Test of speaking in the school-based assessment in an upper secondary school

Andang Saehu[1]
[1] Faculty of Adab and Humanities, UIN Sunan Gunung Djati

Article Info	ABSTRACT
	Speaking is one of language skills taught but not tested in the national examination. It is fully to be prepared and administered by the schools in the school examination. However, the government and public did not have any data-based information about how the test of speaking was developed and administered. The current study investigates the ways the English teachers developed the speaking test in the school-based assessment (SBA) in one of the private upper secondary levels of education in Garut, West Java. The selection of the school was done purposively with considering that the school is accessible by the researcher and A recognition for national accreditation. The result shows that the English teachers had different ways of developing the speaking test. This study suggests that the government needs to design the national-based test specification developed by the experts in order to be used to assess whether students have met the competency standards required for graduation.

Corresponding Author:
Andang Saehu
Email: andangsaehu@uinsgd.ac.id

1. INTRODUCTION

National examination has been a high-stakes test since it has been used to determine the student graduation. Nowadays, another high-stakes test which is used, with the share 40%, to determine the student graduation is school examination or commonly known as School-Based Assessment (SBA). The notion can be seen in the Regulation No.13/2015 stating that the purpose of administering high-stakes tests is to evaluate if English teaching has achieved its goals or objectives. (Mardiani, 2010; Saehu, 2016).

English subject as one of the competencies measured in the school examination covered writing and speaking. A comprehensive study of the result of writing test in the school examination was conducted by Ginting & Saukah (2016). The current research aims to address the lack of a thorough investigation into speaking assessments within Indonesian school examinations.

A number of studies have been reported concerning the implementation of School Examination or commonly known in several countries as School-based Assessment (SBA). The first study was conducted by Cheng, Andrews, and Yu (2011) in Hong Kong. They indicated that the

SBA has recently become part of the Hong Kong Certificate of Education Examination (HKCEE) in English, administered at the edge of Secondary 5 (the fifth year of secondary education following six years of primary education, typically at ages 16-17). The HKCEE explicitly emphasizes that the SBA in English is centered on evaluating students' oral English skills and spans across two academic years: Secondary 4 and 5. Supporting Cheng et al., Fok (2012:14) said that HKCEE SBA component in English will assess students' oral performance. This SBA innovation was introduced into Secondary 4 classroom in 2005-2006.

The next study was conducted by Fook and Sidhu (2012:3) in Malaysia. Malaysia has introduced School-Based Assessment (SBA) or *Penilaian Kendalian Berasaskan Sekolah* (PKBS) into Malaysian schools under the new integrated Curriculum for Secondary Schools of *Kurikulum Bersepadu Sekolah Menengah* (KBSM). The Ministry of Education Malaysia introduced and implemented the SBA to assess students' academic progress, aligning with the Standard-based Primary School Curriculum. This initiative was rolled out in phases starting in 2011 as a component of Malaysia's educational overhaul. In line with this, Mansor et al. (2013) and Furaidah, et al (2015) that under this SBA, teachers are entrusted with increased accountability for crafting high-quality assessments that match their students' learning goals, recognizing them as the most appropriate individuals to evaluate their students' progress. Beginning 2003 the SBA commenced for both Bahasa Malaysia and English Language. Today the SBA focuses on oral English test which is a compulsory component for secondary five candidates taking SPM examination. According to Yusof (2013) and Tong and Adamson (2015) in the SBA, teachers plan their assessment, prepare the instruments, administer the assessment, examine learners' responses and report their progress.

Another study was conducted by Klenowski (2013) in Australia. In Australia, Queensland students receive an A to E report card every semester in every year in every subject. The SBA comprises the National Assessment Program—Literacy and Numeracy (NAPLAN) tests conducted in May. Students then receive their results in September, and the school receives diagnostic information in December or January. The reports are the results of comprehension in reading. Nowadays, Until 2013, the Queensland Studies Authority (QSA) in Queensland was responsible for creating the Queensland Comparable Assessment Tasks (QCAT), which were administered in grades 4, 6, and 9 (source: http://qsa.qld.au/3163.html). These tasks were developed to help teachers evaluate the characteristics required in student work that align with the national achievement standards. The assessment method employed in Queensland's School-based Assessment (SBA) comprises classroom tests and diagnostic tests utilizing Progressive Achievement Tests in Reading.

Of those countries conducting SBA, Indonesia has recently introduced this kind of assessment in 2010. The SBA in Indonesia is almost the same as that of the SBA in that other countries in terms of the responsibility of teachers or schools in planning, preparing, and administering the examination. However, although the articles about the SBA over the countries above tell no clear explanation about how SBA of certain language skills in Hong Kong, Malaysia, and Australia is planned, prepared, and administered, they show that SBA, which has now been administered in Indonesia, has been conducted in several countries. This study attempts to analyze how English teachers implemented the speaking test in the school examination as the government may not know how it is developed and administered.   This research implies the necessity of English teachers to have the examinations be of the highest quality in terms of developing test prompts of speaking skill and administering speaking test.

2. LITERATURE REVIEW

In this part some theories relevant to the conduct of this present study will be discussed. It starts with a description of theories of the relationship between assessment and instruction, followed by the theoretical application of school-based assessment.

2.1   The relationship between Assessment and Instruction

Assessment plays a crucial role in the instructional process, although it is often misunderstood in contemporary educational settings. Many educators mistakenly equate assessment with tests, but they are distinct concepts. Assessment is essentially the process of gathering information to determine what and to what extent students have achieved the intended learning outcomes. Gronlund and Waugh (2009) and Butler (2019) define assessment broadly as encompassing all methods used to evaluate how well students are meeting instructional goals.

In essence, assessment involves teachers continuously gathering information throughout the teaching process, providing opportunities for students to reflect on feedback, and integrating this feedback into their skill development. Tests, on the other hand, are specific assessment instruments conducted at predetermined times within a curriculum. They are just one aspect of assessment used to evaluate student achievement.

Brown (2004) and Miller (2008) differentiate between tests as a component of assessment, emphasizing that assessment covers the complete procedure of gathering, combining, and interpreting formal and informal measurement information. To assess means to gauge the value of something by measuring its various aspects, such as speaking or writing ability, typically using tests as a unit of measurement.

Assessment, tests, and instruction are interconnected in the educational process. The instructional process involves planning instruction, delivering it, and assessing student learning or achievement. During planning, educators identify specific learning outcomes, select appropriate materials, and organize learning experiences cohesively. They then deliver this planned instruction to students, followed by assessment to gauge how well the students have learned. These three steps are closely related, with the planned instruction aligning logically with actual instruction, and assessment reflecting the planned objectives and instructional methods.

2.2   Theoretical Application

Given that this research study focuses on implementing English school-based assessment for speaking skill, it is deemed suitable to integrate the theories discussed above.
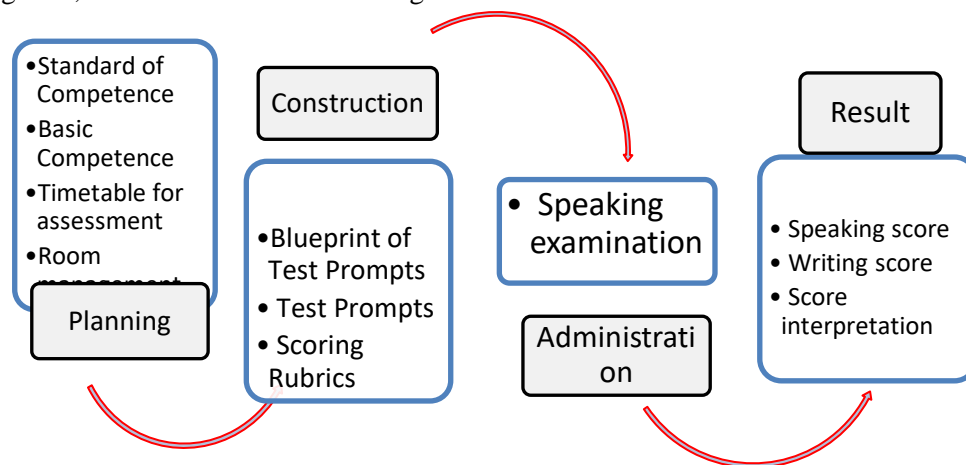

Figure 1. The Application of English School-Based Assessment for Speaking Skill

The process of formally assessing students in English speaking and writing skills is highlighted in Figure 1.3. It emphasizes the importance of systematic implementation, which involves four key stages: planning, construction, administration, and post-implementation.

During the planning stage, English teachers are advised to carefully plan the timing and location for conducting the assessment. They should also review the Standard of Competence and Basic Competence to ensure alignment with the assessment objectives. Moving to the construction stage, teachers begin by designing test blueprints for speaking and writing skills based on the identified competencies. These blueprints serve as a guide for developing test prompts that accurately assess the desired skills (Zhouyang, 2024). Additionally, teachers are encouraged to create scoring rubrics that define proficiency levels for each criterion. These rubrics aid in objectively evaluating students' performance, as they provide a clear framework for assessment. The administration stage involves administering the assessment using the developed test prompts and scoring rubrics. This stage focuses on examining students' speaking and writing skills in alignment with the assessment criteria.

Finally, in the post-implementation stage, teachers process the assessment results by scoring students' performance and interpreting their scores. This step is crucial for understanding students' proficiency levels and identifying areas for improvement. Overall, the systematic approach outlined in the text ensures a comprehensive and fair assessment of students' English speaking and writing skills, ultimately supporting their learning and development in the language.

3. METHOD

Those schools were then studied by distributing questionnaires with about 20 yes/no questions. The use of close-ended questions was intended to enhance the consistency of responses across subjects. The questions are divided into 10 items to question the speaking test develoment and the rest to ask the speaking test administration. The yes or no answers to the questionnaire were then confirmed and clarified for the reasons by interviewing the subjects of the study. The collected questionnaire data were computed in the form of percentage by looking at the total of respondents answering each question, classified by coding the data based on the characteristics of the variables taken from the questionnaire, and interpreted differently depending on the items asked in the questionnaire.

4. RESULTS AND DISCUSSION

The presentation of the research finding is discussed and focused on data collected from interview and documentation. The interview data showed that prior to constructing the test prompt of speaking skill, the English teacher developed the test prompt of speaking skill based on the blueprint or test specification. Those two steps coincide with the statement by the Standardars (American Educational Research Association—AERA, 1999) and by Standards of Operating procedures (SOP) (*Depdiknas,* 2012:2) that test development covers the purpose of the test, the test specification, the test items, the scoring guides, and the procedures.

The interview session asked the teacher "what did you do to design a blue print?" The teachers then explained some stages, consisting of analysing the competency to be tested, setting up the purpose, and delineating the scope and materials. One of the teachers, R#1 said:

"We as the teachers who designed the test for English language skills did some analysis to do with the blueprint design, like analyzing *SK* and *KD,* defining the purpose of test, and selecting the content and skill to be measured."

The statement by R#1 is relevant with the test development stages proposed by Downing (2006) that The test developer needs to initially determine the test's objective that should be maintained, the materials should be tested, the scoring rubrics should be developed, and the sequence tasks should be accomplished. Looking at the statement by R#1 who said "We as the teachers who designed the test for English language skills," this indicates that they work together as a team of test developer. Goral & Bailey (2019) pointed out that the test development process should be made in team: team for developing the test prompt, team for reviewing the prompt, and team for developing the rubrics.

The step of designing the blueprint is also known as the planning stage by Harris and McCann (1994). Here are the steps proposed by them which consist of planning the skill and materials should be tested, and what test format and scoring procedure to be prepared. This planning step can help test developers to increase accountability. Dromerthy (2011) stated that Thoughtful preparation enables examiners to discern the decisions made and the reasons behind them. One of the planning activities is defining the purpose of test. Welch (2006) and Haladyna & Rodriguez (2013) emphasized that the initial step in the test development process is for the test developer to clearly define the test's purpose and content.

Other respondent, R#2 stated that after defining the purpose of the test, he set up the 'indicator' of test prompts of speaking skill. He said"

"Hhmm setting up the indicator was the thing directly done after defining the test purpose. You can see this blueprint we designed (He showed a document). The indicators were set pursuant to the language skill to be tested to the students"

Being curious to the statement by R#2, the researcher then analyzed the documentation of the blueprint. The documentation presents the condition at the first sentence. In this context, the term "condition" refers to one or multiple operational terms employed in the initial sentence. It is written in the blueprint that

Indicator   : Students are able to, when provided with a prompt:
- Remember various phrases for suggesting something.
- Share their viewpoints using different phrases for suggesting something.
- Display proficiency in using different phrases for suggesting something.

The indicator above shows the 'condition—operational words' rather than locating audience—students at the first sentence. Meanwhile, another respondent, R#3, tried to show a document with audience at the first sentence. Take a look at the following indicator:

Indicator   : The students are able to:
      - use accurate grammar, vocabulary, and spelling
      - speak fluently

Dricoll and Wood (2007) defined indicators as a guidance in formulating the expected test prompts. Things that guide the indicators in designing the blueprint are the standard of competence (*SK*) and basic competence (*KD*). By looking at the *SK* and *KD*, an indicator will produce the test prompt that agrees to competency tested. The findings show that every teacher will have different models of indicator: putting a condition—operational verbs or locating audience—students at the first sentence.

The next interview item was "what did you do after designing the blueprint of speaking test?" Responding this question, the teachers said that they validated the test blueprints to the vice principals for curriculum affairs. Weir (2005:1) and Cahyono and Saukah (2015) described validation as the process of gathering evidence to confirm the validity of inferences regarding the test specification's soundness. It means that when creating test blueprints, it's necessary to validate them to determine if the English teachers' designs are valid.

After establishing the test purpose and determining the content and skills to be included based on the test specifications, the test prompts can be constructed. Kopriva (2008), Huang, Bailey, & Chang (2020) explained that the test specification originates from the intended purpose of the test, serving as a roadmap for creating test items or prompts. Thus, the subsequent stage in test development, as advocated by Downing (2006) and Dhami (2018) involved formulating the test prompts that assess the specific content and skill domains outlined in the test blueprints, particularly focusing on speaking and writing abilities in this research. It's essential to take into account the types of assertions or conclusions to be drawn about the targeted knowledge or skills when crafting these test prompts.

Test prompts development, according to Welch (2006:305) and Liu & Stapleton (2015), should be concise, easy to comprehend, and devoid of potential ambiguities. This recommendation appears aimed at aiding learners who vary in their learning speeds and skills; some are quick learners, while others are slow or less proficient. This aligns with Baranovcki's (2007) and Haladyna & Rodriguez's (2013) assertion that explicit guidance on whether examinees should guess answers must be accompanied by ample examples, ensuring that slower or less skilled examinees grasp the problem type. Consequently, if test prompt language is complex or unclear, one cannot reliably attribute poor test performance solely to low proficiency in the assessed skills.

In line with this, Rivera and Collum (2006) emphasize that students' understanding of what to do during a test depends on the clarity and conciseness of the instructions, commonly referred to as the prompt. It is crucial to develop a clear and concise test prompt to maintain the test's validity because unclear instructions can lead to students being unsure about what is required of them, ultimately compromising the test's validity. Luoma (2004) proposed that test prompts should clearly outline the task, method, and criteria for assessment. This suggestion implies that unclear language can lead to students potentially resulting in fail to perform in tests.

The researcher then asked to the teachers about the purpose of validating the test blueprint. They said that after creating the test specification for speaking skill, they must verify it by sharing them with others, reviewing the details gathered in the verification form, and analyzing the data gathered from the verification form. They also said:

"The validator gave some reviews or comments to the blueprint we designed. The comments are constructive judgment for better revision. These cover some mistaker or errors, time allocation, topics to discuss, and test eligibility."

Evaluating blueprints, whether through comments, reviews, or ratings, is crucial as it helps determine the effectiveness of the designed blueprints. Kindler (2002:90) highlighted the significance of external reviews, stressing the importance of allowing others to assess materials independently to identify the program's strong points and areas for improvement. Comparable opinions are shared by Weir (2005) and Bowen, Bowen, and Woolley (2007), emphasizing the need to submit test materials to reviews and revise them accordingly. It's evident that after developing test specifications, multiple reviews are necessary to ensure content quality and fairness. These reviews involve validators collecting information from testing the blueprint, which needs thorough analysis.

This information provides valuable insights to improve the blueprint's quality and align it with institutional examination standards.

The validators reviewed the blueprint of speaking test skill through the validation form, noting several areas for improvement. These included the absence of school identity, class and program details, time constraints, and clear instructions for examinees. They recommended using simple and easily understandable language, revising the examination duration, conducting a trial of the test prompt, and providing options for test topics. Welch (2006:310) and Bloxham, Hudson, Outer & Price (2015) noted that the content reviewer carefully examines every prompt to confirm its adherence to the overall test specifications. Moreover, the content reviewer scrutinizes each prompt to verify that they are all lucid, without ambiguity, and grammatically coherent.

The researcher then asked the participants the issues that guide the prompt development. The following is the general guidelines for prompt developer of speaking skill as displayed in Table 1.

Table 1. General Guidelines for Prompt Developer

Issue	Prompt Development Guideline
Accessibility	- The prompt should be available and understandable to every test taker, regardless of their gender, cultural, or ethnic background. - Variations in how well prompts are handled should correspond precisely to variations in the skills under assessment.
Fairness	- Do not include any statement that could be seen as disrespectful or harmful towards any particular group or its members. - Prompts should accurately represent individuals, ensuring fairness and avoiding stereotypes. - Examinees should have a general understanding of social and cultural issues referenced in the prompts.
Audience	- The prompts must accurately and impartially represent the audience regarding gender, ethnicity, cultural background, and religious beliefs.
Difficulty	- The prompts should possess an adequate level of complexity to facilitate the examinees' advancement in producing original written work. - Offer examinees the chance to surpass mere restatement of the prompt by encouraging originality. - The challenge posed by the language or concepts in the prompt should not hinder examinees' ability to respond effectively.
Background Knowledge	- Test prompts should not favor examinees with specific backgrounds or experiences. - Prompts must be impartial and reachable for examinees from diverse backgrounds.

The next interview item questioned to the teachers was "after successfully sending the blueprint to the expert in curriculum and revising it, what did you do?" The three teachers in the same wavelength stated that they went on constructing the test prompt of speaking skill. R#3 stated that:

"The blueprint which was designed and validated was then used as the basis for developing the test prompt."

To confirm that what has been stated by the teachers, the researcher compared the documentation of the blueprint and test prompt of speaking skill. The comparison was about the

purpose or target of the test, the content to be tested, and the scoring rubric. In addition, R#3 also stated that the test prompt was then tried-out to non-participants.

> "The test takers were the third graders, meanwhile the second graders were those joining try-out to check whether the test prompt was ready for data collection process."

The main goal of the trial is to gather evidence regarding the adequacy of the test prompt's development. The information obtained from the trial provides valuable insights for the test developer to conduct additional revisions. This coincides with Bachman and Palmer (1996:225) emphasizing the significance of the trial phase in assessment. They highlight the importance of testing the students' English speaking and writing skills in a US context through actual trials of the test prompts. Neglecting this crucial phase, as many English teachers do, raises doubts about the quality of the test prompts they create and whether they meet the intended standards.

In a nutshell, the finding on the construction of blueprint and development of test prompt of speaking skill can be figured out in the figure 1.
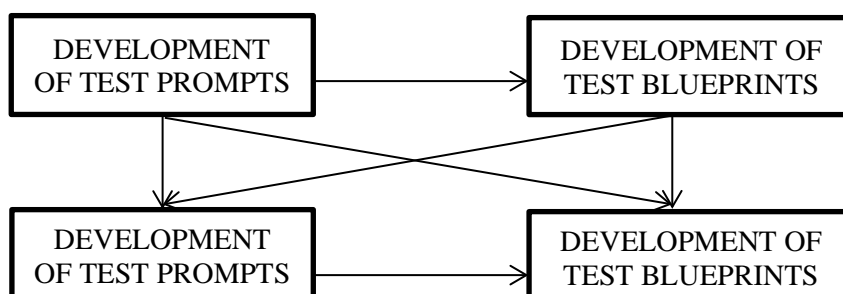

Figure 1. The Development of Test Prompts of English-Speaking Skill

5. CONCLUSION

The English-speaking skill test prompts were created according to the speaking skill test prompt blueprint. Ensuring the quality of both the blueprints and test prompts involves conducting validation and trial runs. Fortunately, the blueprint and test prompt were performed for validation and try-out activities employed by the English teachers. They awared of not skipping the validation and try-out steps will confirm the validity and reliability of the test blueprint and prompt. Skipping the validation and try-out phases could pose a significant risk to the quality of the test prompt.

REFERENCES

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.

Bloxham, S., Hudson, J., Outer, B., & Proce, M. (2015). External peer review of assessment: an effective approach to verifying standards? *Higher Education Research & Development. Vol 34 (6), https://doi.org/10.1080/07294360.2015.1024629.*

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.

Butler, Y.G. (2019). Assessment of young English learners in instructional settings. In X. Gao (Ed.), Second handbook of English language teaching. Springer. https://doi.org/10.1007/978-3-030-02899-2_24.

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.

Dhami, RS. (2018). Practice of testing speaking skill at secondary level. Accessible at https://elibrary.tucl.edu.np/handle/123456789/8308

Downing, S.M. 2006. Twelve Steps for Effeective Test Development. In S.M. Downing & T.M. Haladyna (Eds.) *Handbook of Test Development*. London: Lawrence Erlbaum Associates Publishers.

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: 20th-22nd June, 2011.

Driscoll, A., & Wood, S. 2007. *Developing Outcomes-Based Assessment for Learner Centered Education.* Danville, Virginia: Stylus Publishing, LLC.

Fok, W.K. 2012. *HKCEE English Language Schoool-Based Assessment: Its Implementation at the Frontline,* Durham theses, Durham University. Available at Durham E-Thesis Online: http://etheses.dur.ac.uk/3522/

Fook C.Y. & Sidhu, G.K., 2012. School-Based Assessment among ESL Teachers in Malaysian Secondary Schools in *Journal of the Malaysian Education Dean's Council. Vol 9. 2012.* University Sains Malaysia.

Furaidah, F., Saukah, A., & Widiati, U. (2015). Washback of English national examination in the Indonesian context. *TEFLIN Journal*, *26*(1), 36-58.

Ginting, D., and Saukah, A. (2016). Tests of writing in the school examination in epper secondary school. *Sage Open, p.1 – 7.* DOI: 10.1177/2158244016673130

Goral D., Bailey A. L. (2019). Student self-assessment of oral explanations: Use of language learning progressions. Language Testing*, 36(3), 391–417. https://doi.org/10.1177/0265532219826330*

Haladyna, T.M., & Rodriguez, M.C. (2013). Developing and Validating Test Items (1st ed.). Routledge. https://doi.org/10.4324/9780203850381

Harris, M., & McCann, P. 1994. *Assessment*. Oxford: Macmillan Publishers Ltd.

Huang, B.H., Bailey, A.L., & Chang, YS. (2020). An investigation of the validity of a speaking assessment for adolescent English language learners. *Sage Journals, Vol. 38(3), https://doi.org/10.1177/0265532220925731*.

Klenowski, V. 2013. School-Based Assessment, Standards, Teacher Judgment and Moderation: Messages from Research. In Cambridge Horizones. *School-Based Assessment: Prospects and Realities in Asian Contexts.* Cambridge International Examination.

Kopriva, R.J. 2008. *Improving Testing for English Language Learners*. New York: Routledge Taylor & Francis.

Liu, Fulan., & Stapleton, Paul. (2015). Writing prompt convergence in high-stakes tests: Exploring rhetorical functions and objects of enquiry. *Studies in Educational Evaluation,* Volume 47, Pages 68-75, https://doi.org/10.1016/j.stueduc.2015.09.001.

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.

Rivera, C., & Collum, E. 2006. *A National Review of State Assessment Policy and Practice for English Language Learners.* Hillsdale, NJ: Lawrence Elbaum Associates.

Saukah, A., & Cahyono, A. E. (2015). National exam in Indonesia and its implications to the teaching and learning of English. *Jurnal Penelitian dan Evaluasi Pendidikan*, *19*, 243-255.

Saehu, A. (2012). Testing and Its Potential Washback. In BY Chayono and RN Indah (Eds). *Second Language Research and Pedagogy. P.119-132.*

Saehu, A. (2016). The sketch of teaching speaking in Islamic tertiary level of education. *https://etheses.uinsgd.ac.id/eprints/19759*.

Tong, S. A., & Adamson, B. (2015). Student voices in school-based assessment. *Australian Journal of Teacher Education*, *40*(2), Article 2. doi:10.14221/ajte.2015v40n2.2

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.

Yusof, N.M. 2013. School-Based Assessment: Transformation in Educational Assessment in Malaysia. In Cambridge Horizones, 3 June 2013. *School-Based Assessment: Prospects and Realities in Asian Contexts.* Cambridge International Examination. Pp. 1-25.

Zhouyang, Lv. (2024). Prompt effects on assessment of interactional competence in a group oral test. Accessible at *https://doi.org/10.1111/ijal.12554*