

## The Implementation of English School Examination of Speaking and Writing Skills in The Upper Secondary Level of Education in Bandung, West Java

Andang Saehu<sup>1)</sup>, Utami Widiawati<sup>2)</sup>, Johannes A. Prayogo<sup>2)</sup>, Effendi Kadarisman<sup>2)</sup>

<sup>1)</sup>UIN Sunan Gunung Djati Bandung

<sup>2)</sup>Pendidikan Bahasa Inggris-Universitas Negeri Malang

Jl.Pahlawan No.05 Sukagalih, Tarogong Kidul, Garut. E-mail: andangsaeu@gmail.com

**Abstract:** This study aims at investigating the ways the English teachers develop the test prompts of and the ways they administer the English *US* of speaking and writing skills. This study uses the principles of quantitative and qualitative research designs. Utilizing these principles, this present study involved 73 SMAs (30% of 243 SMAs) which were selected using a stratified random sampling technique, they were 41 SMAs from Bandung municipality and 32 SMAs from Bandung regency. Data were collected using questionnaire, interview guide, and document. The findings showed that all researched schools in Bandung, implemented the English *US* of speaking and writing skills. The administrative aspects, were successfully administered. However, the technical aspects, were not administered appropriately. It is due to the fact that the majority of English teachers did not involve any experts and students for the validation and the try-out of the test prompt.

**Key Words:** English School Examination, test prompt, scoring rubric, score interpretation.

**Abstrak:** Penelitian ini bertujuan untuk meneliti guru Bahasa Inggris dalam mengembangkan soal ujian dan cara guru Bahasa Inggris dalam menjalankan Bahasa Inggris *US* dalam kemampuan berbicara dan menulis. Penelitian ini menggunakan desain kuantitatif dan kualitatif. Penelitian ini melibatkan 73 SMA (30% dari total 243 SMA) yang dipilih menggunakan teknik *stratified random sampling* dan terpilih 41 SMA dari Kota Bandung dan 32 SMA dari Kabupaten Bandung. Data dalam penelitian ini dikumpulkan menggunakan kuisisioner, interview, dan dokumentasi. Hasil penelitian ini menunjukkan bahwa seluruh sekolah di Bandung menggunakan Bahasa Inggris *US* dalam kemampuan berbicara dan menulis. Aspek administrasi dijalankan secara sukses. Namun, untuk aspek teknis masih belum dijalankan dengan baik. Hal ini dikarenakan guru Bahasa Inggris tidak melibatkan ahli dan siswa untuk validasi soal *try-out*.

**Kata kunci:** ujian sekolah Bahasa Inggris, soal ujian, rubrik penilaian, interpretasi penilaian

The government's effort in controlling the teaching and learning of English has been undertaken through the curriculums of English, as a guideline for the teachers, since several years ago. The teaching of English in Indonesia, seen from the history of the development of the curriculum of English for the lower and upper secondary levels of education, dates back from the national independence (Suyanto, 2000). Although the teaching of English has been controlled by a curriculum, the government seems to let English teachers choose the approach and the method of teaching which best suit their students'

needs. In addition to controlling how English is taught, the government also monitors to what extent the teaching of English achieves the objectives of learning as stated in the 2006 Standard of Content. Thus, to identify whether or not the teaching of English reaches its objectives, the government seems to control it through the administration of high-stakes tests (Mardiani, 2010). High stakes tests are those that are considered to be tests used to influence or determine life-changing opportunities for individuals.

In the academic context, tests are used specifically for such purposes as determining whether

or not students will be allowed to study at the college level, permitted to graduate, awarded a scholarship, or even considered worthy of a particular distinction related to language skills (Shohamy, 1996 and McNamara, 2000). Supporting Shohamy (1996) and McNamara (2000), Cheng (2004:320) and Cheng, Andrews, and Yu (2011) state that tests will arguably have a greater influence on teaching and learning in a high-stakes situation, i.e. one in which the test is typically used “to compare and rank individuals, schools, or national systems.” One of high-stakes tests which has been administered in Indonesia since academic year 2003/2004 up to now is National Examination (henceforth *UN*). Another high-stakes test, which has been currently administered in Indonesia since academic year 2010/2011 up to now is School Examination (henceforth *US*). Furthermore, since the Academic Year 2010/2011, the government have let 40% of students’ graduation from a school level to be determined by the scores of *US*. English, especially its speaking and writing skills, is one of the subjects tested in the *US*.

*US* or commonly known as School-Based Assessment (SBA), as the name suggests, is to be prepared and administered by the schools. The decree of Government Regulation of the Minister of Education and Culture No. 59 of 2011 about the Passing Criteria of Students Graduating from Certain Level of Education Unit and the Management of School Examination and National Examination, Article 1, Item 3 defined that “*US* is an activity to measure and evaluate the students’ competence conducted by the scholars for all subject in science and technology fields.” In the context of this study, *US*, especially English, is administered to measure the students’ language abilities in speaking and writing skills. Administering the test means giving the test to a group of individuals (Bachman and Palmer, 1996). As the Standard of Operating Procedure (SOP) and Table of Specification for English *US* of speaking and writing skills are constructed by the schools, they are different from schools to school in the administration of *US* for English. In other words, every school will develop the test prompts of English *US* of speaking and writing and administer the English *US* of speaking and writing differently. This condition led some scholars to worry about whether or not the English *US* of speaking and writing skills which has now been decided to be one of the criteria to determine the students’ graduation is administered under a good design of test prompt.

Since the schools have had 40% share to let their students pass, it seems hard for them to avoid subjectivity in scoring the students’ English *US* of speaking and writing skills as the design of test prompt of *US* is in the form of a subjective test. In line with this, Brindley (2001) states that subjectivity is an issue with scoring constructed-response items because scoring relies on judgments made by human scorers. Thus, this study is worth conducting to gain the information about the development of which it is divided into development of a test prompt and development of scoring rubrics used in the English *US* of speaking and writing skills. The expected information is whether or not the development of test prompts and scoring rubrics are developed based on the consideration of: criteria set up for test developers, experts’ involvement for the sake of test prompts and scoring rubrics validation, and students’ involvement for the try-out purposes. If the collected information meets what has just been considered, the quality of development of test prompts and scoring rubrics might be high.

It is also worth conducting to gain the information about the administration of the English *US* of speaking and writing skills. The information is to find whether or not the principals or vice principals responsible for implementing English *US* of speaking and writing skills set up the criteria for proctors’ involvement of the English *US* and assign professionally the 10<sup>th</sup> and 11<sup>th</sup> graders to be external examiners. The information collected is expected to provide English teachers with a tremendous portrait of what is really going on in the implementation of the English *US* of speaking and writing skills. Further, the portrait of its implementation is expected to contribute to the local government (*Depdiknas Kota Bandung and Depdiknas Kabupaten Bandung*), principals, and English teachers as data based information regarding the implementation of the English *US* of speaking and writing skills. Hence, this study attempts to investigate how English teachers developed the test prompts of English *US* of speaking and writing skills and how English teachers administer the English *US* of speaking and writing skills.

## METHOD

The present study employed quantitative and qualitative approaches. This is in line with the statement by Creswell (2012) that with this

explanatory mixed methods as the research design, the present study aims at answering the research questions: how English teachers develop the test prompt of English *US* of speaking and writing skills and how English teachers administer the English *US* of speaking and writing skills.

Those approaches were utilized for collecting and analyzing the quantitative and the qualitative data and sequentially carried out in two phases. First, this present study carried out a survey in order to gather the quantitative data regarding the percentage of English teachers developing the test prompts of English *US* of speaking and writing skills and administering the English *US* of speaking and writing skills. During this investigation, the speaking and the writing test prompts and the speaking and writing examination administration were among variables that underlie the construct of the implementation of English *US* of speaking and writing skills. While the first variable was related with the technical aspect of English *US* with the 12<sup>th</sup> grader English teachers becoming the target of the investigation, the next variable dealt with the administrative aspect of *US*. Second, this present study also employed a collective case study which was carried out in order to collect and analyze qualitative data about development of the test prompts of English *US* of speaking and writing skills and the administration of the English *US* of speaking and writing skills. Through this collective case study, this study was investigating all researched schools by using multiple data collection techniques (Creswell, 2012) such as interview guide and documents.

The settings of this study were *SMAs* in Bandung, West Java. The choice of West Java was based on the researcher's expectation. The results of this research were expected to improve the quality of English teachers in this province in developing the test prompts used in the English *US* of speaking and writing skills and in administering the English *US* of speaking and writing skills. The population of the study was 243 *SMAs*, consisting of 136 *SMAs* in Bandung municipality and 107 *SMAs* in Bandung regency. While of 136 *SMAs* in Bandung municipality, 27 *SMAs* are state schools and 109 are private ones; of 107 *SMAs* in Bandung regency 19 are state *SMAs* and 88 private ones. Dealing with the selection of the samples—which schools and teachers, the stratified random sampling technique was employed under the desired strata (Creswell, 2012:144) in order that the description of the

characteristics of the population becomes more plausible. In working with the stratified random sampling technique, the samples of school subjects were initially drawn proportionally from each stratum randomly of which the desired sample size was determined to be 41% schools. The result of calculation indicated that there were 100 school subjects selected as the sample of the study consisting of 19 state schools (taken from 11 schools in Bandung municipality and 8 schools in Bandung regency) and 81 private schools (taken from 45 schools in Bandung municipality and 36 schools in Bandung regency, West Java province).

Based on the information collected from *MGMP of Kota and Kabupaten Bandung*, of 100 schools in Bandung municipality and regency, about 300 English teachers were assigned to teach the third grade-students. Initially, those 300 English teachers were selected as the sample of study, but there was no authority to force them all to fill in the questionnaire. The fact is in the process of distributing 300 questionnaires to 100 different schools in Bandung (50 schools of Bandung regency and 50 schools of Bandung regency) of which each school got three questionnaires as the number of English teachers teaching the third grader of students varied, there were only 73 English teachers sent back the questionnaires. Interestingly, those 73 English teachers represented 73 schools from different areas because every an English teacher of a school only filled and sent back one questionnaire. For the sake of interview, seven English teachers filling in the questionnaire sheet unclearly were selected purposively as the key informant. The selected English teachers were taken from one English teacher of state school in Bandung municipality headed by *MGMP*, two English teachers of state schools Bandung regency, two English teachers of private schools in Bandung municipality and two English teachers of private schools in Bandung regency.

To answer the research questions about how English teachers developed the test prompt of English *US* of speaking and writing skills and how English teachers administered the English *US* of speaking and writing skills, the questionnaire, interview guide, and document as the research instruments were employed. The process of collecting data through questionnaire conducted step by step as the area of the schools varied started from May 1 to 31, 2013. There were 37 of 50 schools locating far from the

town especially in Bandung regency were the priority to send the questionnaires. Distributing questionnaires to those schools served 10 days (from May 6 to 15, 2013), but there were only 19 schools sent back the questionnaires. The number of questionnaires gained from 19 schools were 19 questionnaires as every one school sent back one questionnaire. This condition was the same as those of schools (13 of 50 schools) in Bandung regency locating near from town of which they sent back 13 questionnaires. So, the number of English teachers filling in questionnaires in Bandung regency was 32 English teachers with 32 questionnaires. By combining the number of English teachers filling in questionnaires in Bandung regency and municipality, it is known that the number of English teachers sending back the questionnaires was 73 English teachers from 73 schools.

As a supporting instrument to check or confirm unclear answers gained from questionnaire, interview guide was employed. The interview was conducted from July 1 to 30, 2013 after checking the answers of collected questionnaires whether or not the responses are clear enough. It was found that there were so many unclear answers. Knowing this, seven questionnaire sheets showing unclear answers were then selected. Those seven selected questionnaire sheets refer to seven subjects of which they were selected based on the accessibility and information rich. Being accessible implies that the key informants for this study are the English teachers not coming far from the town. Meanwhile, being information rich implies that the key informants were the English teachers who are considered to have much knowledge about the current topic under the study.

Meanwhile, documents which were also used as the research instrument refer to ready-made sources of data, easily accessible to the imaginative and resourceful investigator as to comply with. The kinds of documents relating to this study were referred to the research questions. From the research questions, it is known that the documents needed were blueprints of test prompt of English *US* of speaking and writing skills, test prompt of English *US* of speaking and writing skills, and SOP. Ideally the documents collected should be 73 for each document as the sample of the study was 73 schools. Unfortunately, the collected documents did not meet the ideal expectation of the researcher. They consisted of five blueprints of test prompt of the

English *US* speaking and writing skills, 15 test prompts of the English *US* of speaking skills, 20 test prompts of the English *US* of writing skills, and 50 SOPs.

The data gained through questionnaires were analyzed by: (1) checking the identification of the subjects and by checking the data; (2) classifying the data based on the characteristics of the variables taken from the questionnaire, named coding; and (3) interpreting differently depending on the items asked in the questionnaire. In analyzing interview data, the steps were: (1) transcribing the data based on the recordings of the interview; (2) organizing the transcribed data based on the theme; and (3) drawing a conclusion. The analysis to documents was performed by justifying them with the explanation of the English teachers in describing the ways they developed the test blueprints and test prompts of the English *US* of speaking and writing skills. For example, when an English teacher explained about development of test prompts, their explanation was justified with the existing test prompts s/he gave. So these are to support the main research instrument and research finding.

## RESULT

### Development of Test Prompts of English *US* of Speaking and Writing Skills

Based on the results of questionnaire gained from 73 English teachers and interview from seven English teachers, commonly all schools (English teachers) involved in this study developed the test prompts of speaking and writing skills through two steps and conducted in group; the first step was designing blueprints of test prompts of speaking and writing skills and the second step was developing test prompts of speaking and writing skills.

In working with the first step, although the data questionnaire showed that more than half of the English teachers (53% or 41) seemed not to design the test blueprints, actually when they were confirmed through interview, they said that the blueprints were provided by forum of English teachers (*MGMP*). So, the information is true that those 41 English teachers individually did not design the blueprints as everything was available in *MGMP* of Bandung municipality. The questionnaire also showed that the other English teachers, whose blueprints were not designed by *MGMP*, also designed the blueprints

together with their colleagues. Meanwhile, the rests only some (15% or 11) designed personally the blueprints of speaking and writing skills. This finding implies that the test developers were *MGMP* and individual English teachers. Personally or individually means a school or an English teacher—did not work together with other schools or other English teachers in developing the test prompts of English *US* of speaking and writing skills. The blueprints were designed by analyzing the competency tested, creating indicators of the test prompts of speaking and writing skills, validating the blueprints of test prompts of speaking and writing skills to other people, finding the information collected in the validation form of blueprints, and analyzing the information collected from the validation form of blueprints.

To know how the English teachers analyzed the competency tested, *MGMP Kota Bandung* and six English teachers as their answers to the questionnaire were not clear enough were interviewed. All of them showed the same opinions that they analyzed it by breaking down the Standards of Competence (*SK*) and the Basic Competence (*KD*) whether or not they were taken from School-based Curriculum and whether or not they were appropriate for the students' school level (12<sup>th</sup> grader). While *MGMP Kota Bandung* used all *SK* and *KD* in developing the blueprints of speaking and writing skills, other English teachers commonly selected the *SK* and *KD* of even semester. When they broke down the *SK* and *KD*, they paid attention to the two competencies tested. Having analysed the competency tested, the English teachers also considered time management and topic selection in the process of test blueprint development. The time management of English *US* of speaking is different from that of English *US* of writing skill. The majority of English teachers (78% or 56) decided the length of time for speaking skill based on number of students in a class and types of speaking tasks. There were many English teachers commonly saying that they spent almost a day to examine the students' speaking skill as the number of students were about 35 – 40 students for a class of which a school may have more than one class.

Based on the questionnaire and interview data, it was known that some of the English teachers (14% or 10) set about five minutes for retelling the story or narrating their hobbies, dreams, or unforgettable moments in front of the class. Some others (8% or 6) spent the time differently about 10 to 15 minutes for one group of students for doing

conversation or dialogue, while some others spent 30 minutes for a group for role-playing in a drama. Overall, the time needed for conducting the English *US* of speaking skill is around five minutes to 10 minutes for a student or from 30 minutes to one hour for a group. Meanwhile, all English teachers (100%) had the same length of time for writing examination, namely around two hours or from 60 minutes to 120 minutes. The topics of English examination of speaking and writing skills were selected by the test developers of which the selection was conducted based on the current issues that are being discussed, such as presidential election, public demonstration, the effect of using internet, national examination, folktale, plan after graduation, learning experience, family, friends, etc. The next consideration is about the number of items or prompts designed by the test developers. The decision of the total number of items varied among the schools starting from one to three items. Those deciding three items for writing examination do not mean that the students were asked to response all items, but they were given a choice to finish one of the items.

The questionnaire data also indicated that all English teachers also set up the indicators of the test prompts of speaking and writing skills. Both *MGMP* and other English teachers have the same ideas in designing the blueprints. The ways they developed the indicators of the test prompts: 1) selecting one of the *SKs* and *KDs* of speaking and writing skills; 2) determining the topics relating to the selected *SK* and *KD*—if the *SK* and *KD* of speaking ask to express meaning in short and monologue functional of narrative text, the topic should relate to narrative texts; 3) setting up the indicators using one of two models either the model presenting the condition at the first sentence or putting the audience at the first sentence.

Based on the documents (test blueprints) collected, the majority of English teachers (90% or 66) in Bandung set up the indicator using the second model of which they did not use a sentence, a paragraph, a picture, a graph, etc. as the indicator of test prompt. But, they asked directly the students to do something. As shown in one of the following data.

Indicator : The students are able to write main idea of business formal letter by:

- using accurate grammar, vocabulary, and spelling
- writing the point of letter accurately

- writing the order of information acceptably-writing a well
- ordered and easy to understand the letter using formal language acceptably.

Meanwhile, the minority of English teacher (10% or 7) used the first model of setting up the indicator of test prompt, namely by presenting the condition at the first sentence. The term condition, in the first model, means one or more operational words used in the first sentence as shown below.

Indicator: Given a prompt, students are able to:

- memorize several expressions of proposing something
- express their opinions using several expressions of proposing something
- demonstrate the use of several expressions of proposing something

The next step of test development was developing the test prompts of English *US* of speaking and writing skills. This second step was different from that of the first one of which the 41 English teachers did not personally design the blueprints. In this step, all English teachers individually developed the test prompts based on the designed blueprints. The interviews were then conducted to seven English teachers emphasizing that the test prompts were developed by the English teachers assigned by the school principal or vice principal. Those assigned by the principal were English teachers commonly teaching 12<sup>th</sup> graders and some English teachers teaching the 10<sup>th</sup> or 11<sup>th</sup> graders.

There is a need to see whether or not the test prompts are suitable with the test blueprints. What did they do to develop the test prompts of speaking and writing skills in the English *US*? The answers to the question above were shown in the questionnaire sheets that although the minority of the English teachers (15% or 11 of 73) left the sheet blank—only answered “Yes” without giving any information about the ways they developed the test prompts, the majority of them (85% or 62) developed the test prompts by way of analyzing the blueprints they designed. One of the components they analyzed from the blueprints was the indicators of test. In the process of analyzing indicators of test, their answers showed the same ideas that the indicators of test

prompts should have at least one measurable operational verb and represent the material tested. In all, the test prompts they developed were pursuant to the test indicators they set in the blueprints.

The next step they did after analyzing indicators was developing test prompts of English *US* of speaking and writing skills. As the evidence of test prompt developments, there is a need to show, at least, two samples of test prompts developed by the English teachers as follows.

A. Students must express Monologue in the form of recount text.

Procedures Test

1. Students collect recount text
2. Students tell story of recount text based their own language
3. Students answer question from examiner

B. You will read a dialogue between Budi and a teacher about a celebration of Independence Day. Read the dialogue carefully.

Budi: Iwan, our school will celebrate the Independence Day in grades tenth through twelfth.

Teacher : It's good news. When will it be?

Etc.

Based on the dialogue above, suppose you were Budi and you have to announce the event to the students from grade ten to twelve. Now, write the announcement to the students. Make sure that your announcement is clear, brief, and accurate. (50-100 words).

The former (prompt A), which has many grammatical errors found in the prompt, was the test prompt developed by one of private *SMAs* in Bandung municipality and the latter (prompt B) is by one of private *SMAs* in Bandung regency. The grammatical errors found in the prompt A can be seen in the use of ‘procedures test’. It should be ‘procedure of test’ as there is only one procedure should be followed by the students. Other grammatical errors are there are no articles used in front of ‘story of recount, question, and examiner’. They should be ‘a story of recount, a question, and the examiner.’ Another grammatical error found is in the word ‘based’ which should be ‘based on.’ This occurs because the English teacher did not review his/her prompt to experts in test development. The brief thing in distinguishing those test prompts is by looking at the clarity of test prompts as the examinees need to know what to do during the speaking and writing examinations. One

out of the two test prompts was clearer. In the above cases, prompt B by a private school in Bandung regency was clearer as it tells students what they have to write and where they have to write it whereas prompt A by a private school in Bandung municipality does not tell students what exactly they must do to complete the test or even whether the prompt asks students to write or to speak.

To create the quality of the test prompts of speaking and writing skills in the English *US*, the minority of English teachers (32% or 24) validated the test prompts to the experts in test development, to their colleagues teaching in different classes (Class X and XI), to their colleagues teaching in different schools, to postgraduate students, and to school principals or vice principals by ways of using validation form or using the draft of test prompts. Within the validation form, the experts could give their score or comment for every statement relating to test prompt of speaking and writing skills. The scores used were ranged from 1 to 5 whose meaning is 1 for poor, 2 for fair, 3 for good, 4 for very good, and 5 for excellent. Meanwhile, those sending full draft of test prompt to experts got some information or comments stating the quality of test prompt. Meanwhile, those who did not identify the information said that they could not identify the information collected in validation form of test prompts of speaking and writing skills as the information was perfectly written—the validators stated that the test prompts were good or because the validators did not give any comments on the test prompts.

Another way to create the quality of the test prompts was trying out the test prompts. In relation to trying out the test prompt, one of the questionnaire items asked the English teachers if they conducted the try-out of the test prompt of English *US* of speaking and writing. The primary objective of the try-out was to collect evidence if the test prompt has already been properly developed. The data gained from the try-out served as the valuable information for the test developer to do some further revision. Unfortunately, the finding showed that although the experts or validators suggested the English teachers to try-out the test prompts, the majority of English teachers (90% or 66) did not try-out the test prompt. Based on the interview results, the reason the English teachers did not try-out the test prompt sound the same that the test prompt have been developed favorably and validated to the experts. Interestingly, there were some English

teachers (10% or 7) saying that trying-out the test prompt was time consuming as they were sure that their test prompts were eligible to be tested.

### **Administration of the English School Examination of Speaking and Writing Skills**

Based on the questionnaire data, in the process of administering English *US* of speaking and writing skills, all English teachers administratively knew that the preparation of the test environment (arrangement of time, room, and seat for the examination) has been described in Standards of Operating procedures (SOP) of *US*. The SOP shows that the time (schedule), room, and seat for school examination should be arranged by individual schools. Therefore, all English teachers conducted one month before National Examination. The exact time conducting the English *US* of speaking and writing skills was in March 2013, starting from the first week to the second week of the month, yet there were also schools which had it at the end of February 2013. Others varied from a month, two weeks and a week before the National Examination. The questionnaire data also showed that in deciding how long the test should be, the English teachers did not find any difficulty in doing it. All English teachers decided the length of time for speaking skill differently starting from about 10 to 15 minutes for one student for speaking test, 30 minutes or one hour for a group for speaking test. Meanwhile, all English teachers had the same length of time for writing examination, namely around one to two hours or from 60 minutes to 120 minutes.

The questionnaire data also showed that since the writing examination required much attention to avoid cheating among students, some examiners seemed to involve proctors for examining students' writing skill. The involvement of proctors was administratively regulated in SOP of *US*. It is known that the majority of English teachers (87% or 64) involved proctors during the writing examination. In relation to the number of proctors, the confirmation through interview was conducted to those (7 schools) involving proctors in the English *US* of speaking and writing skills. It is known that the total of proctors was depended on the total class of grade XII of a school. The interview was also asking about the role of the proctors in writing examination. Three of seven stakeholders stated that proctors were involved to help the examiners because the writing examiners

were absent. Some others argued that the existence of proctors was to help students avoid cheating among them and they were expected to take the experience of it as someday they will assess their students' writing skill. However, the majority of school principals (82% or 60) did not involve any proctors in speaking examination. To clarify this case, the interview was conducted to ask for the reasons for not involving any proctors in the speaking examination. Two things in common were firstly due to the fact that the examiner (English teacher) could handle and monitor the examinees while examining individual students or in group students. Secondly, the examiners commonly (65% or 47) located the students outside the classroom and called one or two students or a group of students to have the speaking examination in the classroom.

In relation to the question whether or not the English teachers examined the students' speaking and writing skills by themselves, all English teachers said "Yes" to the questions. It means that they themselves examined the students' speaking and writing skills. The next question relating to this was whether or not they have been trained for examining the students' speaking and writing skills. It was known that the minority of English teachers (only 20% or 14 of 73) got training for speaking skill. The training was about how to develop test blueprint of speaking, how to develop test prompt of speaking, and how to test the students' English speaking pronunciation, clarity, and vocabulary. In addition, the interview with some English teachers showed that the majority of school principals assigned English teachers teaching the 10<sup>th</sup> or 11<sup>th</sup> graders (henceforth called external examiners) in the process of examining the students' speaking skill. The involvement of external examiners was employed depending on the school's policy. In certain schools having limited classes (1 or 2 classes) of 12<sup>th</sup> graders, an examiner has to assess all 12<sup>th</sup> graders' speaking skill, while in other schools the team of examiners is formed. In other words, the involvement of external examiners was due to the fact that the numbers of English teachers teaching the 12<sup>th</sup> graders was less than three English teachers, while the numbers of class were more than three classes. They divided themselves testing different departments—Social, Science, and Language Department. An examiner was responsible for at least a number of examinees in one class, which were around 40 students in certain cases, to examine speaking and writing skills with

the same test prompts. They did not sit together with other English teachers in a class in examining the English *US* of speaking and writing skills. Although the external examiners were involved to examine the speaking skill, it is the fact that the most of external examiners confessed that they were not trained to conduct the speaking examination like what the School-Based Curriculum intended to. This fact is then confirmed through interviewing the principals or vice principals as stakeholders. They admitted that the external examiners were not trained as they were more likely to make use of the panel of the examiners to discuss the ways they administered the English *US* of speaking. Beside that, they also admitted that did not set up the criteria for choosing speaking examiners as they had limited number of English teachers.

Both external and internal examiners performed the speaking examination in different tasks. Some examiners asked examinees to become a reporter and sources of news in group. In a group, a student became a reporter and two other students became sources of news. The students were asked to search news from newspapers, magazines, television, and radio to be used as materials for speaking performance. Other examiners asked the examinees to perform a dialogue. The dialogues, which may consist of expressions of congratulation, gratitude, and compliment, should be made by the examinees. In relation to the situation of the expressions, they were given freedom to choose the situation that they liked. Before performing the dialogues in front of the class, they were asked to memorize and practice the dialogues. In this time, the examiners did not check the students' dialogues. Another examiner was performing announcement individually. In this task, every student was asked to make an announcement containing from six to nine sentences. They were given freedom to choose the situation and topic that they wanted in making announcements. Besides, they were also allowed to search the example of announcement from the internet. Finally they had to present their announcement in front of the class because their announcement would be checked. Among the speaking tasks aforementioned, commonly the examiners admit that they preferred examining speaking skills individually by asking the examinees to present the topics suggested in the test prompt, such as retelling the story, telling their experience, telling their plan after graduation, explaining their opinion about an issue, etc. Before

presenting their topics in front of the class, the students were asked to prepare or memorize the topics they wrote. The consideration for preferring this kind of speaking task was time efficiency. According to them, evaluating the speaking skill individually was easier to do and to give the score than evaluating it in groups.

The writing examination conducted by the English teachers was also performed in different tasks. Some English teachers asked the examinees to create a descriptive text, a report text, a narrative text, and other genres. The students were asked to find a topic they like. Other English teachers asked the examinees to write a 100, 200, or 300-word text based on the topic chosen by the examinees. Other English teachers also examined the writing skill based on the speaking skill. As performed by *SMAN I A*, *SMA Swasta C*, and *SMA Swasta D* Bandung municipality examining the speaking skill by reading aloud a text supplied by the teacher, then right after reading the text, the examinees were supposed to do the writing an at least 100-word text. Another school which also examined the writing skill based on the test prompt of speaking skill was *SMAN E* Bandung municipality. The English teacher asked the examinees to write a script of a folktale which was role-played in groups by them in speaking examination. The different writing tasks affect the time or length of the writing examination in the English *US*. For example, the task of writing that asked examinees to write a 150-word narrative text or other text types commonly spent 90 minutes. This length of time is different from the task of writing that asked examinees to write a script of a drama, a dialogue, or a speech. Although the English school examination was conducted in different tasks and time, the English teachers commonly (95% or 69 English teachers) examined the examinee by asking them to write a text in 90 minutes.

## DISCUSSION

The discussion for the development of test prompt is divided into two: The first one is the elaboration of the design of blueprints of speaking and writing skills. The second part is the discussion of the construction of test prompts of speaking and writing skills. Those two parts are in line with the statement by *the Standards* (American Educational Research Association—AERA, 1999) and by Standards of Operating procedures (SOP) that test

development at least covers some processes, namely delineation of the purpose of the test, development of the test specification (also known as blueprint), development of the test items or the domain (content and skills) to be measured, development of scoring guides and procedures, and assembly and evaluation of the test for operational use.

What has been performed by the English teachers aforementioned coincides with the test development stages supposed by Downing (2006), consisting of plan, test administration, and test result report (to mention some). Obviously, Downing (2006) said that the construction of an educational test should include the stage of planning the test of which in this stage the test developer should first decide the objective of test to be achieved, the construction of test to be measured, the contents or materials to be tested, the test specification to be designed, the scoring rubrics to be developed, and the sequence tasks to be accomplished.

The first stage by Harris and McCann (1994) is called planning and construction stage. They said that in the planning stage, the examiner should plan what performance and materials to be tested, while in the construction stage the examiner should construct the test format and scoring procedure. Supporting Downing (2006), Dromerthy (2011) stated that careful planning tends to increase accountability. This statement shows that a planning can help examiners to identify what was done and why. Fortunately, the finding above also coincides with the statement by Suen and McClellan (2003:2) that defining the purpose of the assessment and delineating the scope of the content and skills to be measured should come first in test development, especially in this study in designing the test specification. Supporting Suen and McClellan (2003:2), Welch (2006) stated that the test developer must begin the development process by specifying the purpose of the test and content or construct to be measured. The aforementioned statements indicate that once the purpose of assessment and the scope of content and skill coverage have been determined, the test specification can be developed. In the process of designing the test blueprint, the test developer needs to set up the ‘indicator’ (Dricoll and Wood, 2007:31) of test prompts of speaking and writing skills before assembling the test prompts.

The next step in test development, which according to Downing (2006), is to construct the test prompts that measure the targeted content and skill

areas laid out in the test blueprints, in which in this study refers to the speaking and writing skills. The kinds of claims or inferences that are to be made about the knowledge or skills of interest must be considered in creating the test prompts. Prompts should be designed to give salient evidence to support these claims. The activity of mapping the *SK* and *KD* coincides with Rivera and Collum (2006) stating that item or prompt construction relates to mapping a sentence for measuring a skill. Mapping sentences, in this research, is similar to mapping *SK* and *KD* as it writes variable elements of a skill. The development of test prompts, as suggested by Welch (2006:305), should be brief, simple to understand, and free from possible ambiguities. This advice is seemingly to be intended to help learners as they are coming from different conditions—some of them are fast learners and some others are slow or least skilled ones. This coincides with statement by Baranovcki (2007) that clear instructions on whether or not the examinees are to guess at answers should be accompanied by sufficient examples to ensure that the slow or least skilled examinees understand the problem type. Therefore, if the language of the test prompts is difficult or confusing, one cannot be sure that poor test performance indicates low proficiency in the skills area being tested.

Having designed the test blueprints of speaking and writing skills and developed the test prompts of speaking and writing skills, the test developer should validate the blueprints of test prompts of speaking and writing skills to other people. In relation to validating the test blueprint and test prompts, Weir (2005:1) defined validation as a process of generating evidence to support the well-foundedness of inferences concerning the test specification. Doing validation, which might be in the form of comment, review, or score, to test blueprints is of great importance as it can delineate whether or not the blueprints designed are well designed and prepared. In line with this, Kindler (2002:90) emphasized that the value of external review, commenting that it is important to let others have access to materials and come to their own independent conclusions about the strengths and weaknesses of the program. The similar ideas are given by Weir (2005) and Bowen, Bowen, and Woolley (2007) stating that having planned the test, the test material should be submitted to review and revised on the basis of review. This becomes clear

that once the test specifications and test prompts have been developed, they must undergo a number of reviews for appropriate content and fairness.

Interestingly, the most English teachers involved in this study skipped this step of validation. They confidently distributed the prompts of speaking and writing skills in the English school examination to the examinees without validating the prompts to the experts. By skipping this step, the test prompts administered to examinees might rise a question whether the content is clear or grammatically incorrect. But, for those validating test prompt, they validated it to the experts in assessment, to colleagues teaching in different class or school, and to the people considered to have much knowledge about school-based assessment. Again, interestingly the majority of English teachers skipped the step of trying-out the test prompt. In line with this, Bachman and Palmer (1996:225) mention that the try-out is indeed very crucial. The fact is they confidently examined the students' speaking and writing ability in English *US* without trying out the test prompt. This implies that since the majority of English teachers skipped the try-out phase of which this is the phase that the test developers see if the test design, along with the administration, has met the expected quality, the quality of test prompts they developed is questionable whether or not the test prompts are well developed.

Meanwhile, the discussion for the administration of English *US* of speaking and writing skills involved a variety activities run by the English teachers. These included administrative aspects, such as preparing the testing environment and materials, involving the testing proctors, and testing for collecting information. These activities coincide with statement by Bachman and Palmer (1996:92) that administering a test involves preparing the testing environment, collecting test materials, training examiners, and actually administering the test. The first activity in test administration was preparing the testing environment. It seems to be consistent with the test blueprint of which it involved arranging time, place, and seat for English school examination. This coincides with Bachman and Palmer (1996:232) stating that one of the steps in test administration is preparing the testing environment in which the place of testing, materials and equipment, personnel, time of testing, and physical conditions under which the test is administered are arranged. In this study,

before administering the test, the examination committee have prepared the time, place and seat for English school examination.

In relation to this study, it has been revealed that all *SMAs* in Bandung for academic year 2012/2013 conducted the English *US* of speaking and writing skills with various preparations. As a result, the management of the examination differed in terms of place, time, and seat. For example, several schools in Bandung municipality and regency (West Java) which examined the speaking skill required a place large enough to create two separate testing areas with different characteristics, one suitable for role-play and one suitable for some forms of oral language assessment such as interviews, story retelling, etc. The speaking examination also required much time under the consideration of practicality. Some of the schools may schedule the English *US* of speaking skill during the day with 40 or 60 minute intervals to avoid fatigue and loss of concentration. Meanwhile, several schools before giving out the writing skill examination, they tried to separate students as much as possible, by moving desks or placing students around the room.

Another thing prepared by the schools was test materials. It is of utmost importance to make sure that before administering the English *US*, all the necessary printed materials and that there were adequate supplies of all test papers, maps, charts, or any other printed material for the number of students who will be taking the test were prepared well. It is also of great importance to make sure whether or not photocopying is of a satisfactory quality and that no errors have been made in printing and preparation. The examiners should inform the students before the test begins if there are some errors such as spelling, punctuation, and repeated questions. Besides, the examiners should inform the students that they should know what materials they need to bring with them to the examination, such as pens, pencils, erasers, etc. The next activity in test administration was involving proctors. Before administering the English *US*, test administrators required to involve the proctors as the assistance assisting them. They also needed to think through what kind of assistance they want the proctors to provide during the examination itself (Bachman and Palmer, 1996). In several literatures, there seems many kinds of assistance provided by the proctors during the examination. For example, Crocker (2006) suggested that as the

examinees arrive at the testing room, the proctors should have them take alternate seats within rows or if possible, empty rows should be left between examinees. This suggestion, in relation to this study, has been seemingly applied by several schools in Bandung in the English *US* of writing skill.

A study conducted by Wijaya (2013) focusing on the roles of proctors in proctoring writing test found that a proctor could be a room checker. It means that as soon as the actual test administration begins, the proctors should walk quietly around the room checking to see that examinees are marking their answers in the prescribed manner. In addition to checking the room, McCallin (2006) stated that throughout the testing period the proctors should frequently walk about the room to make sure that the examinees are not cheating or do not copy or help each other. Thus, a proctor needs to reduce to a minimum any cheating that might go in the classroom, for example by telling them that those caught cheating might be given a zero score or even disciplinary action might be taken.

The final activity in test administration was testing for collecting information. In this activity, testing refers to the act of giving an examination to the testees. In the process of giving the examination, it should be clear that whoever the examiners are, the schools are responsible for training and monitoring staffs in the proper use, administration, and scoring of the chosen assessment, which is especially important with performance-based assessment (AERA, 1999:30). Unfortunately, not all schools in Bandung trained their testers of how the writing and the speaking tests should be administered. Although most of them were not trained, they administered the English *US* of speaking and writing skills with various tasks or activities. O'Malley and Pierce (1999:69-89) listed some forms of oral language assessment, including oral interviews, radio broadcasts, video clips, information gap, story/text retelling, investigations/role-plays/simulations, and debates. Some of the types of task mentioned by O'Malley and Pierce (1999) have already been used in the examination of speaking as the chosen ones were mostly in forms of simulated performance, such as making conversation and giving monologues. Meanwhile, the examination of writing skill has been carried out in an item-response format which can be open-ended.

## CONCLUSION AND SUGGESTION

### Conclusion

Based on the research findings and the discussion in the previous chapters, several points can be concluded.

First, the test prompts of English *US* of speaking and writing skills were developed based on the blueprint of the test prompt of speaking and writing skills. To keep the quality of the blueprints and the test prompts, the validation and the try-out require to be run into operation. However, there was almost no validation and try-out activities employed by the English teachers to the test prompts. The act of leaving the stages of validation and try-out could have a high risk on the quality of the test prompt. The main reason for not validating and trying-out the test prompt is time consuming since the English teachers have already had heavy teaching loads.

Second, the administration of English *US* was run into operation by two aspects. While administrative aspects: the arrangement of time, room, seat for examination, and proctors' involvement were successfully run, the technical aspects: the development of the test prompts and the scoring rubrics seemed not to be facilitated with the relevant empowering activities such as trainings, workshops, and seminars on test development. There are even no criteria set for the involvement of external examiners. Hence, since the relevant strengthening programs were not provided and the criteria were not set for the selection of involving external examiners, the English *US* of speaking and writing skills tended to be administered for the sake of the fulfillment of a mandate from the government.

### Suggestion

The benefit of early interests in the English *US* of speaking and writing skills in determining the students' success or failure is still not much. It is recommended that due to the English school examination which is authoritatively conducted by school or education unit has 40% share to determine the students' graduation, the government should investigate how the English school examination is implemented.

To the policy makers (school principal and vice principal) responsible for assigning English teachers, it is recommended that English teachers' assignment not only be based on those teaching English at the

12<sup>th</sup> graders but also on candidates' academic performance. Regarding the English teacher's selection, those to be selected should be ensured to have high performance in English, for example, by showing the TOEFL or IELTS score with certain minimum score, showing the paper publication, and submitting the documentation as the evidence of being a presenter or a participant in a seminar or alike.

To the policy makers (school principal and vice principal) responsible for implementing English *US* of speaking and writing skills, it is recommended that they should help broaden English teachers' knowledge and ability in test development by asking them to join or conduct seminars, workshops, or trainings on how to develop the test blueprint and the test prompt.

To the English teacher forum (MGMP) considered responsible for controlling or monitoring the implementation of English *US* of speaking and writing skills, it is recommended that due to the perceived importance and role of such a body to help English teachers administer the English *US* run well, preparation for test development should be much paid attention by way of discussing how to implement the English *US*.

Finally to future researchers on English *US*, it is recommended that they investigate more specific topics relating to this research study, such as how the English speaking and writing skills are measured. It is also recommended that they conduct similar studies involving bigger number of informants to refine the theories emerging from the current study. Involving bigger number of informants is based on the consideration that the bigger the sample, the less potential error is that the sample will be different from the population (Cresswell, 2012). This way the generalizability and transferability of the findings can be enhanced and the methodological limitation of this study can be addressed accordingly.

## REFERENCES

- American Educational Research Association & National Research Council, 1999. *Standards for Educational and Psychological Testing*. Washington, DC: National Academy Press.
- Bachman, L.F., & Palmer, A.S. 1996. *Language Testing in Practice: Designing and Developing Useful Language Tests*. New York: Oxford University Press.
- Baranovcki, 2007. *Designing Performance Tests*. Boston: Allyn and Bacon.

- Bowen, N.K., Bowen, G.L., & Woolley, M.E. 2007. Constructing and Validating Assessment Tools for School-Based Assessment Practitioners: The Elementary School Success Profile. In A.R. Roberts & K. Yeager (Eds.). *Handbook of Practice-Focused Research and Evaluation*. New York: Oxford University Press.
- Brindley, G. 2001. Outcomes-Based Assessment in Practice: Some Examples and Emerging Insights. *Language Testing* 18: 393.
- Cheng, L. 2004. *Washback in Language Testing Research Contexts and Methods*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Cheng, L., Andrews, S. & Yu, Y. 2011. Impact and Consequences of School-Based Assessment (SBA): Students' and Parents' Views of SBA in Hong Kong in *Language Testing* 28 (2) 221-249. Downloaded from ltj.sagepub.com at OhioLink on May, 2013.
- Cohen, L., Manion, L., & Morrison, K. 2007. *Research Method in Education*. London: Routledge.
- Conner, C. 1991. *Assessment and Testing in the Primary School*. New York: The Palmer
- Crocker, L. Preparing Examinees for Test Taking: Guidelines for Test Developers. In S.M. Downing & T.M. Haladyna (Eds.) *Handbook of Test Development*. London: Lawrence Erlbaum Associates Publishers.
- Creswell, J. 2012. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Boston: Pearson.
- Downing, S.M. 2006. Twelve Steps for Effective Test Development. In S.M. Downing & T.M. Haladyna (Eds.) *Handbook of Test Development*. London: Lawrence Erlbaum Associates Publishers.
- Driscoll, A., & Wood, S. 2007. *Developing Outcomes-Based Assessment for Learner Centered Education*. Danville, Virginia: Stylus Publishing, LLC.
- Dromerthy, R. 2011. *Implementing Teacher-Based Assessment System: How is the stage?* Paper presented at Seminar on Teachers' Judgment within Systems of Summative Assessment. St Anne's College, Oxford: 20<sup>th</sup>-22<sup>nd</sup> June, 2011.
- Harris, M., & McCann, P. 1994. *Assessment*. Oxford: Macmillan Publishers Ltd.
- Kindler, A.L. 2002. *Survey of the States' Limited English Proficient Students and Available Educational Programs and Services 2000-2001 (Summary Report)*. National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs. Retrieved February 16, 2012. Available at: [www.ncela.gwu.edu/policy/states/reports/searports/0001/sea0001.pdf](http://www.ncela.gwu.edu/policy/states/reports/searports/0001/sea0001.pdf).
- Mansor, A.N., Leng, A.H., Rasul, M.S., Raof, R.A., & Yusof, N. 2013. The Benefit of School-Based Assessment in *Journal of Asian Social Science*. Vol 9, No. 8, 2013. Kuala Lumpur, Malaysia: Canadian Center of Science and Education.
- Mardiana, R. 2010. *The Washback Effect of National Examination (Ujian Nasional) on English Teachers' Perceptions toward Their Classroom Teaching and Students' Perceptions toward Their Learning Attitudes*. Unpublished Paper. Bandung: UPI.
- McNamara, T. 2000. *Language Testing*. Oxford: Oxford University Press.
- McCallin, R.C. 2006. *Test Administration*. In S.M. Downing & T.M. Haladyna (Eds.) *Handbook of Test Development*. London: Lawrence Erlbaum Associates Publishers.
- O'Malley, J.M., & Pierce, V.L. 1999. *Authentic Assessment for English Language Learners-Practical Approaches for Teachers*. Boston, USA: Addison-Wesley Publishing Company, Inc.
- Rivera, C., & Collum, E. 2006. *A National Review of State Assessment Policy and Practice for English Language Learners*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Shohamy, E. 1996. *The Power of Test: the Impact of Language Testing on Teaching and Learning*. NFLC Occasional papers. (College Park, MD, National Foreign Language Center, University of Maryland).
- Suen, H.K., & McClellan, S. 2003. Item Construction Principles and Techniques. In N. Huang (Ed.). *Encyclopedia of Vocational and Technological Education* (Vol 1, pp 777-798). Taipei: ROC Ministry of Education.
- Suyanto, K. K. E. 2000. *Pengembangan Kurikulum Bahasa di Indonesia*. Paper presented in The National Convention of Indonesian Education, Jakarta 19-22 September 2000.
- Weir, C. J. 2005. *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke: Palgrave Macmillan.
- Welch, C. 2006. Item and Prompt Development in Performance Testing. In S.M. Downing & T.M. Haladyna (Eds.) *Handbook of Test Development*. London: Lawrence Erlbaum Associates Publishers.
- Wijaya, A.S. 2013. *Proctoring in the Examination of Writing Proficiency*. Unpublished Paper. Bandung: UIN Sunan Gunung Djati.