12
3.6 The Social and Economic Impact of Unsatisfactory Public Examinations .....	13
<b>4 INTERNATIONAL COMPARISONS.....</b>	<b>15</b>
4.1 The Philippines.....	15
4.2 China .....	15
4.3 Singapore.....	17
4.4 Finland.....	17
4.5 International Studies of Educational Reform .....	18
4.6 International Comparisons of Student Performance.....	18
4.7 Discussion .....	19
<b>5 THEORETICAL AND TECHNICAL QUALITIES OF ASSESSMENT.....</b>	<b>21</b>
5.1 Introduction .....	21
5.2 Validity.....	21
5.3 Reliability.....	22
5.4 Practicality.....	23
5.5 Discussion .....	24
5.6 High Stakes Testing .....	24
<b>6 EXAMINING THE EXAMINATION: THE QUALITY OF UN TEST PAPERS.....</b>	<b>27</b>
6.1 Introduction .....	27
6.2 Evaluation 1: Analysis of 2007/2008 Test Papers.....	27
6.3 Evaluation 2: The 2008/2009 Bahasa Inggris Test Paper .....	27
6.4 Evaluation 3: The 2003/2004 Bahasa Inggris Test Paper .....	28
6.5 Conclusion.....	28
<b>7 THE CURRENT DEBATES .....</b>	<b>33</b>
7.1 Content and Construction of the UN.....	33
7.2 The Usefulness of the UN .....	33
7.3 Administration of the UN.....	33
7.4 The Impact of the UN.....	36
7.5 A Missing Link – High Expectations? .....	40
7.6 Objectivity of the UN.....	40
7.7 Cost Effectiveness .....	41
7.8 The Need for a National Examination.....	41
7.9 Forces For and Against Reform .....	42

<b>8</b>	<b>CONCLUSION AND RECOMMENDATIONS</b> .....	<b>43</b>
8.1	General Conclusion.....	43
8.2	Recommendations for Reform.....	44
<b>9</b>	<b>REFERENCES</b> .....	<b>51</b>
<b>10</b>	<b>SCHOOLS AND INSTITUTIONS VISITED</b> .....	<b>53</b>
	<b>APPENDIX 1: OVERVIEW OF THE UN IN INDONESIA</b> .....	<b>54</b>
	<b>APPENDIX 2: MINIMUM SERVICE STANDARDS, THE NATIONAL STANDARDS AND SCHOOL ACCREDITATION</b> .....	<b>59</b>
	<b>APPENDIX 3: EVALUATION OF RECENT NEWSPAPER STORIES</b> .....	<b>63</b>
	<b>APPENDIX 4: SAMPLE TEST ITEMS</b> .....	<b>67</b>

## LIST OF ABBREVIATIONS

APBN	Anggaran Pendapatan dan Belanja Negara (Budget Revenue and Expenditure)
AusAID	Australian Agency for International Development
BSNP	Badan Standar Nasional Pendidikan (National Education Standards Authority)
CRA	Criterion referenced assessment
GoI	Government of Indonesia
ICW	Indonesia Corruption Watch
IPA	Ilmu Pengetahuan Alam (Science)
JSE/JSS	Junior Secondary Education/School(s)
KKS	Kelompok Kepala Sekolah (Principal's Group)
MCQ	Multiple Choice Question
MA	Madrasah Aliyah (Islamic Senior Secondary School)
MoNE	Ministry of National Education
MoRA	Ministry of Religious Affairs
MTs	Madrasah Tsanawiyah (Islamic Junior Secondary School)
NGO	Non Government Organisation
NRA	Norm referenced assessment
PISA	Programme for International Student Assessment
POS	Prosedur Operasional Standar (Standard Operating Procedure)
Puspendik	Pusat Penilaian Pendidikan (Educational Assessment Centre)
Renstra	Rencana Strategis (Strategic Plan)
SMA	Sekolah Menengah Atas (Senior Secondary School)
SMALB	Sekolah Menengah Atas Luar Biasa (Senior Secondary School for special needs children)
SMK	Sekolah Menengah Kejuruan (Vocational Secondary School)
SMP	Sekolah Menengah Pertama (Junior Secondary School)
SMPLB	Sekolah Menengah Pertama Luar Biasa (Junior Secondary School for special needs children)
sms	Short message service
TIMSS	Trends in International Mathematics and Science Study
UAS	Ujian Akhir Sekolah (End of School Examination)
UN	Ujian Nasional (National Examination)

UNICEF

United Nations Children's Fund

USAID

United States Agency for International Development

## **PREAMBLE**

### **Campbell's Law**

This Law states:

“The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor” (Nichols and Berliner (2005; 2007, p. i).

The Ujian Nasional is a social indicator of student achievement at the end of nine years of basic education. It is also an indicator in the management and development of school quality.

### **A Century of Progress in Examinations?**

It is a very great and more serious evil to sacrifice systematic instruction and a comprehensive view of the subject for the scrappy and unrelated knowledge gained by students who are persistently drilled in the mere answering of questions issued by the Education Department or other governing bodies.

Department of Education of the State of New York, 1906

Cited in: Favato, P., Mathison, S. & Calalano C. (2003). *A murmur of dissent: A story of resistance to high-stakes testing in New York State*. Paper presented at the meetings of the American Educational Research Association, Chicago, IL.

## **ACKNOWLEDGEMENTS**

This Report has been prepared by Robert Cannon and Rina Arlianti at the request of the USAID supported project, Decentralized Basic Education – 3 (DBE-3). The studies for the Report were carried out in Indonesia in October and November 2009.

The authors particularly acknowledge the professional advice and guidance provided by Stuart Weston, Lorna Power and Theresa Wikaningtyas of DBE-3 and the advice and assistance of stakeholders consulted for this study.

Any errors or omissions are regrettable but, of course, as with the whole report, the responsibility of the Lead Consultant and main author, Robert Cannon.

## **DISCLAIMER**

The views expressed in this publication do not necessarily reflect the views of the Government of Indonesia, the Government of the United States of America, and The United States Agency for International Development, or Decentralized Basic Education – 3.

## **IMPORTANT LATE DEVELOPMENT**

On 25 November, 2009, as this study was concluding, the Supreme Court of Indonesia made a major ruling on the Ujian Nasional. The ruling, according to press reports available at that time, is that the government must revoke the Ujian Nasional.

The full implications of the revocation cannot be determined at this time nor is the formal response of the Ministry of National Education known. The ruling adds further credibility to some of the general findings reported here. More information is provided in Appendix 3.

# **1 EXECUTIVE SUMMARY AND RECOMMENDATIONS**

## **1.1 Background**

This exploratory study of the Ujian Nasional (UN) seeks to identify and understand the issues that are paramount in stakeholders' minds and to identify the forces for and against the reform of school assessment. The study reaches broadly similar conclusions to another recent study undertaken by the Australian Council for Educational Research, supported by AusAID in 2008, but goes further than that largely technical analysis by examining in some of the key contextual issues that are of critical significance to the quality of the UN.

The Ujian Nasional (UN), or National Examination, is a Government of Indonesia school examination administered at the end of nine years basic education. The UN assesses four subjects – Bahasa Indonesia, English, Mathematics and Science and its primary purpose is to assist in determining the graduation of students from school and for entry to senior secondary education. A secondary purpose is to provide data to map school and program quality to help determine resources to be given to schools to improve educational quality.

Because of the significant role of the UN in determining these resources and in deciding children's future, it is described in this Report as 'high stakes'.

## **1.2 Methodology**

The study explored issues from the literature, policy documents and with stakeholders in the field. Fieldwork included the gathering and analysis of data and public documents including media reports, visits to schools, small group discussions with students, principals, teachers and parents, workshops with teachers, analytical studies of test papers, meetings with District, Provincial and Central Government officials, educational consultants, university academics, and community organisations. From these data gathering processes clear patterns of issues began to emerge and these patterns are reflected in this Report.

## **1.3 Issues**

The study found congruence between research in the USA on the impact of 'high stakes' testing there and the present conditions in Indonesia where the UN is considered to be high stakes. The research finds that as the stakes associated with a test go up, so does the uncertainty about the meaning of a score from that test and so do the distortions, corruption, and negative backwash effects on the quality of education and on students.

Because of this important finding, the study makes a series of closely linked recommendations designed to reduce these distortions and to increase the quality of the beneficial impacts the UN may have on education and on children in line with developments in some of the best performing education systems in the world today.

The evidence located for this study makes its own, strong case for significant reform of the UN. The study recognises that there is no clear impetus for abandoning the UN. The approach here is to accept that the UN will continue but that substantial change is urgent to address the credibility of the award and to mitigate the negative impacts including on Indonesia's aspirations to improve its international competitiveness through educational excellence.

The key issue to be addressed is to enhance the credibility of the UN and its annual implementation. The credibility of the UN is eroded by four factors.

**Factor 1: A weak foundation of professional knowledge**

The first factor is a weak foundation of professional knowledge about student assessment across the education workforce. This weakness is directly related to widespread concerns about teacher quality and the study notes that government strategies are in place to address this (Jalal, et al., 2009). The weakness in professional knowledge extends to educational administrators who do not have a professional background in education. The wider community is similarly not well informed about assessment and this matter warrants attention in change strategies as well to generate informed debate and demand for better quality in educational assessment.

**Factor 2: Poor professional and ethical standards**

The second factor is the poor standard of professional and ethical behaviour among many who have children in their care. There is much evidence of widespread, systemic and entrenched malpractice in the implementation of the UN. Malpractice invalidates the data the examination process produces. There is also widespread encouragement from adults of student cheating to achieve good results.

**Factor 3: Educational and technical weaknesses in design**

The third factor is educational and technical weaknesses in assessment design. This factor includes the weak alignment of the centralized, national examination system with the decentralized, local school curriculum and the distortions in the educational process created by the UN. It includes the technical design of the examination, test items and implementation procedures of the UN.

**Factor 4: Unacceptable educational risk from high stakes testing**

Finally, in this professionally and ethically weak environment, the high stakes UN increases the scale and distribution of the Factor 2 ethical issues and decreases the educational and administrative utility of the UN. Emerging processes to use examination data in quality assurance and standards assessment only increases the stakes and the distortions.

**1.4 Recommendations for Reform**

The recommendations are based on an understanding of these four factors. These factors need to be addressed holistically, systematically and in the context of Indonesia's reforming and educational system. By recognising the power of assessment, addressing it properly will add impetus to the whole reform agenda by driving improvements deeper and further.

The recommendations are made on the assumption that the UN should continue but in a substantially improved form. There is clear evidence that the UN needs to be improved and little evidence of demand for its abolition at the present time. Three key targets for improvement are proposed having been identified in the study. They are:

**Target 1: A professional, integrated and aligned UN system**

**Target 2: An ethical education system**

**Target 3: A better quality, credible and flexible credential at the end of Year 9**

To help achieve the three targets, seven recommendations are made. Each recommendation is accompanied with a more detailed rationale in Chapter 8, and an estimation of the extent to which donors may help in realizing each recommendation.

**Recommendation 1: Increase and sustain all efforts to improve the quality of the teaching workforce specifically and to improve basic education generally**

This is the foundation for all recommendations. The recommendation is based on the unequivocal experience of positive educational reform in the past two decades, both in Indonesia and internationally.

**Recommendation 2: Increase the supply of high quality assessment leadership**

The most urgent reform is to institute a revised system of governance and management of the assessment process under a single, independent and professional authority that reflects international best practice. The authority must have a clear commitment to the highest integrity and professional standards and the capacity to eliminate direct interference, malpractice and cheating.

**Recommendation 3: Increase the demand in society for high quality assessment**

This is a continuous and long-term goal for the assessment authority to achieve. The experience of children in schools with better quality assessment will build a better-informed citizenry in the long term but increased demand in the short term will also come from:

- Advocacy campaigns to 'socialize' the importance of quality assessment in schools
- A public consultation process to both raise awareness and to seek systematic feedback on assessment proposals
- Exploring strategies to strengthen the appropriate role of school committees and communities in monitoring the assessment of student learning in schools.

**Recommendation 4: Align curriculum, learning and teaching, and assessment**

The concept of 'alignment' describes the essential connection between the curriculum, learning, teaching and assessment. Alignment is fundamental to valid assessment and valid assessment is the foundation upon which any credible national examination system is built. This study finds an unacceptable lack of alignment among the key elements of the educational process in schools and consequential low validity in assessment data. The tension between a centralized, national examination and decentralized school-level curriculum must be addressed.

**Recommendation 5: Increase efforts through research and management to better understand the mechanisms of malpractice and cheating and implement coordinated policies and practices to eliminate it from the education sector**

The full extent of malpractice and cheating in the assessment system is unknown. The critical issue is that it is openly acknowledged, widely known to exist and that its scale and existence means that test scores lack validity and any practical utility. Strategies to address the recommendation are provided in Chapter 8.

**Recommendation 6: Create and develop a professional and ethical environment in which corruption and cheating is denied the opportunity to exist**

The strategies for Recommendation 1, above, provide the foundation for this recommendations to build an ethical system. The best long-term strategy to defeat educational malpractice is to build a quality system that prevents it from happening.

**Recommendation 7: Develop a high quality and credible credential at the end of year 9 through undertaking a systematic and professional review of the UN, with international best practice input**

The present arrangements for the UN are not providing students with a credible credential.

Because of a lack of validity and reliability, the same data is unsuitable for use in educational planning. The present arrangements are distorting the educational process in schools, narrowing the curriculum, and driving teaching and learning strategies away from developing the competencies required in the curriculum towards superficial and trivial learning. The current UN is making it less likely that Indonesian students performance in internationally benchmarked tests, PISA and TIMSS, will lift the nation's position into the ranks of the better performing nations

## 2 BACKGROUND

### 2.1 The Purpose And Scope of the Study

This exploratory study of the Ujian Nasional (UN), the National Examination, has two main purposes.

The first purpose is to review key issues so as to better understand:

- the issues and the forces for and against reform.

The second purpose of the study is to establish a greater understanding of reform needs (if any) so as to:

- improve the UN
- indicate where and how technical assistance from donors might be effective in improving the UN and
- indicate potential obstacles.

The study is undertaken in the broader context of DBE-3's work to improve the quality and relevance of junior secondary education and to support greater harmonization and coordination between the Government of Indonesia and the donor community. In this context, the study therefore focuses on the UN at the end of junior secondary education but many of its observations and conclusions are equally relevant to the Year 12 UN as well. The study of the UN is important because of the powerful influence examinations of this nature have on learning and teaching and, indirectly, on the processes of change and development that projects such as DBE-3 strive to influence in schools.

The study reaches broadly similar conclusions to another recent study undertaken by the Australian Council for Educational Research, supported by AusAID in 2008, but goes further than that largely technical analysis by examining in some of the key contextual issues that are of critical significance to the quality of the UN.

### 2.2 Methodology

This is an *exploratory* study. The major purpose is to identify and understand the issues that seem paramount in stakeholders' minds, and the forces for and against the possible reform of student assessment<sup>1</sup>. The consultants believe this purpose has been achieved.

Having established the basic framework for the study, timelines, and the key concerns of the client, the study proceeded to explore issues from the literature, policy documents and with stakeholders in three locations – North Sumatra, South Sulawesi and Jakarta. Fieldwork included the gathering and analysis of data and public documents including media reports, visits to schools, small group discussions with students, principals, teachers and parents, workshops with teachers, analytical studies of test papers, meetings with District, Provincial

---

<sup>1</sup> The terms 'assessment' and 'evaluation' are frequently used interchangeably. In this Report, assessment refers to the collection of data about student achievements, mainly through formal examinations (the UN) and school exams. Assessment data can be used as part of an evaluation of schools, programs, teaching quality and even the UN itself. Nevertheless, it is noted that in the National Standards for Education, the term 'evaluation' is used to mean assessment. The terms 'test' and 'examination' are used interchangeably to have the same meaning, although examination is the more appropriate for use in the more formal sense of the National Examination' or Ujian Nasional (UN).

and Central Government officials, educational consultants, university academics, and community organisations. From these data gathering processes clear patterns of issues began to emerge and these patterns are reflected in this Report.

It is emphasised that the study is exploratory and the purpose is to identify and understand issues. The study does not claim to have identified all the issues, as this would have been impossible in the limited time available. An issue is put forward only where this has been identified by at least two independent sources or from at least two different locations.

### **3 EXAMINATIONS IN INDONESIAN BASIC EDUCATION**

Basic education comprises primary schools (years 1-6) and junior secondary schools (years 7-9). Both state schools and private schools make up the sector and both are included in this study. Two ministries are responsible for managing basic education, with 82 percent of schools under the Ministry of National Education (primary and junior secondary schools, (abbreviated SD and SMP) and 18 percent under the Ministry of Religious Affairs (Islamic primary and junior secondary schools, MI and MTs). The majority of religious schools are private schools supported by a range of mostly Islamic religious institutions and foundations.

The Ujian Nasional (UN), or National Examination, is a Government of Indonesia school examination administered at the end of nine years basic education and at the end of twelve years of senior secondary education. There is no clear distinction made in nomenclature between these two examinations and this leads to confusion. Discussions about the UN for SMP frequently become confused with references to the UN at the end of senior secondary education. The Ministerial decree uses the terms 'UN SMP/MTs/SMPLB', 'UN SMA/MA/SMALB' and 'UN SMK'.

The term 'UN', when used in this Report, is meant to apply to the examination for students completing basic education in SMP and MTs. Further formal detail about the UN is provided in Appendix 1. Space regrettably does not permit discussion of the UN in relation to special schools (SMPLB) or the Paket B exam in non-formal education.

#### **3.1 Rationale**

The UN is the assessment of student learning at the end of Year 9. It is conducted by the Government nationally for four subjects – Bahasa Indonesia, English, Mathematics and Science (Government Regulation No 19 year 2005).

The primary purpose of the UN is to assist in determining the graduation of students from school and for entry to senior secondary education. The Regulations require more than the UN alone in making this determination. In addition, the students are declared to have passed the education unit of the basic and intermediate education levels only after they have:

- Completed the learning program;
- Achieved minimal marks at the final evaluation for all school-assessed subjects
- Passed the school/madrasah examination; and
- Passed the National Examination (UN).

Further detail on these formal requirements is provided in Appendix 2, Part Five.

The graduation of students from schools is determined by the schools concerned in conformity with criteria developed by BSNP (Government Regulation, 19, year 2005, Part 5).

The second purpose of the UN is to provide data to map school and program quality to help determine resources to be given to schools to improve educational quality. Because of the significant role of the UN in determining these resources and in deciding children's future, it is described in this Report as 'high stakes'.

The consensus view among students, parents, teachers and administrators is that it is necessary to have some kind of UN, but not necessarily in its present form. Among the teachers met, only one argued persuasively that the UN should be abandoned and this was on the grounds of the educational damage it does to her teaching subject, Bahasa Indonesia.

However, there is an equally widespread view that the UN should not be the sole determinant of graduation and that a component of school-assessed work should be a necessary requirement. This view suggests that the Regulation 19 described above may be ignored in too many cases.

Educational administrators argue that the UN is necessary for planning purposes, to set standards and to meet the need for student performance data. Two further reasons are advanced to support the UN. The first reason is that the UN acts as a motivator for student learning. The second reason is that the examination results provide useful planning information. However, this second argument is seen by some as shallow and lacking in seriousness. This is because there is pressure on schools from above to 'achieve' high graduation rates with the threat of possible dismissal if principals do not 'achieve' the expected outcomes for their schools. For students who fail the UN, there is encouragement to complete the examination for informal education 'Paket B' that almost always leads to 'success'.

### **3.2 The Content and Construction of the UN**

#### **Foundations**

The content and construction of the UN is based on the curriculum for junior secondary education. The complex process is presented in full here as it leads to some fundamental conceptual issues that should be addressed. The academic year 2008/2009 is used as a reference.

The foundation for the UN lies in the development of the *Kisi-kisi Soal*, the Table of Specification for Test Items. The steps, followed by the BSNP and the Educational Assessment Centre, in the development of the Table are:

- Identify the Standard of Graduate's Competencies of subjects to be tested in UN from Curriculum 1994, Curriculum 2004 and the Standard of Content. The Standard of Content contains the scope of content and competency levels to reach school graduate's competencies.
- Identify the Standard of Graduate's Competencies of subjects from the three curriculum documents that become the Standard of Graduate's Competencies of subjects to be tested in the UN 2008/2009 academic year.
- The Table of Specification for Test Items, the *Kisi-kisi Soal*, is developed from the Standard of Graduate's Competencies of subjects to be tested in the UN by involving lecturers, teachers and experts in educational assessment.
- Conduct validation of the Table of Specification for Test Items by involving lecturers, teachers and experts in educational assessment.
- Propose the Table of Specification for Test Items to the Minister of National Education to be established as the formal *Kisi-kisi Soal* UN 2008/2009

#### **Alignment**

There is a potential and sometimes real lack of alignment in the tests that emerge from this process. The lack of alignment occurs between the actual content of the school curriculum and what is in the Table of Specifications, and also between the Table and what is in the examination papers.

This lack of alignment, which is essential for a valid test, is demonstrated in Table 3 below. Here, expert reviewers of the four UN test papers, report that only two of the four papers meet the criterion 'Test items truly reflect the *Kisi-kisi Soal UN*'.

The complex challenge of aligning what is actually taught in schools and what may be designed into the Table of Specifications, and therefore the UN, is revealed from an analysis of the curriculum development process. Again, the process begins with the Standard of Content that contains the scope of content and competency levels to reach school graduate's competencies. The Standard consists of a curriculum framework and structure, Competency Standards, and Basic Competencies for each semester and education level.

The Standard of Graduate's Competencies is the statement of graduate's competencies that includes attitudes, knowledge and skills.

### **School Level Curriculum**

The operational curriculum at school level is known as the KTSP – *Kurikulum Tingkat Satuan Pendidikan* – the School Level Curriculum. It consists of educational objectives for SMP/MTs, curriculum structure and content, the education calendar and syllabus for each subject. Finally, the subject syllabus itself contains competency standards, basic competencies, learning content, learning activities, indicators for achievement of competencies for school assessment, assessment, time allocation and learning resources.

The official document: Badan Standar Nasional Pendidikan. 2006. *Panduan Penyusunan Kurikulum Tingkat Satuan Pendidikan Jenjang Pendidikan Dasar dan Menengah*, sets out the following 'Principles in Developing School Based Curriculum'.

The School Level Curriculum, or KTSP, is developed in relation to the relevance for each school of the Standard of Content and Standard of Graduate's Competencies. The School Level Curriculum is developed based on these educational principles:

- Focus on the needs and concerns of students in their local environment
- Diversity and integration
- Responsiveness to the development of science, technology and art
- Relevant to the needs for life
- Holistic and sustainable
- Lifelong education
- A balance between national and regional interests.

### **Educational Professionalism in Dinas Pendidikan**

The process of constructing the School Level Curriculum is supervised by Dinas Pendidikan or by Depag. In several parts of this Report, concern is expressed with standards of teacher professionalism, including their knowledge and skill in assessing student learning. However, the evidence from the field and from references to published accounts should deepen this concern when it comes to considering the professionalism of some senior staff in Dinas Pendidikan. The following quotation from *The Jakarta Post*, Wednesday, 21 January 2009, demonstrates the concern:

"Around half the education agency heads across the country's 33 provinces are 'incompetent' and only secured their positions because of campaigning they conducted for incumbent administration chiefs, a survey says. The Indonesian Teachers Association (PGRI), which conducted the survey, said in Jakarta on Tuesday most education office heads in regions are former members of political teams linked to incumbent governors, mayors and regents. '50 percent of them do not have the sufficient technical competency to head these education agencies,' PGRI chairman Sulistiyo said after meeting with President Susilo Bambang

Yudhoyono at the presidential palace in Jakarta. 'Some of them are former heads of local civilian police units; others are former heads of parks, special planning and public cemetery agencies or even former district heads. In short, many have never handled education issues before,' he said. "

Any assessment system that includes administrators from local police, parks, and public cemeteries in the professional management of education generally, and the implementation of the highly technical domain of assessment specifically, is clearly in very serious difficulty. This difficulty is abundantly demonstrated in the extent of malpractice in the local administration of the UN. It is quite unreasonable to expect people from such backgrounds to be able to comprehend the full *educational* impact of any administrative decisions they may have to make in the course of their duties.

### **Alignment again**

From this presentation of detail it may be argued that what is in the UN may not be well aligned with what has been taught in each school if schools follow the School Level Curriculum. There is evidence that none of these School Level Curriculum principles are reflected in the UN. The argument can be reversed: given that we understand the power of examinations to exert more influence on learning than the curriculum, we find that the present UN compromises all of these School Level Curriculum principles.

The UN does not accurately measure what is taught in the curriculum because the chosen format, Multiple Choice Question (MCQ), does not allow the testing of attitudes and skills. In that sense, the UN measures cognitive skills and can hardly be considered as a valid and reliable indicator of important life skills that may have been developed in the school.

### **School graduation**

The relative importance of the UN against all other elements listed for graduation means that this is the one assessment that drives teaching and learning in Year 9 to the almost complete exclusion of the affective and practical competencies specified in the Standard.

Some will argue, reasonably, that the UN is not the only requirement for graduation and that the skills and attitudes in the School Level Curriculum are assessed in the school Ujian Akhir Sekolah (UAS). This is true, but evidence from some schools suggests that, in reality, the school results may count for little if the UN is passed. There is no evidence in the Regulations about a standard method for combining the outcomes of the UN and school tests to determine graduation, only a statement that students must pass both. This is a critical point for without a clear method there can be no reliability from the range of processes used (or not used). The method needs to be very clearly spelled out for schools, and it is probably why the UN is so popular for graduation purposes.

### **Confused models of assessment**

Hidden in this complex process is an extremely important confusion of purposes and processes that significantly complicates matters even further. The confusion lies in conflating two very different models of assessment.

The first is a norm-referenced assessment model (NRA), designed to assess student achievement on a numerical scale and then make comparisons between students to assign a grade and determine pass-fail. This norm-referenced model, based on the psychometrics of individual differences in large populations, makes a serious, fundamental error in its educational assumptions. That assumption is simply the opposite of what schooling is trying to achieve. That assumption in the NRA is that no matter how successful the learning, how good the teaching, how committed and competent the students may be, the norm-referenced, 'grading-on-the-curve philosophy will guarantee that some students will fail.

This thinking is strongly rooted in Indonesian practice. For example, one recent study suggests that to measure the competency of test participants with a variety of abilities in the UN, the test used should have items with different difficulty levels to be fair and to get accurate results, in other words to set out to differentiate among students rather than testing them to see if they have actually achieved the specified competencies, which is what the Standard requires (Setiadi, 2007).

The aim of good teaching is to work hard to *reduce* the differences between students and help *all* to succeed, not to work towards separation and the inevitable failure of some by 'spreading the scores' on a bell curve. There is no profession other than teaching that sets out with this 'bell curve thinking' that then drives all its efforts to ensure the failure of some, no matter what. In addition, the scores produced by this NRA approach, say a score of 5.5, tell us nothing at all about what has been learned.

The second assessment model is the standards model or criterion referenced assessment (CRA). This is the model implied in the discussion here about competency standards that follows the long process described above, beginning with the Standard of Graduate's Competencies. A careful reading of the National Standards of Education can leave no doubt whatsoever that the intention is to assess *competencies*. Article 66 is one of many articles that make this clear: "The evaluation of learning results as specified in Article 63 para (1) sub c has the purpose of evaluating the competence achievement of the graduates. Appendix 2 sets out these articles in more detail.

Yet, against all of this, the UN is then based on principles of norm-referenced assessment. Scores produced by this approach can tell us nothing about what has been learned and the standards of student competence because the score is referenced to a norm, not a competency standard.

### **The threat from complexity**

The complexity of the UN design and administration processes, together with the complexity and number of laws, regulations and standards poses a threat to the overall quality of the UN. Unnecessarily complex systems almost always lead to administrative failure of one kind or another and also provide opportunities for exploitation by the corrupt and dishonest.

The strategy to address complexity is to focus on simplicity. This means that, all other things being equal, the simplest solution is the best. Test design simplicity should be a key goal, and unnecessary complexity avoided. When alternative strategies are considered, the simplicity principle recommends selecting the strategy that introduces the fewest assumptions, minimal complications and involves the smallest number of administrative entities, a key reason for supporting a single assessment authority.

Simplicity alone is insufficient. Simple strategies must be based on good evidence that the strategy will work. At the moment, the UN seems to reflect the opposite – a strategy based on poor evidence, the maximum number of entities involved in its administration, many complicated regulations, and mistaken assumptions about education and assessment.

### **3.3 Grading**

The rule for graduation is specified ahead of the UN. The standards are set by BSNP and the Educational Assessment Centre does the technical work. There is an attempt to lift standards by changing the rule from one year to the next as shown below:

Participants will pass when they fulfil the UN graduation standard (2007/2008) as follows:

- Minimum average score of 5.25 for all subjects examined, with no score below 4.25,

- or
- Minimum score of 4.00 for one subject and all other subjects minimum 6.00 and will be used to calculate average UN score.

For 2008/2009 the UN graduation standard is:

- Minimum average score of 5.50 for all subjects, with minimum score of 4.00 for maximum 2 subjects and minimum 4.25 for all other subjects.

### **3.4 Weighting**

At the moment, the scores from the UN are not weighted or adjusted according to Provincial or other criteria. A student's score will be the same irrespective of whether as student has attended the best junior secondary school in metropolitan Jakarta or an impoverished madrasah in the remotest island in eastern Indonesia.

The matter of whether or not weighting should be applied was raised with stakeholders. The consensus view from the Provincial, District and school levels is that scores should be weighted. Reasons advanced to support this view include:

- It will stimulate a sense of positive competition within the Province
- National-level standards are too high for some local areas and too low for others
- It is unfair to judge students from many private MTs against the same criteria as the majority of these schools have poorer facilities and poorly trained teachers.

Accordingly some stakeholders argue for a District or Provincial-level system of weighting to compensate for regional disadvantage. However, if this idea were to be implemented, it would introduce another equity issue in that students attending very good schools in a disadvantaged region would receive an advantage over students attending comparatively poor schools in more advantage regions.

But quite apart from the equity issue, weighting according to location or school makes the idea of a 'national examination' nonsense. The UN is a measurement tool. A system of weighting in a national examination to achieve equity goals would be similar in principle to adjusting a thermometer down by a few degrees because it is hot in Indonesia and up a few degrees because it is so cold in Mongolia to make the residents in each place feel better.

### **3.5 UN subjects – Specific Issues**

Recognising that the MCQ test format is a practical solution to testing such a large student population at the national level, there is nevertheless widespread concern about the validity of the inferences that can be drawn from the data the MCQ format generates, the negative impact this test form has on learning and teaching, and the unacceptable quality standards in test items on display in test papers. Similarly, there is shared concern about the lack of reliability of a test that produces results based on a one two-hour test paper of three years of junior secondary education. Including a statistically moderated school assessed input for each subject can make an improvement to both validity and reliability, provided that cheating is under effective control.

Specific issues and concerns expressed by stakeholders with each subject, mainly teachers, are presented below.

#### **3.5.1 *Bahasa Indonesia***

Teachers of Bahasa Indonesia expressed more concerns about the examination for their

subject than any other subject group.

First, teachers feel that Bahasa Indonesia is not regarded highly in schools as it is the one subject that does not have any competitions associated with it that tend to encourage a better level of effort among students. At the time of the UN, Bahasa Indonesia is always the first test to be administered to students and so there is a view that students therefore face greater levels of stress in relation to this subject than others. Finally, there were complaints about the difficulty of test items that are sometimes long reading passages with few questions attached to the passages.

Some teachers believe the UN is consistent with curriculum requirements and that the existence of the UN acts as a strong motivating factor for student learning. Counter to this, is the realization among others that the multiple-choice format cannot test oral language skills and free writing. Teachers point out that local languages are still used in some schools and this weakens Bahasa Indonesia ability. The exam is criticised for being too theoretical and using terms not well understood by students.

### **3.5.2 English**

English shares some of the same issues with Bahasa Indonesia. However, English is widely regarded as a much easier examination than Bahasa Indonesia because of its shorter and clearer test items. Teachers feel that the MCQ test format has a negative impact on learning English as it eliminates the motivation to learn and develop the essential communication skills of writing, speaking and listening. This contributes to the very low level of English language skill so commonly found in schools. There is, as one teacher said, a 'disconnect' between what is wanted by the curriculum and what is delivered via the UN system.

### **3.5.3 Science (IPA)**

Science teachers have the view that their subject is at a distinct disadvantage because students are required to develop competence in three different sciences – Physics, Chemistry and Biology, yet there is still only one two hour paper to test achievement. To improve the situation of IPA, a further test paper has been recommended but one using different types of test items to MCQs. The quality of test items is reported to remain unchanged from year to year and this influences students and teachers to focus on 'training' for the exam and discourages any personal development.

### **3.5.4 Mathematics**

Mathematics teachers are generally more comfortable with the UN than their peers in other subjects. There is general concern, however, with the impact of the test format on student learning. The MCQ test format rewards the idea of rote learning and encourages drilling of students without them learning the fundamental mathematical concepts. Nor does the format require demonstration of problem-solving processes.

### **3.5.5 Other subjects in the curriculum**

There is almost universal agreement among stakeholders that the UN, examining as it does only Bahasa Indonesia, English, Mathematics and Science, harms the attainment of the broader goals of the curriculum and particularly the importance of other subjects in student's and teacher's value system. This is not only because of a perception of lack of prestige among these subjects but also because teaching time is taken away from them to enable UN subjects preparation courses to be conducted in the schools.

## **3.6 The Social and Economic Impact of Unsatisfactory Public Examinations**

A poor examination system has profound, large-scale social and economic effects apart from 'playing games' with children's lives. As data generated by the examination is unreliable and invalid, large numbers of candidates are passing the examination that should not pass and equally large numbers are failing who should not have failed. At junior secondary level, at least society is spared the dreadful consequences of graduating lawyers, airline pilots and civil engineers who should never be passing for reasons of their principles, ignorance or incompetence. But there are serious consequences at the junior secondary level, nevertheless, which are hidden and easily overlooked.

Failure at the UN and then leaving school or repeating a grade places those children at risk in later life through not possessing sufficient life-skills and qualifications to participate effectively in family, social and working life (Mortimore and Mortimore, 1999, 5). Non-completion is an indicator of future problem behaviour and poor mental and physical health (Battin-Pearson, S, et al., 2000: 568). Apart from the personal costs of examination failure and non-completion of schooling there are long-term hidden costs to a society. Early school leaving may result in:

- Decreased health status
- Increased crime
- Increased social welfare costs arising from the above
- Taxation revenue losses to the state
- Reduced participation in community life and contributions to society as an informed citizen,

Indonesian society has a considerable, long-term economic interest in ensuring that children complete a full program of formal education and that the examination arrangements support this goal. At this time, there is much evidence that the opposite is occurring.

## **4 INTERNATIONAL COMPARISONS<sup>2</sup>**

Two questions are posed for consideration when looking elsewhere for ideas, benchmarks and strategies in assessing student learning at school.

The first question should be an important part of any national debate about the purposes of education in Indonesia. The first question is ‘why are we making international comparisons at all; does Indonesia have to slavishly follow world ‘good practice’ or is there something unique and special to be encouraged through education in Indonesia?’ The study believes this is a discussion worth having and it needs to be a public discussion for two reasons. First, to recognise that in a modern, decentralized and democratic state, that it is essential to invite public discussion of key issues as part of the policy development process. The second reason is to assist in the process of socialising issues and contributing to the development of citizens who understand the processes used in the assessment of their children.

The second question assumes that making international comparisons is worthwhile. This question is ‘what can we learn when look at other countries that can help us frame our policies?’ It is impossible to embark on a comprehensive review of international developments here but the following are some useful indicators of what is happening.

### **4.1 The Philippines**

What happens in The Philippines is instructive for Indonesia as both countries share a common set of challenges in geography and in stage of development. The recommendations in Chapter 8 suggest regional cooperation to share common experiences and approaches.

The Philippines has decentralized education. Even though the curriculum is decided at national level, the implementation of the curriculum at school level can be modified and adjusted according the needs of students in the school. National assessments both in primary and secondary education only serve as a measure of school competencies and not as a basis for students’ admission for higher level of studies or for graduation. The primary school assessment given to Grade 6 students is the National Achievement Test, taken by both public and private schools students. The secondary school assessment is the National Secondary Achievement Test. At the national level, the National Educational Testing and Research Centre administer assessment.

The test that determines students’ graduation is held at primary school level and is developed by teachers. Assessment to determine students’ graduation from secondary school is also administered at school level. Teachers develop this assessment and to graduate students must score 75 percent in each subject area. This means that teachers and principals play a crucial role in student’s graduation.

### **4.2 China**

China’s recent experience with school examinations has many lessons for Indonesia. Once again, this is an Asian model of educational change with shared challenges in development, large populations, span of administrative control and geographical area.

---

<sup>2</sup> The kind assistance of Theresa Wikaningtyas, staff, DBE-3 and Lorna Power, Deputy Chief of Party, DBE-3, in preparing the original material for the sections on The Philippines and on China, respectively, is gratefully acknowledged.

The key lessons from China's recent experience lay in decentralization and in the abandonment of a highly centralized, high stakes examination system. Reforms introduced in 2001 have moved China away from a centralized, competitive, superficial and corrupt assessment system that shared characteristics with Indonesia's present high stakes and centralized system. The Chinese system was made worse by the fact that the stakes were pushed even higher as teachers' salaries were linked to students' scores. Moreover, students, teachers and schools were publicly ranked according to scores, compounding the stakes and creating intense competition.

The 2001 reforms were intended to move the system away from a centralized controlled and competitive system to a more relaxed atmosphere. The change introduced a new assessment of "general learning quality" and promoted changes in student assessment from measuring only knowledge and cognitive skills to all-round aspects of development. It was intended to reduce the competition and relieve students of the examination burden and subsequent stress on teachers, schools and families. The move was away from neglecting the diversity of student characteristics to taking these into account.

The following changes to the assessment system have been made. Broadly, examinations changed from being nationally unified to a provincial-unified public examination model. Examinations would be prepared and graded locally. A few of the changes introduced were:

- Public examinations at primary level were cancelled and school-based formative assessment was emphasized.
- Schools were encouraged to set up their own assessment objectives, taking into account the national curriculum standards and the conditions of schools.
- Ranking of schools, teachers and students was prohibited.
- More diverse testing was introduced – teacher observation, student demonstration, teacher–student interaction, student self evaluation and peer evaluation reduced the emphasis on objective, MCQ formats which themselves changed to focus more on problem solving, analyzing, and understanding.

Although China is still in a process of change there are already some positive results from the changes to assessment.

- Assessments are more relevant to the students
- Students are less stressed and many schools report that assessment has become more popular among students. Teachers report that, in general, students prefer the exams under the new reforms and actually are requesting more examinations as they enjoy them and find them useful.
- Cheating or corruption of the examinations in basic education is becoming less of a problem now the stakes for passing or failing the examination have been reduced.

China's experience suggests that change to assessment systems cannot be made overnight and will inevitably meet some obstacles. Lessons learned which could be useful if Indonesia starts to make some changes to the UN are:

- Changes to student assessment require a process of conceptual change for teachers and teachers need to be involved. "Developmental assessment" became a new but confusing concept for teachers and led to misunderstandings.
- Despite their severe limitations, public external examinations were still in most people's mind a fair, reliable and valid approach to assessment. This, and the misunderstanding of the new concept, initially caused conflicts and confusion in the minds of the public and educators alike. These outcomes reinforce the need for the full, transparent and participatory approaches to reform.

As time goes on schools are getting more autonomy and principals can now set their own exams but with guidance from Provinces. Many people are now worried that in reality this

may mean that if you have a good relationship with the head your child will pass!

### **4.3 Singapore**

Singapore, like South Korea, has demonstrated that it is possible to shift from a low to high quality education system within a few decades (McKinsey and Company, 2007). Singapore's ranking in the TIMSS (see Table 1 below) clearly shows this achievement. The common route to this achievement among improving national systems is focusing on teacher quality and professional responsibility. One component of teacher expertise and professionalism is to be able to assess student's work and, significantly, to use the information from a broad range of assessments to provide constructive feedback to students.<sup>3</sup> This focus on teachers is the foundation recommendation in this Report in Chapter 8.

In recent years, the Singapore Ministry of Education has made extensive changes in the areas of curriculum and assessment. The curriculum has been revised to promote critical and creative thinking skills, problem solving and independent learning abilities in students. To complement these changes in the curriculum, significant changes have been made in several areas of assessment that includes a much greater emphasis on assessing coursework and oral communication, and in the changes in item types including project work, science practicals, context-dependent test items, and writing.

### **4.4 Finland**

The Finnish National Board of Education is responsible for supporting school education and for developing the national core curriculum and assessment criteria. The schools decide, with their local authorities, how the core curriculum will be assessed, and they develop their own curriculum, within the national framework, taking account of the local context.

Teachers have considerable respect and trust as a profession in Finland. There is, therefore, a wide degree of responsibility given to teachers for the assessment of students, and flexibility in exercising that responsibility.

What relevance is Finland, a small, cold and high-latitudes European nation to Indonesia? Simply that Finland performs best on international tests of international student achievement and Indonesia performs near the bottom. There are lessons and directions to be learned from their experience. Like other high performing countries, there is greater teacher recognition and responsibility and a move away from centrally controlled multiple-choice testing towards more flexible and valid forms of assessment at school level.

Finland's experience also reinforces a lesson learned repeatedly in helping schools to change so successfully in Indonesia. That lesson is simple and powerful – respect and trust teachers, support them through the processes of change and professional development, strengthen school management, and then allow them to take responsibility for their professional actions. The experience of both Finland and the recent reforms in Indonesian schools demonstrate the power of bottom-up change compared to the poverty and weakness of top-down management and direction of which the UN is a prime example.

---

<sup>3</sup> In a research report that draws on over 500,000 studies of the influences on student achievement, feedback stands out as the single most powerful influence. See: John Hattie, *Teachers Make a Difference: What is the Research Evidence?* Melbourne, Australian Council of Educational Research, October 2003. Available: [www.acer.edu.au/workshops/documents/Teachers\\_Make\\_a\\_Difference\\_Hattie.pdf](http://www.acer.edu.au/workshops/documents/Teachers_Make_a_Difference_Hattie.pdf)

#### **4.5 International Studies of Educational Reform**

Four international studies of educational reform provide a solid basis for the recommendation developed in Chapter 8 as well as confirming Indonesian experience. The studies place great emphasis on the quality of teachers, an issue of considerable concern in Indonesia. One specific matter relates to teacher professional knowledge and skill in assessment and its place in education.

The first study makes a powerful case that if we seek real advances in student achievement and international competitiveness (which Indonesia is committed to in its Strategic Plan), then strategies must recognise and act on the clear finding from years of educational research that excellence in teaching is the single most powerful influence on student achievement (Hattie, 2003, 4).

A second international study to understand why the world's top-performing school systems perform so much better than others, is also relevant here because it again stresses teacher professionalism. The study concludes that three things matter most:

- Getting the best people to become teachers
- Developing them into effective teachers
- Ensuring the system is able to deliver the best possible teaching for every child.

This study, published by McKinsey & Company, 2007, *How the World's Best-performing School Systems Come Out on Top*, makes the following optimistic claim that reinforces the three conclusions listed above: "These [best performing school systems] demonstrate that best practices for achieving these three things work irrespective of the culture in which they are applied. They demonstrate that substantial improvement in outcomes is possible in a short period of time" (Executive Summary).

The same optimistic conclusion is made as in another recent international study, the International Project to Frame the Transformation of Schools: "... we believe, on the basis of evidence ... that all schools can be 'best schools' in the sense that all schools, regardless of circumstance, can be as good as those exceptional few that are currently considered to be the best" (Caldwell and Harris, 2008, 3). In both quotations above, key phrases are underlined to offer encouragement to weaker schools in Indonesia and to refute the inevitable objection that 'things are different here'.

Finally, a fourth study, a comparative review of assessment in Finland and Queensland, Australia, makes a very powerful observation in support of teacher professionalism; that teachers who work within an internal assessment system feel a sense of ownership and responsibility for the system (Queensland Studies Authority, 2008). This sense of ownership and responsibility is increasingly evident in Indonesia where teachers are responding very positively to the introduction of school based management and student-centred active learning, when supported to develop the necessary skills and knowledge. Similar support is indicated in the domain of assessing student learning where both theoretical and practical understanding is not strong and requires significant professional development. This need exists among educational administrators as well.

#### **4.6 International Comparisons of Student Performance**

The Government of Indonesia (GoI) has made strategic policy commitments to educational quality and international competitiveness in the national strategic plan for education, or *Renstra* (Ministry of National Education, 2005, 18).

Internationally, Indonesia performs poorly on educational quality indicators with respect to many countries and particularly with some of its close regional neighbours. Its uncompetitive position is reflected in the outcomes of international testing programs in which Indonesian students regularly participate, so there is a body of longitudinal comparative data available (Arze del Granado, et al., 2007, 26)

The Programme for International Student Assessment (PISA) assesses the reading, mathematics and scientific skills of 15-year-old students. In each three-yearly cycle of testing, one subject is chosen as a focus. In the 2006 test in science, Indonesia ranked 50/57 countries. The PISA report is not all bad news for Indonesia and it notes that Indonesia has seen significant rises in reading performance since PISA 2000 and that mathematics performance was 31 score points higher in PISA 2006 than in PISA 2003 (OECD, PISA 2006, 2007).

In 2007, Indonesia ranked 36 out of 49 countries in the Trends in International Mathematics Science Study (Institute of Educational Sciences, 2007). Close neighbours ranked as shown in Table 1 with comparatively higher achievements noted for Singapore and Malaysia.

**Table 1: International Comparison in the TIMSS, 2003 & 2007**

Country	2003 Score	2003 Rank	2007 Score	2007 Rank
<b>Indonesia</b>	<b>411</b>	<b>34/45</b>	<b>397</b>	<b>36/49</b>
Singapore	605	1/45	593	3/49
Malaysia	508	10/45	474	20/49
Thailand	na	na	441	29/49
Philippines	378	41/45	na	na
United States	504	15/45	508	9/49

#### **4.7 Discussion**

Given these comparatively poor performances, is not surprising that government planning attention has focused on international competitiveness and efforts directed at improving the quality and relevance of education.

The Government has introduced significant reforms in the education sector, which have created a policy environment that has led to important change and improvement (MoNE, 2005, 46). Decentralization delegated responsibility for basic education to District governments. At the same time, complementary laws and regulations to strengthen District and school level management and governance and to improve service delivery. These have been introduced with considerable success. Schools can determine their School Level Curriculum.

Yet, in this highly decentralized environment, there is a paradox. It is well understood that testing influences *what* students learn and the way they are tested determines *how* it is learned. The central government continues to exert enormous influence on learning and teaching through the UN in this supposedly decentralized environment, even where the

School Level Curriculum, is determined (in theory) by each school.

To consider the UN in these terms, it stands as a powerful, centralized, top-down, non-participative and opaque educational institution. It is a metaphor of past educational management in Indonesia, whatever merits can be attributed to centralized testing. It is odd to find a highly centralized assessment system when MoNE has decentralized many of its roles and functions and is more flexible in its orientation. What is especially odd is that the curriculum has been decentralized in a centralized examination system. This makes the alignment of assessment and curriculum unlikely. The Indonesian centralized system is also in contrast to changes made in both China and the Philippines.

There needs to be a serious discussion about the major shifts in educational authority that are transforming Indonesian education and education systems elsewhere. The clear shift in authority and resources away from the centre to schools and Districts in Indonesia is particularly strong. Changes in assessment policy are needed to ensure that it is consistent with this powerful shift. At the moment, assessment is a major blockage to full local responsibility and accountability.

Concurrently, there is a second flow of authority away from the Indonesian education system towards international values, institutions, and models. Some examples are the use of English as a language of instruction; the establishment of government-funded international standard schools; reference to student achievement in TIMSS and PISA tests; and the use of the International Baccalaureate.

This flow of authority towards international standards and values creates diversity in a system that concurrently create tensions with the prevailing, centralized assessment model as internationally-oriented stakeholders seek better quality and more relevant forms of assessment to reflect this diversity.

## 5 THEORETICAL AND TECHNICAL QUALITIES OF ASSESSMENT

### 5.1 Introduction

The study asserts that no progress in the development of a quality assessment system can be made without constant reference to four essential questions: can the system produce valid data, is this data reliable, is the system practical, and is it fair to all students?

A well-functioning examination system will address three standard, technical criteria. These technical criteria are:

- Validity
- Reliability
- Practicality (or utility).

These criteria are central in any consideration of a quality examination system. They are the foundation upon which all informed discussions of examinations must proceed (Chatterji, 2003, 53). The criteria refer ultimately to the quality of the *information* generated by the system, which is the purpose of the system in the first place. A system that does not generate valid, reliable and useful information will have a damaging effect on children's education. This damaging effect is because the process of assessing student learning in schools is the most powerful force driving student and teacher behaviour. Poor systems and poor outcomes will have a negative impact. Students, their teachers and parents understand fully that the examination result is 'high stakes' as decisions about transition to the next level of education, and possibly employment, rest almost completely on the information generated by the UN. The concept of high stakes testing is reviewed at the conclusion of this Chapter.

Poor quality information from the examination system will also distort the school system when that information is used in planning and assessing school quality.

In addition to the technical criteria, it is common to require examinations to meet an important fourth criterion of fairness. An unfair test yields invalid results. It is necessary to assume that all candidates have had an equal chance to learn the material regardless of the school or Province they come from; an assumption that many teachers in remote and rural areas of Indonesia strongly challenge. In addition, test items should be fair and not assume that all students will understand concepts and terms limited in their use to big cities or to Java, an issue complained about in Sulawesi. The review of examination papers in Chapter 6 suggests that fairness issues occurred in the 2007/2008 UN.

Each of the technical criteria will be briefly explained and linked to an analysis of the UN based on the evidence available to the study.

### 5.2 Validity

Validity is generally agreed to be the most important of the three criteria. Validity refers to the accuracy of the conclusions and interpretations we can make from the results or scores from an assessment. Validity is like truth. We can ask if the conclusions we make are truthful, based on the data that the assessment provides. The validity of conclusions and interpretations are adversely affected if the test items fail to match what it is the examination is meant to be assessing.

For example, the evidence in this Report demonstrates that we cannot make valid

conclusions from the data generated by the UN because (i) the test items do not always match what is meant to be assessed (that is, do match the curriculum objectives and content of the junior secondary school), (ii) the tests for Bahasa Indonesia, Mathematics, Science and English, being only 2 hours, do not provide sufficient time to test an adequate sample of objectives and content, (iii) the test papers include some poorly designed test items and a test format (MCQ) that cannot validly test certain valued skills such as the production of spoken Indonesian and English.

These shortcomings of the UN are known as *systematic errors*. Validity is enhanced when systematic errors are eliminated as far as possible from the examination process. The limit to removing known error is the practicality of doing so in a given set of circumstances.

Finally, there is sufficient evidence of such widespread distortion of the examination administration through malpractice that makes interpretations of scores invalid and meaningless. This last point is emphasised. The consistent feedback from stakeholders is that cheating is widespread. Some estimates are that *at least* 50% of candidates cheat or are influenced in some way by the cheating and corruption of others. Even if it is 10%, the following remains true: the scores obtained by those cheating are not a truthful measure of their academic achievement but a measure of their cheating ability. That is, the UN has no *construct validity* because the *construct* being measured is meant to be academic achievement in the subject, not cheating.

### 5.3 Reliability

Reliability refers to the consistency with which an examination measures a student's true knowledge and skills under different test conditions. Reliability is influenced by test design factors and by random factors in the examination environment. Both can be controlled, but not perfectly so.

Design factors can be controlled by:

- Taking an adequate sample of items, tasks, or behaviours from the student and from the same domain set out in the curriculum and related test specifications. As a rule, the greater the number of items, tasks, or behaviours the greater the quantitative estimates of reliability. One advantage of an objective-type, MCQ test is that it is theoretically possible to take that greater sample. In practice, the UN examination of a one two-hours paper cannot yield reliable results because it fails on sample size criteria and also cannot measure many aspects of skill and performance in the curriculum. The UN struggles to succeed when tested against these design criteria.

The matter of low reliability of the UN introduced by design considerations is compounded by environmental factors revealed by this study. Random environmental factors can be introduced by:

- The student. Taking multiple measures of a student's performance over time can improve the reliability of an assessment process by reducing the impact of random factors adversely affecting performance on assessment day (menstruation for girls, variations in mood and motivation, illness or injury, a family crisis, bad weather and flood creating a transport crisis, earthquake, and so on). This is a reason advanced in support of more continuous forms of assessment and school-assessed work.
- The examiner. Using machine scoring of answer sheets eliminates much of the concern about examiner bias and performance. Nevertheless, the study found examples where even the supposed reliability of the machine had been compromised by dirty and spoiled answer sheets and by interference with answer sheets between their completion by the student and scanning.

- The assessment process. Poor and confusing instructions and invigilation of examinations are recurring issues for students. Inappropriate behaviour of exam-room supervisors, the intrusion of monitoring teams and the delivery of incorrect papers or answer sheets are a few of the difficulties students have had to endure with the UN and which compromise the assessment process and reduce reliability.
- Security. Frequent instances of serious breaches of security were reported in the study. These breaches included opening of examination packages, copying test papers and selling them and, after the examination, teachers or invigilators 'correcting' students answer sheets before despatch.

## **5.4 Practicality**

Practicality is concerned with issues of administration, costs, scoring and the resources required for the whole examination process. Practical issues set the limits within which an examination system must operate. To build a well-functioning system, practical issues must be weighed against considerations of validity and reliability, recognising that any compromise to these two essential concepts will necessarily weaken the overall quality of the examinations. There are at least four major practicality concerns.

### **5.4.1 Administrative conditions**

The ease of administering a national examination is a major consideration as complications with administrative accountabilities, test instructions, time required, the availability of materials for the examination will lead relatively inexperienced managers and supervisors to make errors which will inevitably have a negative impact on the reliability of the examination.

The UN experiences serious administrative difficulties because of the number of different agencies involved in the whole process: BSNP (responsible for overall management and supervision), the Educational Assessment Centre (responsible for technical work), Provincial Dinas Pendidikan, printers, District Dinas, National Police (for security and avoidance of corruption), school principals, examination supervisors and the independent monitoring teams, all of which have a management and supervision functions. Two Ministries and their agencies are involved – MoNE and MoRA. This administrative complexity and lack of clear, unambiguous responsibilities and accountabilities is a major contributing factor to political interference, systematic corruption and cheating.

### **5.4.2 Administrative cost**

The paper-and-pencil MCQ tests used in the UN across Indonesia are, in principle, one of the least expensive forms of testing to administer. It does not require additional materials or equipment expenses as some tests do or the training and management of specialist markers, as electronic scanners perform this function.

### **5.4.3 Marking and scoring**

Machine scoring of answer sheets is undeniably a major practical advantage of the MCQ test format as the alternative – selecting, training and managing large numbers of competent examiners across Indonesia is, in practical terms, too complex and expensive to envisage for the large numbers of candidates involved. In principle, fast, efficient and accurate scoring should be the goal without sacrificing the validity and reliability of the examination. In reality, however, there is an inevitable compromise between these criteria.

### **5.4.4 Ease of use of test results**

Test results must have a high level of practical utility to all potential users of them. Information on what scores mean should be clear to all users and interpretive guidelines

made available with the results, something that is commonly missing in UN data sets. It is meaningless to say that a student has a score of 5.5 without placing that score at least within a context of norms and scales.

## **5.5 Discussion**

A well-designed examination system with high validity is the major technical requirement. This will generally lead to high reliability as well. The present UN arrangements, which seem to give highest priority to understandable practicality concerns, compromise the essential technical requirements for validity and reliability. The UN does this by placing emphasis on a relatively short, two hours, MCQ examination that cannot, in principle deliver valid and reliable results. Whatever standard of technical reliability and validity is achieved, however, is most seriously damaged at best or destroyed at worst by systematic malpractice in the examination.

## **5.6 High Stakes Testing**

### **5.6.1 *High Stakes Testing: a Definition***

The UN is described as a 'high stakes' test. 'High stakes' testing refers to tests where the results of a particular test will have a major impact on the candidate. Success leads to significant benefits for the candidate, including progression to the next level of education, scholarships, employment opportunities and recognition. Failure may be a barrier to educational progression, lead to grade repetition, diminished employment opportunities and stigmatisation as a 'failure'. It may also be a major contributing factor leading to early school leaving with profound long-term economic and social consequences for the candidate and his or her family.

The stakes are high for others involved in the UN as well. Teachers' reputations are influenced by test results. Similarly, there are implications for schools and principals, sometimes with financial implications for school quality development; for the Dinas Pendidikan and Kepala Dinas; political implications for the local government and the Bupati and at the Provincial level as well.

In this 'high stakes' environment, most examining systems have established and rigorously implemented elaborate security procedures to protect the integrity of the examination.

### **5.6.2 *The Experience of the United States of America***

Research reported by Nichols and Berliner (2005; 2007), provides strong evidence of a principle of social science known as 'Campbell's Law'. This Law states:

"The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (p. i).

These authors provide lengthy and substantial evidence that "... where high-stakes tests are used to assess students, teachers, or schools, the corruptions and distortions that inevitably appear compromise the validity of the test. As the stakes associated with a test go up, so does the uncertainty about the meaning of a score on the test. That is, in high-stakes testing environments, the greater the pressure to do well on the tests the more likely is the meaning of the score obtained by students or schools uninterpretable" (Nichols and Berliner, 2005, 5).

Applying this principle, the Nichols and Berliner study finds that the over-reliance on high-stakes testing has serious and negative repercussions that are present at every level of the public school system in the United States. A similar conclusion can be reached about Indonesian education as most of the same negative effects are found in the Indonesian UN.

These findings, which can be grouped into 10 categories, are based on an extensive review of news stories, traditional research studies and evidence provide by educators. The data presents an alarming picture of the corrupting effects of high-stakes testing.

**Table 2: The Corrupting Effect of High Stakes Testing (from Nichols and Berliner, 2005)**

	Effect Noted in USA from the Nichols and Berliner Study	Effect Noted in this Study from Fieldwork	Explanation
1	Administrator and Teacher Cheating	√	Corruption of test administration processes.
2	Student Cheating	√	Poll results reveal at 50% of students cheat at least once.
3	Exclusion of Low-Performance Students From Testing		Counselling students to leave school before the test or take equivalency tests so as not to spoil schools' or districts' reputations.
4	Misrepresentation of Student Dropouts		
5	Teaching to the Test	√	Eliminating subjects from the curriculum not being tested.
6	Narrowing the Curriculum	√	Focusing on narrow, mechanical and recall skills; drilling student at the expense of broader and creative curriculum.
7	Conflicting Accountability Ratings		Schools rated 'excellent' by a state failing to be selected into federal programs.
8	Questions about the Meaning of Proficiency	√	Raising benchmarks or lowering benchmarks to achieve politically driven goals.
9	Declining Teacher Morale	√	Teacher loss of sense of professionalism; in hands of administrators and students.
10	Score Reporting Errors	√	Organizations responsible for marking and grading failing to do so correctly.

### 5.6.3 Discussion

These authors conclude their study with these words: “High-stakes tests cannot be trusted – they are corrupted and distorted. To avoid exhaustive investigations into these tests that turn educators into police, this research supports building a new indicator system that is not subject to the distortions of high-stakes testing” (Nichols and Berliner, 2005, iv).

They argue that there are only two choices and these choices are presented here in relation

to the UN. The first choice is to design an assessment system less influenced by the known distortions of high-stakes testing, that is, to reduce the stakes and the incentive to distort and corrupt. That is an urgently needed strategy in the Indonesian education system. This choice builds on well-known strategies for addressing corruption documented in such sources as USAID's *A Handbook on Fighting Corruption* (1999).

In relation to this first choice, dealing with corruption also requires “... a basic framework of anticorruption laws, regulations and institutions in place that serve as the prerequisites or preconditions for all initiatives. As well, government officials and civil society, mass media, and business leaders must have the training, resources, and capacity to act effectively and with meaningful resolve over the long haul if anticorruption initiatives are to be adequately implemented” (Spector, et al, 2009, 2). Further analysis of the framework as it applies in the education sector is suggested.

The second choice is a continuation of the UN unchanged. Improving the grading system, the technical quality of the MCQ test items, technical strategies for identifying cheating in answer sheets (Departemen Pendidikan Nasional, 2009) or even introducing other forms of testing, essays perhaps, into the UN system will ultimately be useless and a waste of time and resources if the stakes remain so high.

As Nichols and Berliner stress, “...if we continue to monitor public education with high-stakes tests then we know full well that we should expect distortions and corruptions of the measurement system; we will be required to regularly investigate how such distortions and corruption occurs and determine who is responsible; and we will also be required to severely punish miscreants when we find them. This will make educators into police, a role they do not desire’ (Nichols and Berliner, 2005, 164).

Can there be any good practices in operating a high stakes testing system? Certainly, there are practices that can be implemented to reduce the worst excesses and negative impacts but this will require the creation of an ‘educational police force’. It is possible too, and necessary to improve the technical quality of the examination papers and administrative process.

But as long as the stakes remain high the incentive to corrupt the process will remain high. The primary source of corruption is not the children – who are the victims – but the way the system is designed and irresponsible adults and government officials at several levels of government. A twin strategy combining a reduction in the incentives to cheat and an increase in the incentives to be honest and professional is necessary. Other countries, as different as China and Finland, have achieved better systems of assessment by reducing the stakes.

## **6 EXAMINING THE EXAMINATION: THE QUALITY OF UN TEST PAPERS**

### **6.1 Introduction**

The test papers that students receive are a key component of the overall quality of a test system. Three estimates of test quality are presented here. First, an analysis undertaken by a team of experienced subject matter experts is presented in Table 3; second, an analysis of the 2008/2009 English test paper by the consultants to this project; and third, a summary of a published analysis of the 2004 English test paper by a local, expatriate English teacher.

### **6.2 Evaluation 1: Analysis of 2007/2008 Test Papers**

A small group evaluation of the UN was undertaken in October 2009. The evaluation was done by four subject matter experts, currently working with teachers in their respective curriculum areas and active in either a university or a development project appointment. Each expert was provided with a copy of the relevant test paper and a schedule containing 15 evaluative criteria drawn from the published literature on good practice in test items and overall test design. The experts were thoroughly briefed prior to working individually on their evaluations. The results of their work are presented in Table 3 on page 29. Although this evaluation strategy can rightly be criticised on the grounds of engaging only one examiner per subject, the results do show some cross-subject consistencies and the evaluation is similar in outcome to the other two evaluations presented here.

In summary, this evaluation shows that:

- Of a possible total of 60 '*agree/strongly agree*' evaluations against the stated criteria, the four test papers together scored only 38 (63%).
- The only criteria that examiners agreed were present in all papers related to the quality of the instructions provided (criteria 11 and 12).
- Only in Mathematics did the evaluator agree with the statement: "This test is a reasonable sample of all content areas of the *Kisi-kisi*".
- On 8 of the 15 criteria, evaluators agreed with two or less of the statements.
- The evaluator of Mathematics agreed with all statements (the best outcome) whereas the evaluator of Bahasa Indonesia agreed with less than half the statements – 6/15 or 40%. Science was 7/15 (47%) and English 10/15 (67%).

An outcome of this standard would be considered unacceptable if it were based on an evaluation of a school-prepared test paper. That this small group of four experienced examiners was unable to provide a 60/60 evaluation (100%) on a *national* examination is a matter of concern and an outcome that warrants further in-depth analysis by others as it has implications for the validity of data that can be generated by these test papers.

### **6.3 Evaluation 2: The 2008/2009 Bahasa Inggris Test Paper**

This 18 pages test paper comprises a title page, a page of directions (both in Bahasa Indonesia) and 16 pages of 50 multiple choice test items, each of four choices. Test item directions are in clear, simple English. All questions are of the context-dependent type. These require the student to read material and then answer a group of from two to four questions. The last four questions in the test differ from all others in that the student is

required to place jumbled words or phrases into a good sentence or paragraph. These last four items could claim to test the Bloom level of synthesis whereas all other test items appear to test comprehension. The test paper is presented clearly and is well laid out.

The directions on the test paper are silent on whether the student must choose one correct answer, the 'best' answer, or all alternatives that are correct/best. Even if this instruction is given on the answer sheet, it is important that it appear on the test paper as well so that the instruction is reinforced, is easily accessible for checking, and the student is not misled into making errors because of item construction faults. In this test paper at least seven questions (14%) have more than one correct option so faulty items is a very real issue for students.

Another (different) six questions are 'wrong' in the sense that they do not follow from the text upon which they are based or the item direction is wrong. For example, one question asks the meaning of the underlined word when in fact two words are underlined. Weaker students may be tricked by such an error. Such errors should not occur at all, however. One advantage of the MCQ format is the potential to test a much greater amount of material by asking more questions. This advantage has been wasted in this test paper by asking sometimes only two questions of rather long blocks of text. In technical terms, asking more questions rather than requiring students to do a comparatively large amount of reading for a few test items could have increased the reliability of this test.

Finally, there are ten questions presented in incorrect English. This occurs in either the context material, question stem or in the list of alternatives. This language error is even larger if the use of Indonesian English rather than Standard English is included. Appendix 4 provides examples of test item faults of this type.

In summary, 19 of the 50 questions, that is, 38% of all questions, have an identifiable error or fault in their construction. This penalises all students except those who are blindly guessing all responses and for whom such errors make no difference. It also introduces a large – and completely preventable – source of error into the scores for the UN for English.

#### **6.4 Evaluation 3: The 2003/2004 Bahasa Inggris Test Paper**

This third evaluation is by an expatriate English teacher. Using a different methodology and an earlier test paper, it reveals that problems identified with the 2003/2004 English paper persist to this day. The evaluator concludes: "The UAN is not very useful. It is not valid, authentic nor interactive and has negative impacts on learning. It is however, reasonably reliable and practical... All in all the UAN fails to be useful because of its test construction which is riddled with mistakes and contains many alternative multiple choice answers that are correct. Hughes (2003, p.2) claims, 'Students' true abilities are not always reflected in the test scores that they obtain.' This is the case with the UAN test" (Ogilvie, M., 2004, 8).

#### **6.5 Conclusion**

These evaluations indicate technical issues with test papers that raise questions about their design features and their capacity to yield information from which valid judgements could be made. The evaluations are of a limited sample of test papers and were undertaken by a small number of evaluators. The consistency of the issues reported in each evaluation suggests that the quality of test papers warrants further and more systematic attention.

**Table 3: An Evaluation of the 2007/2008 UN Test Papers**

Evaluation Criteria and Examples		Evaluation Statement	Evaluator's Strongly Agreeing or Agreeing with Statement				
			English	Math	Science	Bahasa Indonesia	Total Tests Meeting Criteria
1	A test should contain only items that test the subject being tested. Test items should not test knowledge outside the <i>Kisi-kisi</i> or unrelated concepts from other subjects.	Test items truly reflect the <i>Kisi-kisi</i> Soal UN.	√	√			2
2	A test should contain a reasonable sample of content from the <i>Kisi-kisi</i> .	This test is a reasonable sample of all content areas of the <i>Kisi-kisi</i> .		√			1
3	A good test should require students to use several different cognitive (mental) processes (e.g., evaluation, analysis, application).	This test is a reasonable sample of different cognitive levels in the <i>Kisi-kisi</i> .		√	√		2
4	The results from the test items should provide information about the student that leads to valid information about that student. For example, the ability to write should be tested by a writing task.	This test requires the student to demonstrate the skills / knowledge as set out in the <i>Kisi-kisi</i> (such as writing skill) and not something else (such as selecting an answer).		√	√		2
5	Language bias: The test items should be written in clear, simple language that students can be expected to comprehend.	This test is free of language bias.		√	√	√	3
6	Cultural biases: The test has items set in contexts that all students in Indonesia can be expected to understand. Examples where this would NOT be true are: (1) An English test containing the words "breakfast cereal" that remote/rural/poor students cannot be expected to know. (2) A Bahasa Indonesia test refers to 'mikrolet' that students outside big cities may never have heard of (3) The test has a bias towards Western cultural concepts, goods or values.	This test is free of cultural, social, religious or regional bias.		√		√	2

Evaluation Criteria and Examples		Evaluation Statement	Evaluator's Strongly Agreeing or Agreeing with Statement				
			English	Math	Science	Bahasa Indonesia	Total Tests Meeting Criteria
7	Gender bias: A test has items set in contexts that all children can be expected to understand. Examples where this would NOT be true are: (1) tests requiring context knowledge that favours boys' knowledge – such as football or cars, or (2) typical girls' knowledge such as cooking or sewing.	This test is free of gender bias.	√	√	√		3
8	Item distractors: A distractor should never be impossible / silly / irrelevant. For example, a Physics test item about forces should not have a distractor about Indonesian history.	The distractors in the MCQ test items are reasonable; they cannot be eliminated because they are irrelevant or silly.	√	√		√	3
9	Examiners should never attempt to mislead the candidate through the way the item is constructed.	This test is free of items that deliberately mislead or try to 'trick' the student.	√	√		√	3
10	There should be only one correct or 'best' answer to a question. Distractors can be determined to be wrong by a competent student or examiner.	The items in this test have only one correct or 'best' answer.	√	√			2
11	Instructions should be clear and consistent in all sections of the test. Instructions should not be presented in ways that confuse students.	The test instructions are clear and consistent in this exam paper.	√	√	√	√	4
12	Instructions should be free of errors. Instructions should never make incorrect references to section, page or item numbers.	The test instructions are correct in this exam paper.	√	√	√	√	4
13	A test should not present 'barriers' to succeeding. For example, a science test should not use complicated high-level language in ways that makes it a reading or comprehension test rather than a test of science.	This test is constructed in a way that it presents no barriers to responding.	√	√			2
14	Test should not be 'speed tests' - the time required for most students to complete the test should be reasonable.	The number of tasks / items to be completed in the time is reasonable.	√	√			2

Evaluation Criteria and Examples		Evaluation Statement	Evaluator's Strongly Agreeing or Agreeing with Statement				
			English	Math	Science	Bahasa Indonesia	Total Tests Meeting Criteria
15	Tests should be of reasonable physical quality: the test should be printed on reasonable quality paper, in large-enough type for students to read, be professionally laid out and presented, and should not contain errors and mistakes of any kind.	This test paper is of high technical quality.	√	√	√		3
	<b>Totals</b>		<b>10</b>	<b>15</b>	<b>7</b>	<b>6</b>	<b>38</b>



## **7 THE CURRENT DEBATES**

### **7.1 Content and Construction of the UN**

The content and construction of the UN has been discussed in detail in section 3.2. It was shown that the UN is not aligned with content taught in each school if schools follow the School Level Curriculum. There is evidence that the School Level Curriculum principles are not reflected in the UN. Neither does the UN accurately measure what is taught in the curriculum because the chosen format, MCQ, does not allow the testing of attitudes and skills.

Hidden in the complex processes of test construction is a confusion of approaches to assessment that need to be clarified. On the one hand, the UN seems to be constructed on a norm-referenced assessment model designed to assess student achievement on a numerical scale with pre-determined pass levels, but a careful reading of the National Standards of Education can leave little doubt that the intention is to assess students against *competencies* using a standards model. Exactly which model informs the UN is unclear.

The complexity of the UN construction processes, together with the complexity and number of laws, regulations and standards governing it, poses a risk to the overall quality of the UN. Unnecessarily complex systems such as this almost always lead to administrative failure of some kind and provide opportunities for the corrupt and dishonest to exploit them. Simplicity alone is not sufficient. Even simple strategies must be based on good evidence that the strategy will work. At the moment, the UN seems to reflect a strategy based on weak evidence, the maximum number of entities involved in its design and administration, too many complicated regulations, and confused fundamental assumptions.

### **7.2 The Usefulness of the UN**

Notwithstanding its weaknesses, the UN is seen by stakeholders to perform an important function in certifying student completion of nine years basic education. It is seen as the primary means by which entry is determined to senior secondary education for those who wish to continue their school education. In the context of the UN being the 'gold standard' for determining entry to senior secondary school, there is nevertheless a strong body of opinion that is convinced that it should not be the sole criterion and that school assessment of students should also be taken into consideration. This is, in fact, the formal requirement (Appendix 2, Chapter 4, Part V, Passing Marks) but seems to be ignored in many cases as passing the school exams is reduced to a simple threshold requirement only.

The other major use of UN results is as an important input into the development of the quality of schools. School data is organised and published by the Ministry and then used in planning assistance to schools, usually training for the principal and teachers, and in evaluating the performance of the principal. There is, therefore, an incentive for principals to ensure their students perform well in the UN that, in turn, increases the stakes, leading to more malpractice.

### **7.3 Administration of the UN**

The administration of the UN is a massive and complex task. Examination papers have to be prepared; printing tenders prepared and evaluated at Provincial level and printing undertaken under secure arrangements; papers distributed to schools through the Provincial and District

educational administrations; security arranged through the National Police; test supervision arranged in schools; independent monitoring teams arranged and activities monitored; answer sheets collected; secured; scanned and results forwarded to Jakarta for processing from Provinces and finally, the analysis, grading and publication of results.

Schedules have to be meticulously planned to ensure security and timely implementation. In some Districts the consultancy was reliably informed by the District Dinas Pendidikan that test papers must be distributed to remote areas many days in advance of the test dates. All of this occurs within in 33 Provinces, 465 Districts, a total of 39,160 public and private schools administered by two government departments and in 2007/2008 for 2,518,935 candidates, of whom 93% graduated.

The overall administration cost of the UN from APBN alone is estimated by to be about Rp260 miliar or about USD28 million. To this, must be added the significant Provincial, District, and school costs.

This administrative process can be evaluated in terms of the question 'is the glass half full or is the glass half empty?' – in other words, in terms of achievements and failures. Looked at positively, it is a major implementation achievement that the UN runs as well as it does in such a complex environment. Teachers and principals from several schools commented favourably on the smooth running of the UN and did not describe any negative experiences with it at all.

Looked at negatively as in 'is the glass is half empty?' some stakeholders reported cases of significant malpractice and even political interference during the long chain of administrative processes. It is reported to occur among politicians, government stakeholders from senior officials down to classroom teachers and to students cheating. It permeates all the school support structures including the school clusters, teacher and principal's working groups (MGMP, KKG, KKS) and even the 'independent' monitoring teams.

Strategies for cheating are carefully planned by adults to get answers to students before exams, during the exams and to change answer sheets after exams. The examination system operates more like a free market, where results can be bought and sold, than a fair and respected professional process. In some cases, students are so confident of receiving answers that the incentive to do any study at all is removed. Playing video games becomes the substitute for school attendance.

It is, of course, very difficult to corroborate all of these claims. The consultancy noted, however, the significance of the regularity with which stakeholders quite openly and freely reported cheating as though it was a normal and expected feature of the whole examination process. Malpractice and cheating are concerns routinely picked up by the media as documented in Appendix 3. Indonesia Corruption Watch estimates the problem is worse outside Jakarta and worse among weaker, private schools where the incentive to inappropriately boost the apparent success of the weaker school is greater. It is the weaker private schools that enrol more poor students so the costs of cheating are borne disproportionately by poorer parents.

Given the widespread regularity of reports of cheating, the consultancy concludes that there is an issue of major concern here to be addressed. Although by no means universal, it is so widespread as to invalidate the results of the UN. A corrupt system cannot yield valid and reliable data, so an examination system corrupted by systemic cheating has no integrity. A reflection of the almost random nature of scores was revealed by teachers who, when asked about malpractice and cheating admitted their dismay at discovering that known good students had failed whereas weak student had passed. Even the very best technical system cannot yield meaningful data when it is corrupted at so many stages of the examination cycle.

There is a very sad sense of helplessness among honest teachers and principals about this as well. They struggle to uphold moral standards in a corrupt system and earn the displeasure of students and others for their professional integrity. Accounts of teachers giving students the answer sheets, checking and correcting completed answer sheets (and even getting answers wrong thereby lowering a student's score) are commonplace. In one case, when exploring the question of student examination anxiety, one group of teachers openly admitted this was not a problem at all as their students would have the answer sheets anyway, so why worry? Of course, if this is happening, it completely destroys the usual argument that exams are worthwhile because they motivate students to study hard and teachers to teach well.

How this systematic corruption and cheating can be dealt with is generally greeted with silence. It clearly requires attention by all levels of government and agencies involved in the UN. This attention requires evidence-based strategies. However, three global strategies based on proven educational principles can be put into immediate effect to mitigate the worst excesses. They are:

- Reduce the incentive to cheat by reducing the 'stakes' involved in testing. The experience of China demonstrates that this strategy can work and have an effect in the short term.
- Continue to work hard to improve the quality of entrants to the teaching workforce, improve the quality of their formal professional education, their working conditions and their continuing professional development. Concurrently, continue with the improvement of school quality, which almost always will address a wide range of school operational issues such as this. The evidence from the best performing school systems in the world is very clear about the impact of this strategy (McKinsey and Company, 2007; Caldwell and Harris, 2008).
- Consider extending the role of the Dewan Pendidikan, school committees, parents and the community in their monitoring role into the processes of examination supervision; building on their demonstrated success in improving transparency and accountability in school management and finance. However, this strategy will have to be managed with great care if community members have children as candidates in the current UN.

Cheating in examinations has a very long history that dates back to the first tests. Scholars note that cheating was prevalent in the Chinese civil service exams thousands of years ago, even when cheating carried the penalty of death for both examinee and examiner (Wikipedia, 2009). There are two lessons from this; first, Indonesia's problems are not unique and it is common in education systems; second, even the ultimate disincentive for cheating (death) may not work! This suggests that punitive control measures may not be effective.

There is an extensive literature on student cheating and what can be done about it, but in Indonesia this is only one part of a very serious problem that clearly permeates the education workforce, various levels of government, security systems including the police, and printing contractors. Very briefly, US research has shown that there is a strong correlation between forms of academic integrity systems and levels of cheating in schools. Studies have found students who attend schools with honour codes are less likely to cheat. For example, one study found that only 28% of schools with honour codes have high levels of cheating, whereas 81% of schools with judicial boards and punishment have higher rates of cheating (Wikipedia, Academic Dishonesty, 2009).

Administrative failures in examination implementation are also reported from schools. These failures include the late arrival of guidelines from Jakarta, distribution of wrong packages of tests to schools, insufficient test papers or answer sheets – a devastating issue for those candidates affected by this administrative error – and significant disturbance to candidates under examination by the insensitive intrusion of 'independent monitors'. A further failure is

that processes sometimes fail to detect incorrect entry by students of their biodata and the receipt of 'dirty' answer sheets that cannot subsequently be scanned. Finally, ICW noted that late payments from government for invigilation and monitoring service routinely leads to malpractice when people involved seek payments from students and their parents.

Issues emerged with the way the printing of the examination papers was poorly managed. The change from the previous system of direct purchase of printing service to competitive bidding at Provincial level has created a situation where the quality of work declined and security breaches increased.

The current *Renstra* of the Ministry of National Education has a strong conceptual framework for addressing ethical issues. There is very explicit statement of moral values and a statement of policy: "The order is not only the basic protocol but also the direction of conduct for all employees in carrying out their daily activities' (MoNE, 2005, 14). The framework presented in *Renstra* comprises three sets of moral values: input values, including professional and responsible behaviour; process values such as being accountable, giving a good example and respecting the rules; and output values including service excellence, reliability and fairness. What is missing from this commitment is an explicit strategy for implementation and monitoring of these values. The policy is not aligned with any implementation procedure and Chapter 8 addresses this in the recommendations concerning integrity.

## **7.4 The Impact of the UN**

### **7.4.1 *Impact on schools, teachers and principals***

Low scores reflect badly on schools and so there is pressure on principals to avoid this and to achieve a good ranking for their school in District and sub District 'league tables'. A substantial part of the pressure comes from above where governors, bupatis, and wali kota have been reported to seek the 'best' outcomes for schools in their jurisdiction through suspect means.

Low student scores reflect badly on teachers who may be blamed for this outcome. Teachers say they become 'scared', anxious and stressed in this situation, especially when they share their colleagues understanding that there can be no correlation at all between test scores and their teaching or their students' abilities and diligence, unless they make corrupt interventions to improve their student's chances of success. Some teachers wish to be able to hold a strong moral position against the corruption of the process and the cheating but realize that in doing so they may hurt their students' chances of 'success' in a corrupted system. So, whatever the strength of their moral or professional commitment, they feel forced to cheat. They are unable to demonstrate the responsibility and accountability they believe their profession demands of them.

There is also confusion in the minds of some teachers. They report confusion over what will be examined. They also report that their anxieties are increased because of the variable quality of textbooks and, among the books that are available, about the relevance and quality of their content in relation to the content to be examined. Complaints from the Provinces about a Java-centric view of the world in texts have been made. As many teachers base their lesson planning on the texts, rather than the official curriculum, what students actually get from their teachers may be inappropriate for the UN. These matters, they suggest, could be addressed if there was more teacher professional development in curriculum content rather than only in teaching methods and school administration. It also demands better planning from above.

The UN has an impact on teachers' approach to their teaching. Their methods involve a much heavier emphasis on drilling students on facts without attempting to develop deeper levels of

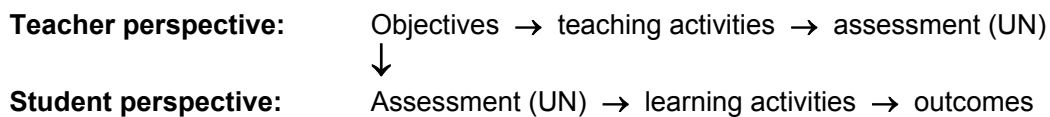
understanding, and a focus on content likely to be tested to the exclusion of other material. These approaches are intensified as the UN gets closer with special tutorial classes added to the day's working program. For some, the tension between their recently mastered approaches to student-centred, active learning and the pressure from the UN to behave in exactly the opposite way is significant and very frustrating.

Most of the pressures outlined here are intensified for private schools. Depending on their reputations to attract future students, they too must do all they can to maintain or enhance their reputations. This simply reinforces all other pressures to cheat. This leads to the creation of 'Tim Sukses' strategies that have a large measure of corrupt practices in them including improperly locating and using test papers and answer sheets before the UN, during the actual UN, and other strategies to 'edit' students' answers after the examination.

#### **7.4.2 Impact on students**

Students clearly appreciate the significance of the UN in determining their futures. Optimistically, they report that it can have beneficial effects by setting standards to be achieved and to focus classroom activities that make both students and teachers more diligent. They have very serious reservations about the UN as well.

Any assessment system will have either a positive or negative impact on students. Educators have realised that the 'backwash' of assessment on learning, almost always seen as negative, may also be positive if good quality assessment is implemented. The essential principle for good quality assessment is that it is aligned with the curriculum. Biggs (1999) explains that teachers and students see the curriculum process very differently. The teacher sees assessment (the UN) as the end of the learning-teaching process where students see it as the beginning. The following diagram shows this difference.



If the teacher's objectives and teaching activities are reflected in the assessment, as indicated by the downward arrow, then the teacher's teaching and student's learning are directed to the same overall goal. In this case, backwash works to support the learning competencies specified in the curriculum.

In the UN, because the specified curriculum competency standards are not well reflected in the UN, the teacher who teaches to help students achieve the required curriculum competencies is working against the student's interests. These interests are to be able to respond, and in some cases, successfully cheat their way through a two-hour multiple-choice examination of knowledge, rather than the specified curriculum competencies. In this case, the objectives, learning and teaching and assessment are not aligned and the backwash is negative.

Students express a number of serious anxieties created by the UN. The seriousness of these anxieties is illustrated by reports from students interviewed of students known to them who have suffered mental illness from the approaching UN and, tragically, one case of reported suicide. Given the large population of UN candidates, the actual numbers of such cases will be much higher than this.

Sources of anxiety include the following: fear of failure and its consequences; slowly

'improving' standards dictated by central government from year to year that students feel they may not reach; the inclusion in the UN of material learned in Year 7 and Year 8 that may have been forgotten; the faking of the answer sheets that can be bought; anxieties from the discrimination against poor students and their families who cannot afford to buy answer sheets; the pressure of being tested in two hours on three year's work; making mistakes in filling in unfamiliar answer sheets and biodata; the extra study load of special preparation classes at school, outside tuition, and parental nervousness and pressure on them at home; the threat that their completed UN answer sheets may be lost or tampered with.

Their anxieties are, of course, picked up by the more sensitive of their teachers who report that many students find the UN 'frightening'. The less sensitive may not even be in the school as a study by Usman, Akhmadi, and Suryadarma discovered in 2004. They found teacher absence rates in Indonesia higher than comparable countries and that it is related to poverty and the location of schools. Teacher absenteeism, not surprisingly, is also negatively correlated with the performance of students. In other words, teachers who are absent compound the assessment issues, particularly for children already experiencing difficulties through poverty and remoteness.

The fundamental issue with so many of these anxieties is that they are of a kind that the students have no control over, no matter how diligent they are, or how much they pay for the best tuition or for the most reliable answer sheets available on the 'market'.

#### **7.4.3 Impact on student learning**

An observation made repeatedly in the professional literature is that high stakes examinations, such as the UN, are the most powerful influence on student learning; much more than the curriculum and more than the teaching approaches that are used – "what gets tested is what gets learnt and how it is tested determines how it is learnt" (McKinsey, 2007, 36).

One respondent in the group interviews succinctly summarized the answer to this issue by saying: "Successful cheating is an important life skill in Indonesia." To which another added: "The other life skill is to avoid being caught." These reflect the management truism "what gets rewarded, gets done". Apart from focusing student attention on the goal of passing the UN, it is very difficult to identify important life skills that the UN does assess. Certainly, life does not require an ability to answer multiple-choice questions unless one seeks a life as a professional student!

One of the most common arguments advanced in support of the UN in all groups was that it acts as a powerful motivator for student learning. Teachers frequently noted in discussions that students 'enjoy' the UN as it gives them a focus and clear structure for their classroom activities, which is a poor reflection on the quality of teaching preceding the UN. Some students are able to take advantage of their place in one-roof schools to benefit directly from the experience of students who preceded them and others rely on working together with their friends on shared examination challenges. Parents and students report that the UN year can bring them closer together as they work on achieving good results. Others, of course, feel excessive pressure from anxious parents who may be constantly nagging them to study hard.

For years educators have railed against this 'anti-educational' view of assessment although there is growing acceptance of its central truth and inevitability, and a realisation that the motivating effect of assessment can be put to positive educational use as Biggs explained above.

However, the currently implemented UN is teaching children a range of values that truly professional and moral educators would deplore. According to the true 'experts' on this matter,

the students themselves, among the negative things that students are learning are:

- How to cheat; success in life depends on being a successful cheat
- No matter how hard you work or how little you work, your result may be truly random
- There is no clear reward in the school system for honesty and integrity, either for students, or for their teachers, principals, and others in the system who have professional and moral responsibility for them
- Successful learning, and the kind of academic learning that Indonesian society most values is how to answer multiple choice examination questions
- Not to take responsibility. Failure in the UN is the fault of others, especially teachers and the system
- Lack of trust. This is developed very effectively through
  - their experience of working with extremely poor quality test items that lead them to confusion or wrong answers,
  - learning that some answer sheets for sale are fakes (in other words, the corruptors are corrupting a corrupt system)
  - through the suspicion that teachers who may change their UN answer sheets before they are submitted for scanning or who may send answers by sms to candidates during the examination will get answers wrong and place the student in a worse situation

The experiences of Year 10 senior high school students who completed the most recent UN are especially revealing. The only positive thing they had to say about the impact of the UN was that it focused their study and they worked more seriously than in Year 7 and 8. However, there is considerable doubt about what they were learning and how they were going about this, given the distorting effects of high stakes testing on learning strategies. Their further comments reflect a number of other key issues as follows:

**The quality of the UN.** Students assert the UN is ineffective in what it does. The quality of test papers was poor, contained material not in the curriculum, and lower quality than the semester tests prepared by their school. It is of concern that better quality school exams were noted in several locations.

**Cheating and corruption.** The whole experience was a test of skills in cheating. The UN is of no use as most students get access to the answer key. However, this can be expensive; questions and answers were available the night before for Rp 5,000,000. Wrong answer keys are circulated which contribute to failure and to further anxiety. During the actual test, answers are circulated by teachers via sms or by dropped notes. Mobile phones were allowed into the examination room. Students say teachers can be easily bribed to support the cheating process. The whole process is ruining the morals of children and, they claim, has led to one known death and stresses students to the point they cannot function effectively.

**Extra tuition.** Outside classes are available but these can be expensive at Rp 2,000,000 for a year's tuition. Schools also conduct additional tuition classes but students complained that teachers do not often turn up for these classes, as they get no additional payment.

#### **7.4.4 Impact on parents**

As a group, parents value the UN and see a national certificate as something to be valued over a local or school award that may not have much value, particularly in other parts of Indonesia. On the other hand, they contradict this position by telling you that the common national standard is unfair particularly when their children attend poor quality schools in disadvantaged locations where education is often under-valued.

The UN brings students, parents and schools together. This is strengthened when the school

provides opportunities for parents to learn how they can support their children during the examination year. Parents believe that the UN acts as a good motivator for student study.

Outside tuition classes, resources in district libraries, and the availability of for-sale tests and answer sheets, places a financial burden on parents. Some buy additional books and computers to support their children. Together with the availability of for-sale answer sheets, parents recognise that there is an examination 'industry', some of it legitimate and some illegitimate, built around the UN. This industry seems to include significant numbers of 'moonlighting' teachers and other education sector professionals.

The stresses of this additional financial burden are felt particularly among the poor who may not be able to pay for the extra help to support their children. Without this help, the children of poor parents, who are said to need more help, are placed at an even greater disadvantage compared to the children of wealthier parents. In addition to financial stresses, parents naturally worry about their children and the results they will achieve. Many will buy the answer sheets but worry further that they may be not helping at all because of the circulation of fakes.

### **7.5 A Missing Link – High Expectations?**

What is absent is a fundamental principle of good quality education. This principle is to set high expectations for what students should achieve. The McKinsey review of best performing school systems stresses, "all of the top performing and rapidly improving systems have curriculum standards which set clear and high expectations for what students should achieve" (McKinsey, 2007, 35).

In Indonesia, the grading scheme (see section 3.3) sends a confused signal. A higher expectation is communicated from one year to the next as shown, however the basic pass-fail principle remains. Pass-fail schemes are understood to depress overall levels of achievement as there is no reward for the really hard-working and competent to receive better than a pass (Mehrens and Lehmann, 1973). A move away from this scheme towards a graded scheme (A,B,C,D,F) or scores referenced to a curriculum standard would help to address this matter.

Apart from the above, the strongly influential UN is sending powerful signals of both low academic standards and low moral standards. These standards are communicated through the corrupt behaviour of some adults managing the UN on the one hand, and by the low standard of item construction and errors in the test papers themselves. Students learn to expect that error is normal and acceptable in official documents such as the UN, and that the idea that 'near enough is good enough' is acceptable in the Indonesian education system. Appendix 4 presents some examples of this issue.

### **7.6 Objectivity of the UN**

There are three ways in which objectivity can be analysed. The first way relates to the nature of the tests used. As the UN relies entirely on objective-type test formats, specifically multiple-choice items, it is technically described as an 'objective' test.

The second refers to the way in which the test is constructed. Based as it is upon the framework of the kisi kisi, the construction of the test can be described as 'objective', that is, not determined by the attitudes or preferences of the test-constructors but by reference to a formal, officially approved framework. Thirdly, the use of optically scanned answer sheets removes the subjectivity of human test markers and so here also the UN can be described as 'objective'.

However, this account fails to take account of the potential and actual distortions known to exist in the examination process. For example, the outcome for a student that has been provided with a test paper and answer sheet prior to examination day cannot be reasonably described as 'objective'.

A most revealing indicator of lack of objectivity was agreement among several teachers interviewed that there was no correlation between their pre-examination estimation of a student's results and the actual outcome. They say that often very good students fail and very poor students pass. Moreover, Depag complains that year-to-year variation in scores is too high to be believable. This variability occurs both between schools and within regions. Therefore using test results for planning purposes is unreliable at best and highly misleading at worst.

### **7.7 Cost Effectiveness**

The study has been asked to report on the cost effectiveness – the most cost effective way to assess. Cost effectiveness analysis is an evaluation procedure that assists in choosing among alternative courses of action based on a consideration of the costs of the alternatives (Education Encyclopedia-StateUniversity.com, 2009). All educational decisions face budgetary and resource constraints so decision makers require estimates of cost in addition to estimates of educational or administrative benefits.

Cost effectiveness is determined with a clear goal and a set of alternatives for reaching the goal. This necessitates measuring, or estimating, both the effectiveness of the alternatives and the costs of those alternatives. Such analysis is impossible in the present circumstances for four reasons:

- It is not possible, at the present time, to construct a set of clear alternatives as the goals are by no means settled and remain in a state of flux (see Ministerial Statement).
- Clear alternatives to the present UN which can meet agreed goals need to be constructed before their potential costs and effectiveness can be estimated
- Attempts to identify costs in this study failed because the necessary financial data were not readily available. The analysis is exceedingly complex as costs are borne at the three different levels of government and in schools. Disaggregated data is not readily available or available at all.
- Finally, having established here that the present UN is technically and administratively ineffective, it cannot, by definition be 'cost effective' even if all costs were known.

### **7.8 The Need for a National Examination**

Stakeholder groups consulted were strongly of the opinion that there is a continuing need for a national examination and very few suggestions were offered about any viable alternative. The reasons most commonly advanced for the retention of the UN, in some form at least, were that it provided evidence of the completion of nine years basic education and motivated both students and teachers to work more diligently. A common response reflects the weak foundation of professional knowledge of assessment noted in Chapter 8. That response is that we value a national examination but, on the other hand, we also want it decentralized to at least Provincial level to account for local needs and characteristics.

The consultants believe that a good quality examination can serve a useful purpose in the education process and notes that many good quality education systems have such an examination. However, there are two questions that deserve to be publicly discussed.

The first question is ‘should the UN be awarded on an the basis of an external examination only as at present, or should the award contain a school-assessed (but moderated) component?’”

The second question is ‘ what is the merit of continuing with the present UN when its technical shortcomings and corrupted processes invalidate any reasonable conclusion that can be drawn from the data it provides?’

## **7.9 Forces For and Against Reform**

The use of the term ‘forces’ suggests there may be powerful groups or individuals championing the case either for or against reform. The study was not able to clearly discern such forces but accepts they may exist and have not been identified in this exploratory study. Certainly there is recognition of the need for reforms from the new Minister and among Ministry officials and no sense of the UN being ‘off-limits’ to donor assistance was detected.

Rather there seems to be disorganised and somewhat weak pattern of forces that would like to see constructive change but seem bowed by the sheer size, difficulty and complexity of the task. One of the most powerful lobbies for change is the media. It is clear from a brief analysis of the newspaper media that they would like to see change, but again it is not in any way systemic or a ‘campaign’ but rather a seasonal pattern of stories of frustration that reflect the present year’s UN. A very small sample of such stories is identified in Appendix 3.

Another organized group seeking change is non-government organisations of which the Indonesia Corruption Watch (ICW) is a good example. Two groups, the Advocacy Team for National Exam Victims and the Education Forum have recently (14 September, 2009; but made public on 25 November 2009) achieved notoriety by succeeding in a legal challenge to the conduct of the UN and having the Supreme Court of Indonesia requiring the government to revoke the UN. The implications and outcomes of the Court’s ruling have yet to be determined. Appendix 3 cites the news reports confirming this decision.

Possibly the greatest resistance to reform will come from the ‘rent seekers’, meaning those in the system who are corruptly extracting personal value from that system without making any ethical or professional contribution themselves. Where these people are organised and structured at local level in the gugus and KKS, for example, it is likely that reformers will face resistance there.

Rather than only looking for forces for and against change in terms of people, groups and organizations, an additional approach is to use the economic constructs of supply and demand to structure a consideration of the forces for change.

Significant change will be facilitated if there is demand for it and at the moment there seems to be little widespread, coordinated demand for change. Certainly, the new Minister has called for improvements in the quality of the UN (Jakarta Post, 7 November 2009) and his call deserves all the support it can get by building demand in the profession and in the community. Equally, the supply of expert knowledge and technology from donor or government sources or in some kind of contracting arrangement with a specialist assessment organization must be in place to respond effectively to the demand.

## **8 CONCLUSION AND RECOMMENDATIONS**

### **8.1 General Conclusion**

The study concludes that there is sufficient evidence from this exploration of the UN to confirm that the UN is a high stakes examination where the pressures on students and on schools to perform well contribute to an intolerable level of examination malpractice, invalid data outcomes and negative backwash effects on the quality of learning and teaching in junior secondary schools. There is also evidence, presented throughout this Report, of serious deficiencies in the educational and technical preparation of the UN.

Because of these findings, the study makes a series of recommendations designed to reduce distortions and to increase the quality of the beneficial impacts the UN may have on education. The key issue to be addressed, according to both senior Ministry officials and school-level stakeholders is to enhance the credibility of the UN and its annual implementation. The credibility of the UN is eroded by four factors. The four factors are interlinked and mutually reinforcing. For example, Factor 1 is the foundation for all other Factors and Factor 4 and 2 clearly contribute to each other.

#### **Factor 1: A weak foundation of professional knowledge**

The first factor is a weak foundation of professional knowledge about student assessment across the education workforce. This weakness is directly related to widespread concerns about teacher quality and the study notes that government strategies are in place to address this. The weakness in professional knowledge particularly extends to those educational administrators who do not have a professional background in education and whose knowledge of assessment seems to be limited to their experiences as school students. The wider community is similarly not well informed about assessment and this matter warrants attention in change strategies as well.

#### **Factor 2: Poor professional and ethical standards**

The second factor is the poor standard of professional and ethical behaviour among many who have children in their care. There are many dedicated and truly professional people in the education workforce and this criticism is not of them. Rather, it is of those who participate in the widespread, systemic and entrenched corruption of the administrative processes in the implementation of the UN and of those who support and encourage student cheating.

#### **Factor 3: Educational and technical weaknesses in design**

The third factor is educational and technical weaknesses in design. This factor includes the weak alignment of the examination system with the curriculum and the distortions in the educational process created by the UN. It includes the technical design of the examination, test items and implementation procedures of the UN.

#### **Factor 4: Unacceptable educational risk from high stakes testing**

Finally, in this professionally and ethically weak environment, the high stakes UN increases the scale and distribution of the poor professional and ethical standards issues and decreases the educational and administrative utility of the UN, putting at risk the futures of young people.

The recommendations are based on an understanding of these four factors. These factors need to be addressed holistically, systematically and in the context of Indonesia's reforming and improving educational system. By recognising the power of assessment, addressing it properly will add impetus to reform by driving improvements deeper and further.

This section proceeds on the assumption that a UN should continue at least in the short-term but in a substantially improved form. There is clear evidence that the present UN needs to be improved but little evidence of demand for its abolition at this stage. This may, however, be due to the fact that no alternative has yet been offered to schools. The new Minister of National Education has recognised the reform needs in his 100-days plan (see Appendix 3).

## **8.2 Recommendations for Reform**

The evidence located for this study makes the case for significant reform of the UN. The study recognises that there is no clear impetus for abandoning the UN. The approach here is to accept that the UN will continue for the foreseeable future but that substantial change is urgent to address the credibility of the award and to mitigate negative impacts on the system as a whole, on children and on Indonesia's aspirations to improve its international competitiveness through educational excellence.

The recommendations, while focused primarily on the UN, are also intended to develop the quality of school-based assessment, the teachers, and the educational administrators working in assessment-related areas of the education system.

The recommendations include specific strategies that can be implemented collaboratively by donors and government agencies. Consideration of the recommendations can take place within a continuing policy dialogue with the government on improving the quality of education. For USAID, engaging in such a dialogue, especially with BSNP, the Educational Assessment Centre and the Directorate General of Primary and Secondary Education, Ministry of National Education, is a recommended strategy.

What would an improved UN system look like? Improvement requires the construction of a set of key targets for an improved, simplified UN and proposals on strategies to achieve those targets. The key targets for improvement proposed here have been identified among the factors listed in the conclusion above and are summarized as follows:

**Target 1: A professional, integrated and aligned UN system**

Addresses: Factor 1 and Factor 3.

**Target 2: An ethical education system**

Addresses: Factor 2.

**Target 3: A better quality, credible and more flexible credential at the end of Year 9**

Addresses: Factor 3 and Factor 4.

### **8.2.1 *Target 1: Strategies for a Professional, integrated and Aligned UN System***

**Recommendation 1: Increase and sustain all efforts to improve the quality of the teaching workforce specifically and improve basic education generally**

**Potential for donor assistance:** High. Strategies for sustaining changes achieved in donor assisted basic education development projects is warranted as is building further on achievements in districts and schools in disseminating and mainstreaming the good practices arising from these projects. In future work, greater emphasis must be given to the development of high quality approaches to the assessment of student learning.

**Rationale and Strategies:** This is the foundation for all recommendations. The recommendation is based on the unequivocal experience of positive educational reform in the past two decades, both in Indonesia and internationally. The key strategy is to continue and to strengthen the professional knowledge and skills of teachers *and all administrators* working closely with children, particularly their skills in assessing student learning.

This Report cannot be silent on weak professionalism in the education system, and so draws attention to the findings of recent international studies on improving schools and school systems. These studies are discussed in section 4.5. Excellent skill in the educational assessment of student learning is an essential, non-negotiable component of teacher professionalism.<sup>4</sup> It is also a non-negotiable skill requirement among those educational administrators who are working in support of curriculum, assessment, schools, students and teachers. The paramount importance of attention to teacher professional development in assessment is one of the clear lessons learned from the discussion in section 4.2 of China's reform experience.

The emphasis on *educational* assessment is difficult to exaggerate. The neglect of the *educational* in assessment is most apparent in the negative backwash effects of the UN on learning and teaching in schools. It is apparent in the lack of alignment between curriculum objectives, learning and teaching methods, and assessment (see section 7.4.2 which explains the impact of alignment on students).

Too much of the debate on assessment is dominated by administrators, psychometricians, test-item development specialists, and now by quality assurance and 'standards' specialists. All these disciplines have important *parts* to play in educational assessment. It is stressed that the focus has to be *education*, not psychology, the advanced statistical manipulation of data, standards, or quality assurance concerns. It is only the experienced and professional educator with a strong background in teaching students that can see the *whole* picture of student learning and its assessment in its complex social and curriculum context.

A sense of responsibility in assessment practices is clearly missing among too many Indonesian teachers and administrators in the present UN system. This is because the stakes are so high and they have such limited involvement. This leads inevitably to a loss of teacher morale, an outcome of high stakes testing noted by Nichols and Berliner (2005; 2007). Responsibility need not be missing. The successful implementation of school based management and the shift towards student-centred active learning has demonstrated unequivocally that Indonesian teachers and local communities will respond responsibly, enthusiastically and professionally to the opportunities placed before them when they are supported to do so (DBE-1, 2009). Section 4.4 makes the clear link between this kind of success in Indonesia with the successes reported in high-achieving education systems such as in Finland. Donors have a continuing role to play in this type of supporting work.

**Recommendation 2: Increase the supply of high quality assessment leadership by supporting GOI to establish an independent examinations authority**

**Potential for donor assistance:** Uncertain. This is an internal political matter. Donor assistance may support the work of a professional partner as proposed below.

---

<sup>4</sup> The importance of this professional ability in assessment is stressed here. In his summary of over 500,000 studies of the influences on student achievement, feedback stands out as the single most powerful influence. See: John Hattie, *Teachers Make a Difference: What is the Research Evidence?* Melbourne, Australian Council of Educational Research, October 2003. Available: [www.acer.edu.au/workshops/documents/Teachers\\_Make\\_a\\_Difference\\_Hattie.pdf](http://www.acer.edu.au/workshops/documents/Teachers_Make_a_Difference_Hattie.pdf)

**Rationale and Strategies:** An urgent reform is to institute a revised system of governance and management of the assessment process under a single, independent and professional authority that reflects international best practice that is now implemented in many countries. It is now widely understood that the performance of the public examination function under direct Ministerial control is not effective in modern, democratic states and that this form of governance exposes the function to unacceptable levels of bureaucratic and political interference.

The authority must have a clear commitment to the highest integrity and professional standards and the capacity to eliminate direct interference, malpractice and cheating. It should have access to the latest technologies to assist in achieving an innovative corruption-proof system. It should also have the capacity to exert clear, strong professional and administrative leadership in the assessment of student learning in Indonesia at all levels of government.

To achieve a rapid transformation from the current situation, the proposed authority should seek the support of an acknowledged, internationally reputable strategic partner into the medium term. The partner will be able provide outstanding professional support and advice in the building of an Indonesian examination authority of recognized excellence.

The selection of such a partner must be done with great care. In Indonesia's situation, the potential partner must be able to demonstrate experience in the broadest sense of working at the intersection of curriculum, learning and teaching and assessment. Experience in the technical aspects of assessment alone will not be sufficient. Partners may be available close to Indonesia and could, conceivably, form a partnership consortium through existing regional cooperation arrangements enabling Indonesia to learn from and with its neighbours such as the Philippines, Australia, New Zealand and Singapore.

An early task for the new assessment authority will be to review, integrate and simplify assessment policies and regulations and to address the complexity and uncertainty in the system.

### **Recommendation 3: Increase the demand in society for high quality assessment**

**Potential for donor assistance:** Moderate. This recommendation could be integrated with Recommendation 1, particularly in relation to school committee experience where DBE-1 is at the forefront of development in this approach to management. Donor involvement in media campaigns and public consultations may be unwelcome.

**Rationale and Strategies:** This is a continuous and long-term goal for the assessment authority to achieve. The focus will be on both public examinations, such as the UN, and on the quality of school based assessment. The most comprehensive and effective strategy to combat widespread misunderstanding of assessment practices and standards that will touch all future citizens is to work towards ensuring that all students in the education system experience high quality assessment during their school years. This experience will come gradually from the other improvements nominated here, and particularly from improving professional standards.

In the short-term, however, additional strategies will be essential. Three strategies are proposed to support demand for high quality assessment, enhance transparency and accountability and also to contribute to a continuing process of review and development of assessment:

- A carefully constructed media campaign to 'socialize' the importance of quality assessment and to explain to parents what they can expect from schools in the assessment of their children's learning. How they may support these expectations by contributing to the work of their schools through school committees should also be

socialized.

- A public consultation process to both raise awareness and to seek systematic feedback on assessment proposals in Indonesia as one input into a wider process of review and policy development. The lessons learned about public consultation in China's reform experience (see section 4.2) need to be integrated into this process.
- Exploring strategies to strengthen the appropriate role of school committees in monitoring the assessment of student learning in schools and in strengthening school-based demand for professional and ethical practices.

#### **Recommendation 4: Align curriculum, learning and teaching, and assessment**

**Potential for donor assistance:** High. Assistance can be provided to drive alignment upwards through better teacher training and professional development approaches and downwards by assisting both the Educational Assessment Centre and Curriculum Centre to work on the identified issues of alignment.

**Rationale:** One of the most prominent researchers in this area in the past 20 years, John Biggs, introduced the concept of 'alignment' to better describe the essential connection between the curriculum, learning, teaching and assessment (Biggs, 1999). Alignment is fundamental to valid assessment and valid assessment is the foundation upon which any credible national examination system is built. The further tension between an unaligned, centralized, national examination and the decentralized school-level curriculum must be addressed.

The single assessment authority is key to achieving alignment because it will have the educational capacity and responsibility to ensure this is achieved as the foundation for examination validity. Two other strategies are required:

- Reviewing the teacher-training curriculum to strengthen and place emphasis on the importance of aligning curriculum, learning, teaching and assessment, and on the development and use of assessment methods appropriate for different contexts and subjects, and on the interpretation and use of assessment information.
- Reviewing all training materials used in the professional development of teachers by government, donors and NGOs to reflect the same elements identified for strengthening in the teacher-training curriculum above.

#### **8.2.2 Target 2 Strategies for an Ethical Education System**

**Recommendation 5: Increase efforts through research and management to better understand the mechanisms of cheating and implement coordinated policies and practices to eliminate it from the education sector**

**Potential for donor assistance:** Moderate. May be a sensitive area. However, some assistance in research would be acceptable, as would anti-corruption strategies through improved governance and management, building on donor experience in these domains. Anti-corruption and malpractice strategies can be designed into professional development and management processes.

**Rationale:** The full extent of cheating in the assessment system is unknown. The critical issue is that it is openly acknowledged, widely known to exist and that its scale and existence means that test scores lack validity. Several stakeholders have independently estimated that (a) although it is not universal, that (b) at least 50% of all candidates would be directly or

indirectly (because of the actions or inaction of others) involved in cheating in any one year.

As explained in section 7.4.2, the currently implemented UN is teaching children a range of values that moral educators would deplore.

In section 7.3 it was observed that the current *Renstra* of the Ministry of National Education has a strong conceptual framework for promoting ethical issues. What is missing from this framework is an explicit strategy for implementation and monitoring of these values. The policy is not aligned with any implementation procedure. It is recommended that this missing link receive attention in the forthcoming period. Consideration does need to be given to the tolerance of those adults who are involved in malpractice and who encourage and assist in student cheating by addressing this, when proven, very much more decisively than is currently the practice.

Managing cheating requires the development of evidence-based strategies that focus on reducing the incentives to cheat through reducing the 'stakes' involved in testing. There is an extensive literature on student cheating and what can be done about it that can assist in its management. Mere adjustments to technical procedures and statistical analyses that identify patterns of cheating are not the solution to a deep-seated, systemic and cultural issue.

**Recommendation 6: Create and develop a professional and ethical environment in which malpractice and cheating is denied the opportunity to exist**

**Potential for donor assistance:** See Recommendations 1 and 2.

**Rationale:** The strategies for Recommendation 1, above, provide the foundation for this recommendations to build an ethical system and will not be repeated here. In summary, the best long-term strategy to defeat educational malpractice is to build a quality system that prevents it from happening.

Part of the prevention strategy must be an urgent review of the processes which use data from examinations in determining the allocation of resources for school quality improvement and the development of standards ratings. The news item dated 24 November 2009 included in Appendix 3 has to be considered as an early warning of the further damage such use of data will do in the education of children and in increasing the 'stakes' rather than reducing them.

**8.2.3 Target 3 Strategies for a High Quality Credential at the End of Year 9**

**Recommendation 7: Develop a high quality and credible credential at the end of year 9 through undertaking a systematic and professional review of the UN, with international best practice input**

**Potential for donor assistance:** High. High-level technical assistance in the curriculum process generally and in educational assessment specifically will be necessary. Donors may also be able to support a regional cooperative arrangement to foster the sharing of change and development strategies as well as technical expertise and this cooperation could be integrated into the recommended review process.

**Rationale:** The present arrangements for the UN are not providing students with a credible credential. Stakeholders may believe they have a credible award but in fact the data (scores) on student learning are neither valid nor reliable. That situation strengthens the case for more school-based assessment. It is illogical to argue that one two hour multiple choice test is going to provide higher quality data than the judgement of a teacher who has been working with the student for at least a year.

Because of this lack of validity and reliability, the same data is unsuitable for use in educational planning. In the case of planning, even valid and reliable test scores should not be assumed to be a strong indicator of school quality or school needs.

The present arrangements are also distorting the educational process in schools, narrowing the curriculum, and driving teaching and learning strategies away from developing the competencies required in the curriculum towards superficial and trivial learning demanded for success by the dominant multiple-choice test format.

The current UN is making it even less likely that Indonesian students performance in internationally benchmarked tests, PISA and TIMSS, will lift the nation's position from among the world's worst performing nations.

To address these challenges, the new assessment authority will need to develop guidelines for a systematic and professional review of the UN with international best practice input to assist in the following tasks:

- Reduce the stakes in the current system. This may include the introduction of student choice in a range of examinable subjects, not only the present four (and soon to be five with Social Science added) and the introduction of competency standards to reflect the curriculum. Reporting in terms of bands of achievement rather than pass-fail will have educational benefits and also reduce the stakes.
- Reduce the emphasis on one multiple-choice examination per subject, improve the technical quality of all assessment methods (including, also, MCQs if retained) and increase the range of assessment methods to ensure that curriculum competencies are assessed to achieve valid and reliable data.
- This reduction in emphasis on one multiple-choice examination per subject will necessitate the inclusion of a school-assessed component and a rigorous methodology of statistical moderation. The Minister of National Education Regulation No 39 year 2007 on National Standard Final School Examination (Ujian Akhir Sekolah Berstandar Nasional (UASBN) for primary schools in 2007/2008 provides an established model to achieve this immediately.
- Revise the grading system to remove the depressing effect of the present pass-fail system and introduce performance bands of achievement.

The inclusion of a moderated school-assessed component suggested here would go some way to addressing the paradox concerning decentralization discussed in section 4.7 by giving some real responsibility to schools, who after all, are meant to be implementing their own school curriculum.



## 9 REFERENCES

- Battin-Pearson, S., Newcomb, M.D., Abbott, R.D., Hill, K.G., Catalano, F., and Hawkins, J.D. (2000). Predictors of Early High School Dropout: A Test of Five Theories. *Journal of Educational Psychology*, 92, 3, 568 – 582.
- Arze del Granado F.J., Fengler, W., Ragatz, A. and Yavuz, E. (2007). *Investing in Indonesia's Education: Allocation, Equity, and Efficiency of Public Expenditures*. The World Bank, Poverty Reduction and Economic Management (PREM), Jakarta, Indonesia. Available: <http://mpr.ub.uni-muenchen.de/4372/> Accessed: 1 October 2009.
- Australian Council for Educational Research (ACER). (2008). *The National Examinations, Indonesia. An Assessment Strategy for Indonesia Education Sector Assessment (ESA) Project*. ACER for AusAID, Brisbane.
- Biggs, J. (1999). *Teaching for Quality Learning at University*. SRHE, Buckingham.
- Caldwell, B. and Harris, J. (2008). *Why Not the Best Schools?* ACER Press, Melbourne.
- Chatterji, M. (2003). *Designing and Using Tools for Educational Assessment*. Pearson Education, Boston.
- Cost Effectiveness in Education - Methodology, Examples, Use of Cost-Effectiveness Analysis*. (2009) Available: <http://education.stateuniversity.com/pages/1887/Cost-Effectiveness-in-Education.html> Accessed: 11 November 2009)
- DBE-1. (2009). *Implementing School-Based Management in Indonesia, The DBE-1 Experience. Impact Study 2009*. Unpublished draft report to USAID.
- Departemen Pendidikan Nasional. (2009). *Panduan Kebijakan Pemanfaatan Hasil Ujian Nasional Untuk Perbaikan Mutu Pendidikan. Ujian Nasional Tahun 2008/2009*. Jakarta.
- Hughes, A. (2003). *Testing for Language Teachers* (2nd Ed.). Cambridge University Press, Cambridge.
- Hattie, J. (2003). *Teachers Make a Difference; What is the Research Evidence?* Australian Council for Educational Research, Melbourne. Available: [http://www.acer.edu.au/documents/RC2003\\_Hattie\\_TeachersMakeADifference.pdf](http://www.acer.edu.au/documents/RC2003_Hattie_TeachersMakeADifference.pdf) (Accessed: 19 September 2008.)
- Institute of Educational Sciences, US Department of Education (2007). Trends in International Mathematics and Science Study (TIMSS). Available: <http://nces.ed.gov/timss/index.asp> (Accessed: 1 November 2009).
- Jalal, F. et al. (2009). *Teacher Certification in Indonesia: A Strategy for Teacher Quality Improvement*. Ministry of National Education and the World Bank. Jakarta.
- McKinsey and Company. (2007). *How the World's Best performing School Systems Come Out on Top*. Available: [http://www.mckinsey.com/clientService/Social\\_Sector/our\\_practices/Education/Knowledge\\_Highlights/Best\\_performing\\_school.aspx](http://www.mckinsey.com/clientService/Social_Sector/our_practices/Education/Knowledge_Highlights/Best_performing_school.aspx) Accessed: 16 October 2009.

- Mehrens, W.A. and Lehmann, I.J. (1973). *Measurement and Evaluation in Education and Psychology*. New York, Holt, Rinehart and Winston.
- Ministry of National Education (MoNE) (2005). *Strategic Plan, Ministry of National Education, The Republic of Indonesia, 2005 – 2009*. Jakarta.
- Mortimore, P. and Mortimore, J. (1999). Improving Educational Performance of At-Risk Youth. In: *Preparing Youth for the 21<sup>st</sup> Century: The Policy Lessons from the Past Two Decades*. Proceedings of a Conference organized by the OECD and US Department of Labor and Education, Washington, DC.
- Nichols, S.L. and Berliner, D.L. (2005). *The Inevitable Corruption of Indicators and Educators Through High-Stakes Testing*. Education Policy Research Unit, Arizona State University, Tempe, Arizona. Available: <http://epicpolicy.org/files/EP5L-0503-101-EPRU.pdf> [also available and updated in Nichols, S.L. and Berliner, D.L. (2007). *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*. (2007). Harvard Education Press, Cambridge, Mass.]
- OECD (2007). *PISA (2006)*. Available: <http://www.pisa.oecd.org/dataoecd/15/13/39725224.pdf> (Accessed: 1 November 2009)
- Ogilvie, M. (2004) *Evaluation of the UAN*. Available: <http://re-searchengines.com/art05-4.html> (Accessed 5 November, 2009)
- Queensland Studies Authority. (2008). *The Assessment Systems of Queensland and Finland*. Research Report 1. Brisbane. Available: [http://www.sqa.org.uk/sqa/files\\_ccc/PNP\\_ResearchReport1\\_TheAssessmentSystemsFinlandQueensland.pdf](http://www.sqa.org.uk/sqa/files_ccc/PNP_ResearchReport1_TheAssessmentSystemsFinlandQueensland.pdf) Accessed: 12 November 2009.
- Setiadi, H. 2007. Accurate and Fair Ujian Nasional (UN) for Schools in Indonesia which are Varied in Condition and Quality. *Jurnal Pendidikan dan Kebudayaan*, 065, Maret. Available: [http://www.depdiknas.go.id/publikasi/balitbang/066/editorial\\_j65.html](http://www.depdiknas.go.id/publikasi/balitbang/066/editorial_j65.html) Accessed 23 October 2009.
- Spector, B.I., Johnston, M. and Winbourne, S. (2009). *Anticorruption Handbook; Final Report*. USAID, Washington, D.C.
- Usman, S. Akhmadi, and Suryadarma, D. (2004). *When Teachers are Absent: Where Do They Go and What is the Impact on Students?* The SMERU Research Institute. Jakarta.
- Wikipedia. (2009). *Academic Dishonesty*. Available: [http://en.wikipedia.org/wiki/Academic\\_dishonesty](http://en.wikipedia.org/wiki/Academic_dishonesty) Accessed: 3 November 2009.

## 10 SCHOOLS AND INSTITUTIONS VISITED<sup>5</sup>

Location	School/Institution	Persons met
Jakarta	Directorate General of Primary and Secondary Education, Ministry of National Education	See footnote
	Baden Standar Nasional Pendidikan (National Education Standards Authority)	
	Directorate of Junior Secondary Education	
	Educational Assessment Centre	
	Central Library	
Jakarta	UNICEF	
	Indonesia Corruption Watch	
	DBE-3 (Examination Review Panel)	
Makassar, Sulawesi Selatan	Dinas Pendidikan, Propinsi	
Makassar, Sulawesi Selatan	Kantor Wilayah, Departemen Agama, Sulawesi Selatan	
Makassar, Sulawesi Selatan	SMA	
Makassar, Sulawesi Selatan	SMP	
Pangkep, Sulawesi Selatan	SMP	
Pangkep, Sulawesi Selatan	MTs	
Medan	DBE-3	

<sup>5</sup> Participants in meetings, particularly in schools, were given an assurance of confidentiality given the sensitivity of topics discussed. Individual's names have therefore neither been recorded nor – as a result – listed here. In this final version of the Report the names of schools have been removed for the same confidentiality reason. The DBE-3 office keeps a confidential record of schools visited.

## **APPENDIX 1: OVERVIEW OF THE UN IN INDONESIA**

### **Nomenclature and Formal Responsibilities**

The Ujian Nasional (UN), or National Examination, is a Government of Indonesia school examination administered at the end of nine years basic education and at the end of twelve years of senior secondary education.

There is no clear distinction made in nomenclature between these two examinations and this leads to confusion. Discussions about the UN for SMP frequently become confused with references to the UN at the end of senior secondary education. The Ministerial decree uses the terms 'UN SMP/MTs/SMPLB', 'UN SMA/MA/SMALB' and 'UN SMK'.

In Government Regulation No 19 year 2005 on National Education Standards, the responsibility of Ujian Nasional is assigned to the Badan Standar Nasional Pendidikan (BSNP), the National Education Standards Authority. In implementation, BSNP cooperates with relevant institutions in central government, provincial government, district government and educational institutions. In the Minister of National Education Regulation No 34 year 2007 on National Examinations for School Year 2007/2008, the Pusat Penilaian Pendidikan (Puspendik, the Educational Assessment Centre), assists BSNP in technical matters, such as managing the item bank, developing the test paper masters, scoring of UN results and overall data management of the examination system.

The present analysis is concerned only with the UN that is administered at the end of nine years basic education. This UN is for Sekolah Menengah Pertama (SMP) and Madrasah Tsanawiyah, (MTs).

For simplicity, the term UN will be used in this report and it is meant to apply only to SMP and MTs.

### **Rationale**

The UN is the assessment of student learning at the end of Year 9. It is conducted by the Government nationally for four subjects – Bahasa Indonesia, English, Mathematics and Science (Government Regulation No 19 year 2005). The results of the UN are used for two main purposes, determining student progress and for school quality management.

The primary purpose is to assist in determining the graduation of students from a program or educational institution to higher levels of education, particularly senior secondary education. The regulations require more than the UN alone in making this determination.

A second, but very important purpose, is to provide data to map school and program quality and as a tool in determining assistance given to schools to improve educational quality.

### **Synopsis of Governing Laws and Regulations**

The principal laws and regulations governing the UN are:

1. Act No 20 year 2003 on National Education System: general statements on evaluation of student's achievement.
2. Government Regulation No 19 year 2005 on National Education Standards: description of each of the national education standards, the standard of educational

assessment for basic and secondary education consists of – assessment of learning outcomes by teachers, assessment of learning outcomes by educational institutions, assessment of learning outcomes by Government.

3. Minister of National Education Regulation No 22 year 2006 on Standard of Content for Basic and Secondary Educational Institutions: provides the minimal content for the level of minimum competencies to reach the minimum level of graduate's competencies.
4. Minister of National Education Regulation No 23 year 2006 on Standard of Graduate's Competencies for Basic and Secondary Education Institutions: standards of graduate's competencies is used as assessment guide in determining graduation of students. It is the minimum graduate's competencies for educational institution, the minimum graduate's competencies for groups of subjects and the minimum graduate's competencies for each subject.
5. Minister of National Education Regulation No 6 year 2007 on Implementation of Minister of National Education Regulation No 22 year 2006 on Standard of Content for Basic and Secondary Educational Institutions and Minister of National Education Regulation No 23 year 2006 on Standard of Graduate's Competencies for Basic and Secondary Education Institutions: for implementation of Standard of Content and Standard of Graduate's Competencies.
6. Minister of National Education Regulation No 20 year 2007 on Standard of Educational Assessment: definitions of terms, principles of assessment, technique and instruments used for assessment, mechanism and procedure of assessment, assessment by teachers, assessment by school/madrasah examination, and assessment by Government (Ujian Nasional).
7. Minister of National Education Regulation No 34 year 2007 on National Examination for SMP/MTs/SMP-LB, SMA/MA/SMA-LB, and SMK for School Year 2007/2008<sup>6</sup>: definition of terms, objectives, use of UN results, criteria for student eligibility to participate in UN, subjects examined in UN, Standard of Graduate's Competencies for 2008, item bank managed by Pusat Penilaian Pendidikan, examination papers reviewed and set by BSNP, printing of examination papers, implementation of UN, responsibility of the Minister, BSNP, governors, higher education institutions, bupati/walikota, Indonesian Ambassadors, schools/madrasah, Independent Monitoring Team, scanning of answer sheets, scoring of UN results, passing grade, budget, confidentiality, and consequences of violations.
8. In relation to violations, Article 17 of this Regulation 34 requires that "Individuals, groups, and/or institutions involved in implementation of UN are required to keep secrecy, honesty, safety, and the smooth operation of the UN implementation and states that those "that conduct violations or irregularities will get sanctions according to law and regulations". Students caught cheating are considered to have failed. Chapter IX of BSNP Decision No 984/BSNP/XI/2007 sets out further sanctions as follows: student cheating – initial warning which, if ignored, leads to removal from examination room and failure; room supervisors, members of the Independent Monitoring Team,

---

<sup>6</sup> Issued each year for respective school year.

and UN organisers – will not be involved in future UN activities, replaced, and organisers dealt with according to laws and regulations.

9. BSNP Decision No 984/BSNP/XI/2007 on Standard Operation Procedure for National Examination for SMP, MTs, SMP-LB, SMA, MA, SMA-LB, and SMK School Year 2007/2008<sup>7</sup>:
  - a. Participants - criteria for participants of UN, registration of participants,
  - b. Organizing of UN - components of UN organizers at central government and their role and responsibility, components of UN organizers at provincial government and their role and responsibility, components of UN organizers at district/city government and their role and responsibility, UN organizers at school/madrasah and their role and responsibility.
  - c. Examination papers – development of graduate’s competencies standard for 2008, development of *Kisi-kisi*, development of master copy of UN paper, sending of master copy to Provinces, printing of UN papers, storage and distribution to all district/city in the Province.
  - d. Implementation of UN – schedule of UN, schedule of announcement of results, examination room, supervisor in examination rooms, participants order.
  - e. UN results – collection of answer sheets, processing of UN results.
  - f. Graduation from UN – criteria for graduation.
  - g. Monitoring, evaluation and reporting.
  - h. Budget for implementation of UN – components for budgeting, budget for UN at central government and regional government.
  - i. Sanctions – for UN participants, room supervisors Independent Monitoring Team, UN organizers.
  - j. Attachment – explanation on criteria for graduation from educational institutions.

There is another regulation for school/madrasah examination : Minister of National Education Regulation No 5 year 2008 on School/Madrasah Examination for School Year 2007/2008. This regulation is issued each year with necessary adjustments.

### ***Discussion***

This long list of laws and regulations is an indicator of the complexity in managing a large, public national examination system. Given the added complexity of three levels of government, the very significant challenges of Indonesian geography and distance, and the SMP/MTs school population size (2,518,561 candidates in 2007/2008), it is a tribute to all involved that the process works as effectively as it does.

An issue that is not explored here at all is raised as a question: “Do the complexities named here create administrative and educational difficulties that could be resolved by undertaken a thorough review of the whole administrative system to achieve a greater degree of simplicity?”

### **Construction and Implementation of the UN**

The responsibility for administering the UN rests with the BSNP assisted by Pusat Penilaian Pendidikan (Puspendik). In the administration of the UN, BNSP is responsible to:

---

<sup>7</sup> issued each year for respective school year

- a. Establish the central level UN organizer
- b. Conduct quality assurance of examination paper packages
- c. Prepare master copies of papers with Puspendik
- d. Coordinate implementation of UN that is honest
- e. Monitor the appropriateness of printing companies selected by Governors
- f. Conduct supervision of scanning of the answer sheets
- g. Establish an Independent Monitoring Team for UN
- h. Coordinate the monitoring of UN
- i. Develop and issue Standard Operation Procedure for UN
- j. Evaluate implementation of UN
- k. Report on the implementation of the UN to the Minister.

The process by which the test is constructed is laid down in POS issued by BSNP.

The development of the examination follows these steps:

- a. Development of standard of graduate's competencies for respective school year
- b. The standard is proposed to Minister who will issue the decree
- c. *Kisi-kisi* (table of specification) is developed from the standard of graduate's competencies
- d. Development of master copy of UN papers. Number of items and time allocation for each paper is listed in the POS.

Development of master examination copy:

- a. Identify and select test items according to *Kisi-kisi* from the national item bank
- b. Assemble the UN test papers and consider the difficulty of items
- c. Determine the exam packages with consideration for equality among packages
- d. Check the exam packages for equality of difficulty, quality and validity
- e. Develop the layout for exam packages
- f. Provide coding for the exam packages
- g. Duplicate master copies for the Provinces

Distribution and printing:

- a. The master copies are sent to Provinces.
- b. Technical criteria for printing companies in the Provinces are issued from the central level. The governor will determine (through a bidding process) which printing company is selected according to established criteria.
- c. Printing of the exam papers in the Provinces.
- d. Exam papers are distributed to districts. Time of distribution is as close as possible to UN, (depending also on distance of district to capital of Province).
- e. Organizers at district/kota distribute UN papers to schools/madrasahs daily according to what subject is being tested on the day

### **The Administration of the UN in Schools**

At the conclusion of the actual examination, another carefully planned administrative process begins with the collection of answer sheets.

- a. After the exam is finished, the answer sheets are put into an envelope and sealed
- b. The principal will put all the envelopes with answer sheets into a bigger envelope
- c. These envelopes are sent to UN district organizers

- d. From the district the envelopes will be sent to UN provincial organizer
- e. BSNP and Puspendik has developed a system and software for data of UN candidates, scanning of answer sheets, analysis and reporting of UN results

At Province level:

- a. The answer sheets are checked and classified according to subjects
- b. Answer sheets are scanned
- c. One week after UN conducted the scanning results are sent to central/national level. At the latest 3 weeks after UN all scanning results have been sent
- d. BSNP and Puspendik conduct scoring and assessment of UN results
- e. BSNP and Puspendik send the UN results to UN provincial organizers
- f. UN provincial organizers print the accumulative list of UN results
- g. UN provincial organizers print report of UN results

Graduation:

Criteria for SMP: average minimum of 5.25 for all UN subjects, no score below 4.25 or minimum 4.00 for one of the subjects, while other subjects are minimum 6.00

District/Kota can determine a higher passing grade than the criteria above

### **The UN Results**

The results of the UN are used for consideration in:

- a. Mapping of program quality and schools
- b. For entry to a higher level of education
- c. To assist in determining the graduation of students from a program or educational institution
- d. As a tool in the supervision and assistance given to an educational institution in its effort to improve education quality

It is important to appreciate the difference between successful completion of the UN and graduation. A passing grade in the UN is essential for graduation and should be only one of four components for graduation (although reports from Districts and schools suggest this is not always the case).

Article 72 of the Government Regulation No 19 year 2005 on National Education Standards states that a student can graduate from an educational institution when they have:

- a. Completed the educational program
- e. Satisfactory scores are achieved at the final assessment for all subjects in Religious education and noble character, Citizenship and personality, Aesthetic subjects, and Physical education, sports and health
- f. Passed the school/madrasah examination for subjects in science and technology
- g. Pass the UN.

## **APPENDIX 2: MINIMUM SERVICE STANDARDS, THE NATIONAL STANDARDS AND SCHOOL ACCREDITATION**

### **Minimum Service Standards (MSS)**

- MSS for Education do not stand alone; they are closely aligned with both National Standard of Education (NSE) and School/Madrasah Accreditation.
- MSS for Education is developed based on National Standards of Education (NSE) and Instrument for School and Madrasah Accreditation.
- MSS is an implementation strategy for achieving NSE step-by-step.

### **National Standards for Education (NSE) – see also below.**

- The NSE were developed under Law 20/2003 and Government Regulation 19/2005 by an independent Board (BSNP) and covers 8 aspects of schooling: (a) standard of content; (b) standard of process; (c) standard of teacher, principal and education personnel; (d) standard of equipment and infrastructure; (e) standard of management; (f) standard of financing; (g) standard of evaluation; and (h) standard of graduate competence.
- The NSE contains comprehensive and detailed standards for each level of schooling – primary, junior secondary, senior secondary and vocational high schools – that are expected to be fulfilled by 2014. Implementing MSS is the first stage towards ensuring achievement of the NSE.

### **School Accreditation**

- The National Board of Accreditation (BAN-SM) developed the accreditation instrument for schools and madrasah based on NSE.
- The accreditation status of a school/madrasah is based on their overall score (A to D) derived from a series of rating scales. (A = highest; D = lowest). Rating of D or below is unsatisfactory.
- Schools who do not meet the lowest level (D) on the accreditation are the target group for MSS, together with schools that have not yet taken part in the accreditation process but are thought to be at risk.

**Source:** Basic Education-Sector Capacity Support Program (BE-SCSP). May 2009. *Component 1: Central, Provincial, and District Capacity Development for Decentralization Progress Report*

## National Standards for Education (NSE): Key Articles and Comments

### CHAPTER I GENERAL PROVISIONS

#### Article 1

**Definition:** Educational evaluation standards refer to the national standards of education related to the mechanism, procedures, and instruments for evaluating the learning results of the students.

**Note:** Part Two, Evaluation of Learning Results by Teachers, Article 64, confuses *what* is being evaluated by stating: 'The evaluation of the learning results in the group of subject matters related to science and technology is measured through tests, assignments, and/or other forms that are to conform with the characteristics *of the materials being evaluated*. Materials are not being evaluated, it is, by definition 'the learning results of the students'.

### CHAPTER V COMPETENCE STANDARD OF GRADUATES

#### Article 26

The group of subject matters related to science and technology at the SMP/MTs/SMPLB/Package B or any other similar forms, is meant to provide a basic competence in science and technology and to cultivate critical, creative and independent scientific thinking.

The competence standard of graduates at the basic education level is targeted to lay down the basis of intelligence, knowledge, personality, noble character, and skill to live independently and to be able to continue studies.

#### Part Four Evaluation of Learning Results by the Government

#### Article 66

The evaluation of learning results as specified in Article 63 para (1) sub c has the purpose of evaluating the competence achievement of the graduates on national basis and in certain subject matters within the group of subject matters related to science and technology, and is undertaken through a national examination.

#### Article 68

The results of the national examination are used as one of the considerations for:

- a. charting the quality of the program and/or the education unit;
- b. the basis of the selection in order to enter the next education level;
- c. determining the graduation of the students in the program and/or education unit;
- d. developing and providing of assistance to the education unit in an effort to improve the quality of the education.

#### Part Five

### Passing Marks

The students are declared to have passed the education unit of the basic and intermediate education levels after:

- a. they have finalized the entire learning program;
- b. they have achieved minimally good marks at the final evaluation for all subject matters covering the group of subject matters related to religion and character building, the group of subject matters related to citizenship and personality, the group of subject matters related to esthetics, and the group of subject matters related to physical, sport and health education;
- c. they have passed the school/madrasah examination with regard to the group of subject matters related to science and technology; and
- d. they have passed the National Examination.

The graduation of students from the education units is determined by the education units concerned to conformity with the criteria developed by BSNP and laid down in a Ministerial Ordinance.

**Source:** Government Regulation of the Republic of Indonesia, Number 19, Year 2005, Concerning The National Standards of Education.



### APPENDIX 3: EVALUATION OF RECENT NEWSPAPER STORIES

Location of Story	Source of Story	Headline	Story Synopsis
<p><i>West Sumatra, Riau NTT</i></p>	<p><i>Jakarta Post 23 June 2009</i></p>	<p><i>National exam pass rate declines in W. Sumatra, Riau</i></p>	<p><i>Provincial education offices will evaluate the drop in pass rates, whether it is because of an increased number of schools or student's ability to absorb the subjects. Schools for the disabled contributed the highest number of failures. NTT recorded a zero pass rate in 10 junior high schools. Provincial education head reported that 70.25% passed this year compared to 46.36 last year.</i></p>
<p style="text-align: center;"><b>Comment</b></p> <p>This item reflects poorly on the professionalism of the provincial offices, assuming the reporting is accurate. In the first case, blame for failure is placed on an unlikely factor – numbers of schools and on students, some of whom are disabled, which is a common strategy, instead of accepting responsibility. No mention is made of teaching or administration that might enter the evaluation. The zero pass rate in 10 schools is also an unlikely result and suggests, if a true measure of student achievement, very poor teaching, management and supervision by Dinas Pendidikan of schools there. The situation in the 10 schools would have been known and addressed if local management had been professional. A between year variation in pass rates of 24% indicates serious data errors rather than a true outcome.</p>			
<p><i>National</i></p>	<p><i>Jakarta Post 23 June 2009</i></p>	<p><i>Violations taint exams despite higher grades</i></p>	<p><i>Discusses pass rates. National high school exams are tainted with violations. Examples cited include questions being leaked, procedural mistakes at regional level, students openly cheating and teachers correcting results.</i></p>
<p style="text-align: center;"><b>Comment</b></p> <p>The article highlights administrative incompetence and corruption. Nevertheless, results are presented in the article with great precision “93.74% passed the exam”, and uncritically. The evidence of teacher and administrator corruption or incompetence is clear but the Ministry is not reported as having anything to say to address this issue. Teacher correction of results has been corroborated by educators interviewed in this study – sometimes with sad results. In one case it is alleged that an attempt to correct students’ responses to ensure a pass backfired when the teacher concerned, through incompetence, changed correct responses to incorrect responses leading to student failure. This is one plausible explanation for the zero pass rates in NTT noted in the previous story.</p>			

Location of Story	Source of Story	Headline	Story Synopsis
National	Jakarta Post 1 October 2009	Govt mulls over asking students to take more graduation exams	The government is considering whether to put in place more exams for students graduating from junior and senior secondary schools, to assess their physical skills and behaviour. Education expert Arief Rachman said the country still needed a national exam, but the government should take into consideration students' varying levels across schools in different regions to avoid cheating. The UN results are also used to assess the quality of schools across the country.
<b>Comment</b>			
Three issues are mentioned in this article, first that the Ministry would like to add Religion, Sports and Arts as subjects in the Ujian Nasional. Second, that the government should take into consideration the various levels of schools across the country in determining scores. The present system requires all students score an average of 5.5 to graduate and this leads to cheating. Third, students' scores, which may not be the real results and are corrupted, are used to categorize the quality of schools. Sadly, there is no discussion of the quality of assessment. More exams does not equal better exams.			
National	Kompas 7 November 2009 and Jakarta Post 7 November 2009	[Kompas] Entry test to University deleted UN has to be improved first	The new Minister of National Education has announced his 100-day plan. He hopes that the entry test to University can be deleted in 2012 because the selection for university entrance is integrated with UN. But before this system is implemented, UN (SMA/MA/SMK) implementation must improve. The University Rectors will assist in monitoring the problem points such as printing and distribution. Problems in UN implementation have to be straightened out first and build trust to UN. When there is trust in the UN then the separate entry test to university can be deleted.
<b>Comment</b>			
Too many problems in the implementation of UN have occurred during the years, doubts of quality of test items, cheating by students, teachers, and administration personnel. When the government has revised the system and universities trust the UN results, then the entrance test can be integrated with the UN. Although it is planned for 2012, if no major changes happen starting now then it is unlikely that the integration will be successfully implemented in 2012. The impact of the needs of university entry must be considered because of the known influence they will have on school curricula. There has to be a debate on the question: 'With such a small percentage of students from senior secondary school proceeding to university, why should universities have such a powerful influence on school level education?'			

Location of Story	Source of Story	Headline	Story Synopsis
<i>National</i>	<i>Jakarta Post 24 November 2009</i>	<i>Abuse against students rises: Commission</i>	<i>The National Commission on Child Protection says that instances of violence against children committed by teachers had increased dramatically compared to past years.</i>
<p style="text-align: center;"><b>Comment</b></p> <p>The observation made in the present story sends a very clear warning about the negative impact of high stakes testing and the further use of test scores in school quality management: “The violence occurred because teachers were under increased pressure to achieve targets set by their schools as part of efforts to maintain schools’ quality, while receiving salaries below standards set by the government.”</p> <p>This matter is very strongly linked to the recent report prepared by Cannon and Arlianti, (2009), <i>Transition to and Participation in Junior Secondary School</i>. Jakarta, USAID/DBE-3. This report finds abuse a significant issue in early school leaving.</p>			
<i>National</i>	<i>Jakarta Post 26 November 2009</i>	<i>Government insists exam must stay</i>	<i>The Minister of Education will seek a case review against a Supreme Court ruling which requires the government to revoke the final national examination. The ruling is in favour of a lawsuit claiming the government had failed to protect the rights of citizens to pursue education at a higher level because they did not pass the examination.</i>
<p style="text-align: center;"><b>Comment</b></p> <p>The story reflects community concerns about the UN. One of these concerns is reflected by the reported comment of the Head of BSNP: “There has been prevalent fraud in the implementation of the national exam, but we conduct evaluations and improvements every year.”</p> <p>The matter is unresolved and is continuing.</p>			



## APPENDIX 4: SAMPLE TEST ITEMS

### From the 2007/2008 Bahasa Inggris (English) Examination

#### Question 7

The stimulus material provided is from a warning sign that states: *Do not aim the camera directly at the sun or other intense light resources that could damage your eyesight.*

The word “resources” is used incorrectly (should be “sources”). “Resources” is then used in only one of the four choices available, leading the test-wise student directly to the answer.

Faults: incorrect use of English; very poor item construction.

#### Question 18

The stimulus material provided is an advertisement for an English teacher, which, among other information, states: *Required Exp/C TEFL preferred.*<sup>8</sup> The question asked is:

Which qualification must the teacher candidate have?

- A. Explicit
- B. Expired
- C. Expressive
- D. Experienced

One difficulty with this question is that it rewards guessing. The candidate can have no reasonable knowledge of what was really intended by the advertiser when using “Exp/C”. It is not an abbreviation that is commonly understood such as RI, TNI or USA. It is true that none of A-D is actually a “qualification” and so the correct answer should be “C TEFL”. However, A, C and D could be considered as plausible choices in the context of the stimulus material.

Faults: Construction rewards guessing; unreasonable expectation set in stimulus material.

### From the 2008/2009 Bahasa Inggris (English) Examination

#### Question 1

The stimulus material states: *Read the following question and answer question number 1.*

Due to many sharp corals, the visitors are prohibited to swim along this area.

1. Where do we usually find the text?

- A. At the lake
- B. At the beach
- C. In the river
- D. In the swimming pool

---

<sup>8</sup> C TEFL means Certificate of/in TEFL (Teaching English as a Foreign Language) but this information is NOT provided in the test paper.

The English in the statement has several errors. Errors in a national examination paper are bad enough, but in the very first question, errors are potentially very damaging to student confidence and test performance. This error in this position indicates very poor test preparation standards

The first language error is that we find text on a printed page or on a sign, not 'in' or 'at' the places listed.

"The visitors" is incorrect, as it is not intended to refer to a particular group of visitors, as implied here, but to all, or any, visitors. "Visitors" is the correct English in this case.

Also, the use of "along this area" is inappropriate. "Along" means following the boundary of something. People would not swim "along" the water, but "in" the water where the sharp coral grows. Other reasonable doubts about the English used in the item can be raised; these doubts are with the plural "corals" and the incorrect use of "prohibited to".

Faults: Incorrect use of English.

### **Question 9**

Question 9 is taken from a supplied text about Earth Day and the environment.

*Question 9. ... or even reverse the damage we've done to this beloved earth.*

What does the word "damage" in the sentence mean?

- A. Injury
- B. Hurt
- C. Destruction
- D. Reparation

The text discusses "Earth Day". It refers to "Earth" the planet, not a clod of earth a student may find in their garden at home, as used in the question stem. "earth" should be capitalized.

A, B and C can be correct; it could be either injury or hurt that have similar meanings, whereas destruction seems too strong, although children could select this given the emphasis on environmental issues these days. D is a poor distractor as it comes from a completely different class of meanings (technically, it should be a 'plausible distractor').

Faults: Incorrect English (capitalization); very poor set of alternatives and distractor.

### **From the 2007/2008 Science Examination**

#### **Question 19**

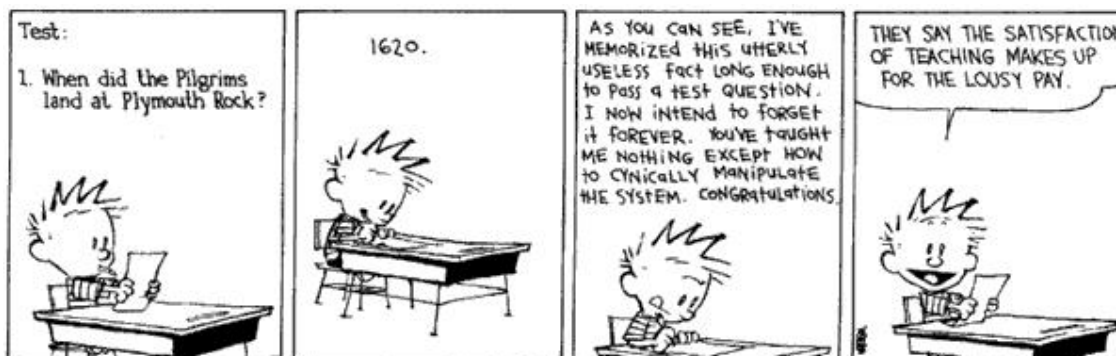
Question 19 presents a clear, well-drawn diagram showing the relative position of the planets in relation to the sun. Planets are labelled from 1 to 8 according to their position. Students are then asked to select the location of Mars from one of four choices.

Question 19 is a good example of a well-designed question, however, it illustrates testing at the lowest cognitive level of simple recall. It is this kind of question that encourages and rewards learning of simple facts and the 'drill-and-practice' type teaching that the curriculum

is supposed to be moving away from – but is not because of this type of testing.

The following Calvin and Hobbes cartoon<sup>9</sup> illustrates the issue with this recall-of-facts type of testing. It also illustrates the ‘life skill’ developed through the UN described in section 7.4.3.

**“You have taught me nothing except how to cynically manipulate the system. Congratulations.”**



© Universal Press Syndicate

## Question 20

Students are required to read four lines of detailed stimulus material that explains the four seasons in North America and Europe, for example, that winter is from December to March, and so on. The question is then asked about the cause of the changes in seasons. Four options are presented about the rotation or revolution of the Earth and Moon.

Faults: 1. The stimulus material is actually unnecessary in relation to the question asked and could have been very considerably shortened to make the test less of a ‘reading test’. The time could have been allocated instead to asking another question thereby increasing the potential reliability of this test. 2. The stimulus material actually contains wrong information; the specification of seasons provided does not follow standard international practice in naming the months in each of the four seasons experienced in temperate latitudes.

## General Comments

These examples illustrate item technical errors. Errors should never appear in a *national-level* examination. These technical deficiencies occur *within* a deficient approach to testing student achievement through the use of the multiple-choice format. The examples also highlight the difficulty in constructing good quality multiple-choice items. These issues are serious enough, but the multiple-choice format is then emulated in local try-out tests and in classrooms where the general standard of items is often so much worse than illustrated here.

This clear evidence of the serious weaknesses in the use of multiple-choice testing in Indonesia should be sufficient to cause an urgent re-think about how to assess student learning. Relying on multiple-choice questions, alone, because they are *practical* is absurd. The evidence shows that current practice is both technically unsound and is having a major, negative impact on education and on children. More or better quality items will not solve the problem. The end result is invalid and unreliable data from which to grade student achievement in Indonesia’s education system.

<sup>9</sup> Available: <http://progressiveboink.com/archive/calvinhobbes.htm>