

## **UN: TO BE CONTINUED OR ABORTED?**

by  
Tanzil Huda<sup>4)</sup>

**Abstract:** One of the qualities of education concerns with content standard. The implementation of the standard is the form of the issuance of Permen (Minister of Education's Decree) number 22 and 23/2006 which becomes the basis of conducting the teaching learning at school including its evaluation. *UN* (National Exams) as the final form of the evaluation has been expected to be the measurement of the students' achievement of the content standard. But the controversies arise to respond the implementation of *UN*. However, *UN* is supposed to be the most valid and dependable form of measurement of the content standard of students' achievement and of school.

Keywords: *UN* (National Exams), highstakes test, content standard

### **INTRODUCTION**

As an attempt to improve the quality of education, the government of Indonesia through the Ministry of Education has issued several important policies. Some of the important policies which have great influence toward day to day schooling are the minister of education's decree (*Permen*) number 22/2006 about content standard (*standar isi*), *Permen* number 23 about the standard competence of graduates (*standar kompetensi lulusan /SKL*), *Permen* number 24 about the implementation of *Permen* number 22 and 23. The implementation and accomplishment of these *Permen* have basically become the basis of conducting the teaching learning at school.

The issuance of *Permen* on content standard and standard competence logically requires the issuance of policy which regulates the evaluation of the attainment of the content standard. The Indonesian Republic Act number 20 year 2003 about the System of National Education, chapter XVI article 57 verse 1 states that evaluation is held to control the quality of national education and it

also serves as the accountability of an educational institution in conducting the educational process conducted in the institution.

The Minister of Education, then, through the Decree number 20/2007 about Standard of Evaluation states that there are several ways of measuring students' achievement of the content standard. They are daily quizzes/tests, mid semester tests, final semester tests, grade promotion tests (*ujian kenaikan kelas*), school examinations, and *Ujian Nasional* (National Examination or *UN*). The different term of the exams administered to *Sekolah Bertaraf Internasional* or International Standard School is *UASBN (Ujian Akhir Sekolah Berstandar Nasional)*.

The daily tests are held periodically to measure students' achievement after they finish studying one or more basic competence (*Kompetensi Dasar*). The midterm semester tests are carried out after 8-to-9-week-learning activities to measure the students' achievement toward all the basic competencies and the indicators covered within that studying period. To assess the result of a whole semester's learning achievement, final semester tests are used. Next, the grade promotion tests are held at the end of even semester to gauge the students' achievement in that semester in schools which use package systems. The coverage of the tests includes all the indicators representing the basic competencies covered in that semester. The school/*Madrasah* final examinations evaluate the students' final achievement in the subjects whose final achievement is not measured by *UN*, such as the subjects within the religion, ethics, personality, and citizenship groups. The exams also cover the cognitive and psychomotor domain of those subjects. The last, *UN* is conducted by the government to measure the achievement of the national standard of education. The subjects tested are those which belong to the science and technology group. Among the five ways of measuring students' learning achievement, the last one has induced unending controversies among our community.

This paper discusses *UN* as a high stakes tests, people's opinion about *UN*, what research says about high stakes tests, our stance toward the issue of *UN* discussing whether *UN* should be continued or aborted.

## UN AS HIGH STAKES TEST

In the pursuit to increase the quality of education in Indonesia, the government has implemented the *UN* in junior and senior high schools since 2002/2003. According to the *Permen* no. 20 /2007 about Standard of Evaluation, scores from *UN* are used to serve several purposes. Some of which are as follows.

First, on the school level, the scores are used as one of the criteria to determine students' graduation from junior and senior high schools. The government issues different passing level every year, and it usually gets higher for the next examination in the following year. Through *Permen*, based on the recommendation of *BSNP*, a state body under the coordination of Ministry of Education, the passing level of *UN* has been raised up year to year. In 2007, the passing level was 5.00. In 2008, it was 5.25 and in 2009 it becomes 5.50. Students as the participants of the exams (examinees) will be given *Surat Keterangan Hasil Ujian Nasional (SKHUN)*, a document which states the scores that the students obtain from their *UN*. This document is issued by the exam administer or the students' school.

Second, the scores, as stated in the *SKHUN*, are also used for school admission purpose to the next level of education, such as from junior to senior high school. Third, on the level of government policy making, scores from *UN* serve as data based on which the government can develop a map of schools' abilities in meeting national standards of education. Fourth, the data are also used as a basis to issue the policy on providing needed support and resources to relevant schools.

Considering the decisive consequences following the result of *UN*, especially to the students, *UN* can be classified as a high-stakes test. A test can be categorized as high-stakes test when the result of the test bears decisive consequences to the test takers, such as grade promotion or retention, passing or failing, being denied or admitted entry to the next level of schooling, etc. The determinant consequences are the 'high stakes' put on the test result. In the case of *UN*, scores from *UN* have determinant effect of deciding whether students fail or pass from their 3-year study in junior high or senior high school and flexibility in choosing a favorite school for

their next level of education. This high-stakes side of *UN*, in fact, is the aspect that has made *UN* an unending controversial issue.

### ***UN*: HIGH STAKES AND STANDARDIZED TEST**

High-stakes tests can take any form of assessment, such as essay type tests, open-ended questions, performance tests, a hands-on activity, etc. For large-scale test involving big number of test takers a high-stakes test usually take the convenient form for test administrators: a standardized multiple choice test.

According to Wikipedia (2007) high-stakes tests are normally given as [standardized tests](#) (all examinees take the test under reasonably equal conditions) to afford all examinees a fair and equal opportunity to pass. Hopkins, Stanley, and Hopkins (1990) states that standardized tests impose certain controls on testing conditions. The conditions under which the tests are taken are to be standard for all examinees. The directions, time limits, and other controls of standardized tests are prescribed.

The result of the standardized test, in which standard questions are given under uniform conditions, can provide meaningful basis for the government and those concerned with education to evaluate and compare the educational performance of students and schools.

*UN* can probably be categorized as standardized test because uniform questions are given to all students regardless of the local educational process, problems, and uniqueness. All students across provinces take *UN* under the same condition, directions, and time limit

High stake tests, however, should use best standardized tests. Hopkins, Stanley, and Hopkins (1990) assert that best standardized tests are carefully developed and refined by means of editorial writing and item analysis so that virtually every item functions appropriately. Most intrinsic ambiguity is removed, and implausible distracter are deleted. Hopkins, Stanley, and Hopkin (ibid) further explains that the reliability of standardized test should be greater than that of teacher-made tests. Concerning with *UN*, there has been no information about the degree of reliability of *UN* that can be obtained.

However, the drawback of a single standardized test in terms of its reliability can be compensated with teacher-made tests. Hopkins, Stanley, and Hopkins (1990) state that a number of and frequency of teacher-made test can easily compensate for the lack of high reliability in a single test. In the case of *UN*, therefore, the *UN* as the standardized achievement tests are partners rather than competitors. The two tests serve somewhat different purposes and provide complementary information. Both kinds of tests are needed for an adequate evaluation of education achievement of individual student, schools, and provinces and the country.

## THE CONTROVERSIES

Since its first implementation in 2002/2003 (Irawan, 2007), *UN* has inflicted continuing controversies over the effect of *UN*. People have expressed disagreements and criticism more than agreements. Objections against *UN* are usually triggered by the role of *UN* score which serves as the only or more determinant factor in determining students' passing or failing.

People argue that the harm induced by *UN* outweighs the benefit. *Koalisi Pendidikan Bandung* (2008) found that *UN* consumed a lot of school financial budget. On the average, junior and senior high schools had to spend from 17 millions rupiahs up to 70 millions rupiah to prepare the students facing the *UN*. *UN* is also accused of disregarding the development of the character building of whole person focusing too much on behavioristic cognitive goal (Maliki, 2009), and confusing administrative goal with educational one (Alam, 2008).

So far the arguments, however, are more frequently based on personal opinions instead of being based on solid data resulted from accountable studies. Comments about the (negative) effect of *UN* are usually very general not specifically directed to the teaching of certain subjects, such as English so that the comments contribute very little to the improvement of both national examination and the quality of the teaching and learning of certain subject.

A surprising statement was released by Minister of Education predicting that the result of 2009 *UN*, in terms of students who fail the exams will increase. He added that *UN* should not be only a measurement of students' competence but also that of all parties involved in the exams. This claim was supported by Commission X

of People Representatives that the number of students fail the exams will rise up 12 % of the total number of examinee (the number of students failed in 2008 *UN* was 10 %). It was suspected that the so far *UNs* were carried out honestly. It indicates that *UN* still retains problem that it does not convince us as a standard indicator of education.

### **EFFECT OF *UN***

In the context of teaching and learning, influence exerted by external tests such as *UN* on teaching and learning process is frequently referred to as washback effect. Messick (1996) in Brown (2000) defined washback as ". . . the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning". Brown (ibid) himself synthesizing the definitions of washback concludes that washback clearly has to do with the effect of external testing on teaching and learning process in language classrooms.

Different from almost completely negative comments about the effect of *UN*, the definitions of washback quoted above implies the neutral aspect of washback: that washback has the potential power to produce negative or positive influence toward teaching and learning. Brown (2000), for instance, organizing information from many sources, proposes 4 general strategies which expectedly can turn the influence of external tests, like *UN* in this case, into positive washback. Brown argues that test design strategies, test content strategies, logistical strategies, and interpretation strategies can help tests produce positive washback.

### **LEARNING FROM WASHINGTON**

#### **High Stakes Test and Education Reform**

Washington's education reform, which was adopted by the state legislature in 1993, was designed to affect both school and classroom practices. Washington's system includes state-wide standards for what students should know and be able to do—called the Essential Academic Learning Requirements (EALRs); tests to evaluate student knowledge and progress toward standards—called

the Washington Assessment of Student Learning (WASL); and a mechanism to hold schools accountable for student performance. The WASL test in writing consists of two writing prompts of different genres. Each prompt is scored using two WASL-specific scoring rubrics, one that emphasizes content, organization and style, and one that emphasizes conventions.

After EARL and WASL were implemented, a survey to see the influence of EARL and WASL was conducted. Teachers reported changes in their allocation of time to writing, the emphasis they placed on specific aspects of writing, their teaching methods, and their students' learning activities.

Both EARL and WASL were clearly influential, although there is some evidence that more teachers focused on the WASL content and format than the EARLs. Explicit test preparation for the writing exam (e.g., using released items from previous tests) was not widespread. However, a focus on tested content and format was evident in teachers' reports of classroom practice.

To the extent that the tests broadly represent the domain of writing and the scoring rubrics broadly reflect the characteristics of effective written communication, a focus on the tests should not be substantially different than a focus on the standards. The WASL test in writing achieves these goals more than a multiple-choice test of writing would do, because students must produce an essay, not merely fill in blanks, identify mistakes, or complete other writing related tasks that can be assessed using a multiple-choice format.

Given the limited amount of class time available and the large number and breadth of the content standards, the teacher focused more on the tested subject: writing. According to the teachers, replacing or supplementing multiple-choice tests with more performance-based assessments has led to a dramatic increase in the amount of writing students do in school—both as part of language arts instruction and as part of instruction in other subjects.

The most dramatic finding of the survey is the reallocation of instructional time from non tested subjects to tested subjects. This is strong evidence that the tests are driving change more than the standards. Washington adopted standards in eight content areas, but the survey shows increases in time for only those subjects that are tested.

The result of this survey, thus, gives us an example how an educational reform was attained through the implementation of statewide standards and state exams.

### **UN: CONTINUED OR ABORTED?**

It is actually not easy to take a position whether to continue or to stop the implementation of UN. High stakes decision should be based on solid data taken from accountable studies. There should be studies investigating whether UN has resulted in intended good washback or unintended negative one. So far, we have not been able to obtain any information related to the washback of UN and the government measures following the result of UN. However, in the absence of such important basis, we recommend that UN be continued for the following reasons or under the following considerations.

High stakes standardized tests have the potentials to drive teachers to teach according to what is going to be tested like the example of Washington reform. Making use of the potential, the government should develop UN carefully so that all the test items can function as expected, such as forcing teachers to teach according to the standard competence. By launching good UN, the government can carry out reform agenda.

In the first purpose, scores from UN play a decisive role in determining students' passing or failing. In this context, final decision should be left to the teachers. UN is administered only once and the UN as a high stakes standardized achievement test may lack of reliability in evaluating the students' true achievement. This possible lack of reliability needs to be compensated with the teachers' frequent assessments which according to Hopkins, Stanley, and Hopkins (1991) may have higher degree of reliability than one single standardized test. In addition to that, leaving room for teachers and schools to make decision is in line with the decentralization policy which gives more autonomy to schools, which is popular under the term: school based management.

After implementing UN several times, the government should have had the overall picture of the national standard achievement of schools throughout Indonesia. The government should have been able to identify schools which repeatedly underperform.



Studies should be conducted to investigate why certain schools repeatedly underperform and provide needed resources accordingly.

## REFERENCES

- Alam, S. 2008. *Instrumen Ujian Nasional sebagai Penentu Kelulusan Berpotensi Merugikan Siswa*. (on line) (<http://www.duniaesai.com/pendidikan/>) (retrieved on December 2008)
- Brown, James Dean. 2000. University Entrance Examinations: Strategies for Creating Positive Washback on English language teaching in Japan. *Shiken: JALT Testing & Evaluation SIG Newsletter*. Vol. 3 No. 2 Jan. 2000 (p. 4 - 8). (online) ([www.jalt.org/test/bro\\_5.htm](http://www.jalt.org/test/bro_5.htm)) (retrieved on December 15, 2008).
- Hopkins, D. Kenneth., Julian C Stanley and B.R. Hopkins. 1990. *Educational and Psychological Measurement and Evaluation*. Needham Heights, Massachusetts: Allys and Bacon.
- Irawan, Ade. 2007. Menimbang Ulang Kebijakan Ujian Nasional. *Koran Tempo, 17 April 2007*. (online) ([www.unisosdem.org](http://www.unisosdem.org)) (accessed on December 8, 2008).
- Koalisi Pendidikan Bandung. *50 Juta Rupiah per Sekolah untuk Persiapan UN*. (online) (Kompas.com.) (accessed on December 8, 2008.)
- Komunitas Pendidikan Papua. *5.50 Standar Ujian Nasional 2009*. 21 Desember 2008. (online) (Papuapos.com.) (accessed on May 27, 2009)
- Maliki, Z. 2009. Kenaikan Standar Kelulusan Unas. *Jawa Pos, Sabtu 3 Januari 2009*.