PERFORMANCE-BASED ASSESSMENT AS A CURRENT TREND IN ELT: INVESTIGATING ITS WASHBACK EFFECTS ON SECONDARY-SCHOOL STUDENTS LEARNING

Sumardi
SMA Negeri 1 Sragen
arif_sumardi74@yahoo.co.id

Abstract

Assessment is an integral part of an English teachers' instruction since it enables them to visualize the extent of successful instructions. However, teachers might not pay much attention to their role to help their students improve their learning. Even, many English language teachers seem to focus more on traditional formats of assessment such as multiple choice, true false, matching type, etc. and pay little attention to the washback effects of these kinds of assessment on students' learning. This paper argues that a traditional assessment has at least two fundamental flaws. There is a little chance to identify the students' factual skills as it only focuses on formal assessment. In addition, it only focuses on achievement of high scores under test conditions. This orientation can be misleading to the nature of language instruction itself. Therefore, this research article explores the washback effects of performance-based assessment as a current trend in English Language Teaching (ELT) in secondary-school students' learning. To investigate this issue, in-depth interview, classroom observation, and document analysis were employed as the data collection methods. The merits of the research are to indicate how to arrange performance-based assessment and to identify its washback effects in the students' learning. The results conclude that performance-based assessment encouraged and motivated students to learn more. Students' perception in this type of assessment was positive and they strived to be more successful in their learning.

Keywords: performance-based assessment, washback effects, student learning.

Abstrak

Penilaian merupakan bagian terintegrasi pada pemberian tugas guru bahasa Inggris karena penilaian memungkinkan para guru untuk menunjukkan keberhasilan penugasan yang mereka berikan. Namun demikian, para guru tersebut mungkin kurang memberikan perhatiannya pada perkembangan belajar siswa. Bahkan, banyak guru Inggris yang nampaknya lebih fokus pada format penilaian tradisional seperti soal-soal pilihan ganda, benar dan salah, mencocokkan dan seterusnya dan kurang memberikan perhatian pada efek dari penilaian ini pada pembelajaran siswa. Penelitian ini menunjukkan setidaknya dua kekurangan utama pada penilaian-penilaian tradisional. Ditemukan bahwa dalam penilaian tradisional kurang merepresentasikan kemampuan siswa yang sebenarnya karena hanya fokus pada penilaian formal. Selanjutnya, penilaian tradisional hanya berfokus pada pencapaian nilai-nilai tinggi pada sebuah tes yang tujuannya dapat mengaburkan tujuan pemberian tugas yang sebenarnya. Oleh karena itu, penelitian ini mengeksplorasi dampak penilaian pada penilaian

berbasis kinerja sebagai tren terkini dalam pengajaran bahasa Inggris dalam pembelajaran siswa sekolah menengah. Untuk meneliti isu tersebut, wawancara mendalam, observasi kelas dan analisis dokumen digunakan dalam metode pengumpulan data. Tujuan penelitian ini adalah untuk mengidentifikasi bagaimana cara menyusun penilaian berbasis kinerja dan dampaknya pada pembelajaran siswa. Hasil penelitian ini menunjukkan bahwa pemilaian berbasis kinerja memberikan motivasi siswa untuk lebih giat belajar. Persepsi siswa dalam jenis penilaian ini positif dan mereka berusaha untuk lebih berhasil dalam belajar.

Kata Kunci: *penilaian berbasis kinerja, dampak pemberian tugas, pembelajaran siswa*

1. Introduction

Language assessment is a field of study under the umbrella of applied linguistics (Safa and Goodarzi, 2014) and its main focus is the assessment of first, second and foreign languages in school, college, or university context. The assessment may include listening, speaking, reading, writing, or may be an integration of two or more of these skills. There are at least two modes of language assessment that English language teachers may follow to assess a student's learning. The first one is traditional standardized achievement tests. It is mostly in the format of paper and pen tests such as multiple-choice, true false and matching type tests. These kinds of test are used for high-stake decision in assessment and accountability system. Traditional standardized achievement tests are still widely used by English language teachers throughout Indonesia. The second mode is authentic assessment. It refers to a form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills (Mueller, 2014) and the importance of such assessment of real-life-situation-based performance becomes more crucial in identifying students' language abilities (Aksu Atac, 2012). One of the formats of authentic assessment is performance-based assessment. Performance-based assessment can be used by English language teachers to assess a student's learning by following numerous methods of assessment such as oral presentation, story-telling, dramatic reading, English debate, etc. It is a more challenging format of assessment to develop and implement than those of others as assessment is a key to student engagement (Irwin and Hepplestone, 2012). These two kinds of assessment have their own strengths and weaknesses to assess students' learning outcomes in English Language Teaching (ELT) classroom.

A traditional assessment, especially in the format of multiple-choice test, is still considered as a more accessible assessment technique for English language teachers. Although this kind of assessment can be relatively difficult to develop, it is easier to administer than that of performance-based assessments. When a multiple-choice test is applied to assess students' learning outcomes, teachers may objectively score students' learning outcomes as teachers' subjectivity has no influence to the results of scoring. Additionally, these test results offer accountable measures for students' learning outcomes. However, scholars have identified limitations of multiple-choice testing. Abedi (2010) states that traditional assessments do not afford an opportunity for students to present a comprehensive picture of what they know and are able to do in content areas. In addition, multiple-choice test as one of traditional assessment formats usually fails to assess higher order skills and other skills essential for functioning in schools or work settings (Haney & Madaus, 1989; Neill & Medina, 1989; O'Neil, 1992; Wiggins, 1989 in Pierce & O'Malley, 1992). In addition, multiple-choice

tests are generally incomplete since they portray an individual at a single moment in time within particular context (McTighe & Ferrara, 2011). Due to these limitations of traditional assessment, there has been a current trend and a growing interest among English language teachers to deploy a performance-based assessment as one of authentic assessment formats since this mode of assessment is in accordance with the requirements of current curriculum in Indonesia, i.e. Curriculum 2013.

Performance-based assessments may help fill the gaps left by traditional assessment approach. Research by Bass, Magone & Glaser (2002) identified that performance-based assessments allow all students, especially those with different language backgrounds, to engage in cognitively complex activities such as generating strategies, monitoring work, analyzing information, and applying reasoning skills. Earlier research by Linn & Burton (1994) revealed that performance-based assessments have appeal as mode of effective assessments and are better reflections of criterion performances that are of importance outside the classroom. Related to English language assessments, performance-based assessments can actually help identify language factors that influence assessment outcomes (Abedi, 2010). Similarly, Goldschmidt, Martinez, Niemi, & Baker (2007) found that performance-based assessment as the open-ended assessments improve the chances for English language learners (ELL learners) to engage with language production and learning. In addition, it also offers unique opportunities for ELL learners to express their knowledge in a broader sense than the limited linguistic opportunities given to them in traditional multiple-choice items. In short, performance-based assessments may be used by English language teachers to assess the students' real language performances.

Washback effects associated with traditional and performance-based assessments can further add to the assessment of students' learning outcomes in ELT classrooms. Washback is the effects of tests and/or assessments on teaching and learning (Dorobat, 2007) and can be positive or negative (Cheng & Watanabe, 2004; Brown, 2004) both for students and English language teachers in ELT classrooms. The positive washback may occur when teachers, after deploying performance-based assessments, try hard to reflect on their method of teaching and then modify or make it better to facilitate students to learn better in ELT classroom; or students, after taking the test, may be well-motivated to learn more to fulfill their needs in learning English as a foreign language. Conversely, the negative washback occurs when students are discouraged or less motivated to learn after taking the test and teachers dominate their teaching by coaching for testing session or examination (Dorobat, 2007). Another research by Darling-Hammond (2006) found that performance assessments that require students to evaluate and solve complex problems, conduct research, write extensively, and demonstrate their learning in projects, papers, and exhibitions have proven keys to motivating students and attaining high levels of learning. To summarize, a well-constructed design of performance-based assessment may motivate students to engage totally during the class and encourage teachers to facilitate effective student learning in ELT classrooms.

A current trend of assessing students' learning outcomes emphasizes on the learning by using performance-based assessment for assessing students' real language performances. It has become a central issue for English language teachers who are required to apply authentic assessments to assess students' learning outcome (Ditjen Dikdasmen, 2015). In addition, Curriculum 2013 requires that English language teachers shift their assessment choices to include traditional assessments and performance-based assessments. That is why performance-based assessments are now getting their popularity among the English language teachers to measure students' communicative competence. By using such performance-based assessments, English language teachers may ask students to perform real-world

tasks that demonstrate students' real-life competences in employing English as a means of communication.

Acknowledging the importance of performance-based assessment to assess the students' learning, this research article explores the washback effects of the application of performance-based assessment as a current trend in ELT classrooms for secondary-school students. The secondary-school students were selected as the subjects of this research since they had relatively adequate levels of language proficiency to communicate both in oral and written English compared to those of primary-school students. Their speaking performances are categorized into interactive speaking performances, i.e. oral communicative competences which include multiple exchanges and/or multiple participants who interact with each other in two forms of interaction – transactional and/or interpersonal language interactions; and extensive speaking performances, i.e. oral communicative competences which include extensive oral language productions such as speeches, oral presentations, and story-telling, during which the opportunity for oral interaction from listeners is either highly limited – perhaps to nonverbal responses – or ruled out altogether (Brown, 2004). Therefore, they are supposed to have adequate oral communicative competences. For this purpose, the subsidiary research questions are formulated as follows:

a.   How should performance-based assessments be designed to assess the secondary-school students' learning in ELT classroom?
b.   What forms of washback effects were evident following performance-based assessment for secondary-school student learning?

2.   Methods

This study aimed at exploring the washback effects of performance-based assessment as a current trend in English Language Teaching in secondary-school students' learning. This study focused specifically on investigating the washback effects of performance-based assessment which was deployed in speaking class. Thirty two students in the twelfth grade of a secondary school in Sragen, Central Java volunteered to participate in this research project. They have received formal English instructions for six years or more, three years at SMP and three years at SMA, and their language ability ranged from elementary to intermediate levels. Students' initial motivation to enroll in the speaking class and to speak in English was also identified as low.

Assessment is quite important in language instructions and it is not an end in itself but a vehicle for educational improvement (Dorobat, 2007). In addition, assessment is most effective when it reflects on understanding of learning as multidimensional, integrated, and revealed in performance over time (Brown, 2004) and should be authentic (Callison, 1998). Therefore, performance-based assessment is part of an authentic assessment deployed in English language instructions. To investigate its washback effects in secondary-school students' learning, a performance-based assessment was used by the researcher. He also serves as an English teacher at the ELT classroom where a research was conducted, by asking students to perform oral presentations and was designed following several stages: (1) teachers provided firstly a stimulus related to the topics to be discussed during the instructional process; (2) teachers established the students into groups consisting of three or four students in every single group; (3) each group of students was given the opportunity to identify the varieties of social issues containing aspects of the pros and cons and occurring in the local communities; (4) each group of students carried out an intensive discussion with regard to the social problems to identify why those problems provoke the pros and cons; (5) each group of students was, then, given an opportunity to deliver an oral presentation for 10 to 15

minutes; (6) as one of the groups had completed a presentation, their classmates were given the opportunities to deliver a range of questions in connection with the materials presented; and (7) each member of the group, who was doing a presentation, ought to provide responses to the variety of questions submitted by their classmates. These stages were designed as the hierarchical activities to realize or to manifest a performance-based assessment in speaking class.

Empirical data were collected through in-depth interview, classroom observation, and document analysis. The in-depth interviews served as verbal justifications from the participants for what were observed in the classroom and for what were unknown in the classroom observation (Widodo, 2016). The classroom observation data were video recorded, while interview data were audio recorded. Finally, documents related to the theories underlying this project were analyzed to provide a complete depiction on how this project should be conducted.

The empirical data were qualitatively analyzed. This qualitative analysis entails an interpretative enterprise. To discern multifaceted qualitative data, micro-interaction analysis using Anderson's interaction framework (2009) as cited in Widodo (2016) was deployed to examine the collected data garnered from classroom observations. All the data were transcribed and reviewed, and the data showing participants' actions, moves, and interactional patterns were tabulated. The data were categorized into moment-by-moment interactions, i.e. teacher–student and student–student interactions; characterizations of such interactions, i.e. how much both the teachers and the students valued such interactions; and interactional patterns, i.e. how students interacted with their peers. These three layers of the analysis allowed the author to capture some emerging findings relevant to the two research questions.

3.   Research Findings and Discussion

Having collected the data, the selective data were hereafter analyzed to identify the washback effects of performance-based assessment in secondary-school students' learning. Two finding themes were identified: (1) students' initial motivation to speak in English; and (2) the washback effects of performance-based assessment in students' learning. The findings are presented in narrative way (Sandelowski, 1991) accompanied with discussions. A narrative approach is presented as a framework to analyze and produce explanatory stories from mass of data gained from the documents, interviews, and observation so that the information, events, and happenings gathered from the documents, interviews and observation would be meaningful and coherent.

3.1   Students' Initial Motivation to Speak in English

The students' initial motivation to speak in English for most students at school was identified all very low. Generally, the students were afraid of making any mistakes to speak in English that affected their motivation and self-confidence to speak. A preliminary observation identified that the students tend to keep silent when they were asked to speak in English. Some of them looked nervous; whereas others seem to be unconfident to speak in English. In line with the results of observation, WDT said, "I'm afraid to speak in English. Everyone laugh at me ... my English is bad". Similarly, AKB said, "Speaking in English is not easy. I can't speak in a good grammar ... the grammar is different from – that of the grammar of – Bahasa Indonesia". This condition has resulted in inadequate speaking abilities. The students' motivation to speak and their ability to speak in English were the two interrelated-variables. The higher the students' motivation, the better the students' English proficiency. Similarly, Aidogmus, Aksu & Rich (2014) state that a motivation is very closely linked to

the students' ability to communicate orally. In addition, Quadir (2014) also points out that the motivation has strong correlation with verbal or oral communication skills, especially for students who are learning English as a foreign language. The latter opinion explicitly confirmed that a teacher is supposed to be able to design an appropriate mode of assessment in order to motivate the students to learn better and to communicate well and spontaneously in English. This is a very important issue because students' abilities to communicate orally and spontaneously are the major indicator of the students' success in learning English.

The students seem to be unmotivated to speak in English. Based on initial interviews conducted by the researchers, there were numerous reasons why the students' motivation to communicate orally was so low. A student, IBN, stated, "I'm less motivated to speak In English. My vocabulary mastery is limited". Similarly, DND claimed, "I'm not really sure with my English. ... I'm not confident enough to speak. I'm afraid that my sentences are wrong". In short, the students' low motivation to speak in English was due to several reasons. Among those reasons were that (1) the students were afraid of making any mistakes when they wanted to communicate orally in English; (2) the students' mastery on English vocabularies was very limited; and (3) the students' level of confidence to speak in English was very low as well. Another dominant factor leading to the students' low motivation was that a mode of assessment used by teachers tended to be monotonous and mechanistic. To measure students' speaking skill, the teacher relied solely on a mode of retentive dialogue-based assessment. Students in groups were asked to carry out a discussion to prepare a script of dialog and they hereafter put it into practices in front of the class. This mode of assessment seemed to be monotonous and mechanistic since the teacher repeatedly used this mode of assessment and what the students expressed in the dialogue were the expressions and statements that have been formerly prepared and memorized. As YGA remarked, "… I get bored with the speaking class. Practicing dialog doesn't challenge me to speak spontaneously". Another student, RMT, claimed, "... usually we are asked to prepare a dialogue with a group of friends then practice it in front of the class in 5 - 7 minutes". When the assessment of speaking was going on, the teacher just asked the students to set up a dialogue script with their friends in group and hereafter to practice it in front of the class. It did not challenge the students to learn actively during the speaking class in order to achieve the desirable learning outcomes.

This activity took place repeatedly when speaking classes were carried out with the result that students were not challenged and motivated to speak in English spontaneously. In performance, the students seemed to be able to speak in English fluently, but the students' fluency in speaking English was identified as a result of students first memorizing the utterances they prepared before practicing a dialog in front of the class. Thus, the students were generally not able to communicate spontaneously. In addition, the teacher did not provide students the proportional opportunities to speak in English as well. That was why the teacher needed to cope with those problems by implementing a particular treatment in order to motivate the students to learn more and to allow them to be more active during the speaking class. Students' high motivation is believed to have a good impact on students' abilities to speak in English. In order that students' motivation and speaking skills improve better, English teachers need to create an exciting and enjoyable learning activity so that students are motivated to learn well and not feel bored. This effort can be done if teachers are able to choose the appropriate teaching method in accordance with the students' learning needs.

There are numerous ways teachers may carry out to improve students' learning motivation to speak in English, one of which is using varied and comprehensive modes

of assessment (Williams & Williams, 2012) and an appropriate mode of assessment is believed to affect students' learning motivation (Cheng & Watanabe, 2004; Dorobat, 2007; Anderson; 2009). Therefore, a performance-based assessment was used in this research to overcome students' low motivation to speak in English. By using this method, it was hopefully that students' motivation to speak in English could improve better. The increases of students' motivation are possible since this mode of assessment provides opportunities for groups of students to communicate in three different directions, i.e. communication among students in the groups, communication with students in the group with their audiences, and communication between students in the group with their teacher. These directions of communication are different from those of a dialogue-based mode of assessment that is commonly used by teachers to assess students' oral communication. In this method, the flow of communication often occurs only in one direction, i.e. communication between interlocutors.

To overcome the students' low motivation to communicate orally in English, the teacher took an action by asking students to perform oral presentation in group, as one of performance-based assessments, in front of the class. This mode of assessment allows students to communicate orally in various directions spontaneously. The action was conducted in two parallel cycles. At the first cycle, a group of three or four students were asked to present their works in front of the class. At this stage, every single group of students only presented their works without any questions delivered by their peers or classmates or audiences. Subsequently at the second cycle, after a group of students presented their works, their peers or audiences were hereafter given some opportunities to ask questions towards the materials presented by that group. To stimulate the audiences to convey various questions for the presenters, the teacher promised to provide additional scores as extrinsic rewards (Girmus, 2011) for those who actively asked the questions. The additional scores were provided as a form of appreciation to the students' courage to convey the questions. Those various actions conducted at the first and second cycles were eventually intended to improve the students' motivation to speak in English.

3.2   The Washback Effects of Performance-Based Assessment on Student Learning

During the first cycle of activities, the students conducted the oral presentations in groups for 15 to 20 minutes. Before doing so, each group of students made some preparations by undertaking an intensive discussion with their peers in group. They discussed everything related to the materials that they wanted to present in front of the class. They had to prepare several slides of power point as well. During this preparation phase, it was observed that the students showed a very good cooperative attitude because every group member could provide and receive their peers' opinions and thoughts. This cooperative attitude might certainly be one of the triggers to increase the students' achievement. This is in line with the opinion of Gillies & Boyle (2010), who state that the cooperative attitude which emerged during the learning process, can improve students' academic achievement. Thus, the cooperation among the students is very important to support the success of students' learning.

In addition to the cooperative attitude, the discussions done by the students also allowed the students to respect to each other. Each member of the group respects any thoughts and opinions put forward by others and seeks to respect to a variety of opinions and thoughts delivered by other group members as well. This mutual respect allowed the students to create a positive relationship among the group members with the result that the learning environment became more conducive. Similarly, Bucholz & Sheffler (2009) state that the mutual respect among groups of students create a positive relationship among those students so that the

learning environment will be more conducive. The cooperative attitude and mutual respect among the students were the positive washback effects arising from the implementation of performance-based assessment. The emergence of those students' behavior has alluded the students' learning becoming more productive and enjoyable.

After the preparatory process was finished, each group of students presented the group's work in front of the class. At the end of the presentation, there were question and answer sessions between the presenters and their audiences. The questions are supposed to be in relation to many things which were formerly presented by the group. The processes of question and answer were carried out spontaneously with the result that the students who asked the questions and the members of the group who answered those questions could communicate orally without using a variety of expressions and statements that had been prepared in advance. In other words, those students carried out oral communications in English spontaneously. The process of spontaneous oral communication was certainly capable of directing the students to be able to communicate better in English. Therefore, Buo and Waibel (1996) state that the process of spontaneous oral communication can train students to communicate with the target language automatically and independently. This statement implies that the process of spontaneous communication is capable of directing students to be able to communicate fluently in the target language.

It was, however, identified that the students' motivation to communicate orally in English observed in the first cycle of this research did not increase significantly. The results of the observations indicated that there were only a few students in speaking class dared to express their thoughts, opinions, and ideas in the form of questions addressed to the group who carried out the oral presentation. Motivation was an important factor that might affect the success of students' learning. Therefore, the students' motivation needed to be improved at the second cycle in order that their abilities to speak in English might be improved as well. An effort to improve the students' motivation was then done by providing a positive stimulus in the form of additional scores for the students who dared to provide their thoughts, ideas and opinions. The stimulus was expected to increase the students' motivation to be better in the subsequent cycle.

During the second cycle, it was identified that the students' motivation was much better than that of the students observed at the first one. On the one hand, the student's motivation identified at the first cycle was still in a category of fairly high as only a few students who actively conveyed their ideas, opinions, and thoughts addressed to the members of groups who had carried out an oral presentation. Many students still tended to be passive. In contrast, the students' motivation at the second cycle turned into high motivation in which most of students seized and scrambled an opportunity to convey their ideas, thoughts, and opinions in the form of questions addressed to members of other groups who had carried out the oral presentation. Moreover, students had a positive perception on the oral presentation as a mode of teaching and assessing speaking skills. DE said, "In my opinion, I am happy with such a method – Oral Presentation method. ... we can discuss and argue with the members of other groups. Not sleepy, Sir." Similarly, ACDNW confirmed, "I like the presentation, sir. Because we see that learning can be more active. Many of us are scrambling to ask questions". This suggested that the students' motivation in the second cycle turned into greatly higher. The students' high motivation occurred since the oral presentations were accompanied with positive stimulus in the form of additional scores provided by the teacher for each student who dared to express their ideas, thoughts and opinions orally addressed to the members of other groups who had done an oral presentation. Thus, it was understood that the stimulus that was rightfully provided to the students would be able to increase their motivation to

learn more and more. Furthermore, Cheung (2001) states that the stimulus is a very effective key to the success of students' learning outcomes. Cheung also explains that the stimulus can increase students' motivation to learn English. Therefore, teachers need to provide a good stimulus in order to improve students' motivation to learn better and to increase the effectiveness of the process of learning as well.

Those analyses revealed several positive washback effects of performance-based assessment in secondary-school students' learning. Based on the results of in-depth interviews and systematic observations, it was identified that performance-based assessment promoted positive washback effects on students' behavior and way of their learning with respect to the following matters (1) it eliminated the students' boredom and tedium during the speaking class; (2) it promoted the students' cooperative attitude; (3) it allowed the students for high order thinking as they had to speak spontaneously in English; (4) it increased the students' motivation to learn more and more; (5) it increased the students' quality of learning; (6) it provided the students a real-life communication in English; and (7) it caused to emerge the students' curiosity about the course contents. Therefore, those findings imply that oral presentation as one of performance-based assessments could be an effective pedagogical measurement tool which could well replace the traditional standardized achievement tests as these efforts were planned to maximize students' learning outcomes and students' development as well.

4.   Conclusion

This study revealed that performance-based assessment has positive washback effects in students' learning in ELT classroom. These positive washback effects might not have resulted using the traditional standardized achievement tests where the students just choose the options provided in a particular set of the test. In other words, there might be significant differences between performance-based assessment and traditional standardized tests concerning their washback effects in students' learning in ELT classroom. Performance-based assessment was proven to have the positive washback effects as students were motivated well to learn while the traditional assessments were not comparatively of the same positive washback effect. Furthermore, the positive washback effect of performance-based assessment underscores the effectiveness of this language assessment mode as an alternative substitute to traditional assessment methods in educational measurement.

There are some implications for the results of the study. First, English language teachers may need to reflect on their assessment practices to determine whether their current assessment practices help to improve their students' language learning processes. Second, if teachers rightly assume that the main purpose of all assessment practices is to foster students' learning, and if they further believe in the efficiency of the alternative modes of assessment for the intended learning, it is justifiable to compliment the traditional assessment procedures with the alternative modes like that of performance-based assessment procedures.

References

Abedi, J. (2010). *Performance assessments for English language teachers*. Stanford, CA: Standford University, Standford Center for Opportunity Policy in Education.

Aidogmus, Y; Aksu, R; & Kaya, M.F. (2014). The relationship between communication skills of geography teachers and students' level of motivation towards geography lesson. *International Journal of Academic Research in Business and Social Sciences 4 (5)*

Aksu Ataç, B. (2012). Foreign language teachers' attitude toward authentic assessment in

language teaching. *The Journal of Language and Linguistic Studies*, *8 (2)*

Anderson, K. T. (2009). Applying positioning theory to the analysis of classroom interactions: Mediating micro-identities, macrokinds, and ideologies of knowing. *Linguistics and Education*, *6 (20)*.

Bass, K.M., Magone, M.E., & Glaser, R. (2002). *Informing the design of performance assessments using a content-process analysis of two NAEP science tasks*. (CSE Technical Report No. 564). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Brown, H.D. (2004). *Language assessment: Principles and classroom practices.* San Francisco: Longman Publisher

Burns, A. (2010). *Doing action research in English language teaching*. New York: Routledge

Callison, D. (1998). Authentic assessment. *School Library Media Activities Monthly 14 (5)*

Irwin, B & Hepplestone, S. (2012). *Examining increased flexibility in assessment formats.* Retrieved from http://shura.shu.ac.uk/6529/ 1/Irwin%20Flexible_assessment_ formats_FINAL.pdf

Bucholz, J.L. & Sheffler, J.L. (2009). Creating a Warm and Inclusive Classroom Environment: Planning for All Children to Feel Welcome. *Electronic Journal for Inclusive Education, Vol. 2, No. 4 [2009], Art. 4*

Buo, F. D., and Waibel, A. (1996). FeasPar - a feature structure parser learning to parse spoken language. *In Proceedings of the COLING '96*, Kopenhagen.

Cheng, L., & Watanabe, Y. (2004). *Washback in language testing: Research context and methods.* New Jersey: Lawrence Erlbaum Associates, Inc.

Cheung, C.K. (2001). The use of popular culture as a stimulus to motivate secondary students' English learning in Hong Kong. *ELT Journal Vo. 55/1*

Darling-Hammond, L. (2006). No Child Left Behind and high school reform. *Harvard Educational Review, 76*(4)

Ditjen Dikdasmen. 2015. *Panduan penilaian untuk sekolah menengah atas*. Jakarta: Kementerian Pendidikan dan Kebudayaan

Dorobat, D. (2007). *The methodology of evaluation and testing.* Athena: Ministerul Educatiei si Cercetarii

Gillies, R.M. & Boyle, M. (2010). Teachers' reflections on cooperative learning: Issues of implementation. *Teaching and Teacher Education 26 (10)*

Girmus, R.L. (2011). *How to motivate your students*. Articles was presented at New Mexico State University -Grants Round -Up Conference, September 30 – October 1, 2011

Goldschmidt, P., Martinez, J. F., Niemi, D., & Baker, E. L. (2007). Relationships among measures as empirical evidence of validity: Incorporating multiple indicators of achievement and school context. *Educational Assessment, 12 (3)*

Hine, G.S.C. (2013). The importance of action research in teacher education programs. I*ssues in Educational Research*, 23(2), 2013: Special Issue

Haney, W., & Madaus, G. (1989). Searching for alternatives to standardized tests: Whys, whats, and whithers. *Phi Delta Kappan*, *70 (87)*.

Kemmis, S. & McTaggart, R. (1988). *The Action Research Planner* (3rd ed.). Geelong,

Melbourne: Deakin University Press.

Linn, R.L., & Burton, E. (1994). Performance-based assessment: Implication of task specificity. *Educational Measurement: Issues and Practice, 13(1)*

McTighe, J., & Ferrara, S. (2011). *Performance-based assessment in the classroom.* Retrived from http://jaymctighe/wordpress/wp.context/uploads/ 2011/004/performancebased-assessment-in-the-classroom.pdf

Mueller, J. (2014). *Authentic assessment toolbox*. Retrieved from http://jfmueller.faculty.noctrl.edu/toolbox/whatisit.htm

Neill, D.M., & Medina, N.J. (1989). Standardized testing: Harmful to educational health. *Phi Delta Kappan, 70(97)*.

O'Neil, J. (1992). Putting performance assessment to the test. *Educational Leadership*, *49(8)*

Valdez Pierce, L., & O'Malley, J. M. (1992). Performance and portfolio assessment for language minority students. *NCBE Program Information Guide Series*, *14 (9)*

Quadir, M. (2014). Relationship between learners' motivation factors and speaking strategy factors to learn oral communication in English. *The English Teacher 43 (3)*

Safa, M.A, and Goodarzi, S. (2014) The Washback Effects of Task-based Assessment on the Iranian EFL Learners' Grammar Development. *Procedia - Social and Behavioral Sciences 98 (14)*

Sandelowski, M. (1991). Telling stories: Narrative approaches in qualitative research. *Journal of Nursing Scholarship. 23 (3)*

Widodo, H.P. (2016). *Engaging young learners of English in a genre-based digital storytelling project*. Cambridge: Cambridge University Press

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, *70 (13)*.

Williams, K.C. & Williams, C.C. (2012). Five key ingredients for improving student motivation. *Reseacrh in Higher Education Journal*. Retrieved from http://www.aabri.com/manuscripts/11834.pdf

(Endnotes)
1.   *This article was presented at International Conference on Educational Research and Evaluation at Yogyakarta State University, May 29 – 30, 2016*