

THE ANALYSIS OF AUTHENTICITY IN SUMMATIVE ASSESSMENT IN AVOCATIONAL SCHOOL IN BANDUNG

Fauziah

*Fakultas Keguruan dan Ilmu Pendidikan Universitas Jabal Ghafur Sigli
fziah05@yahoo.com*

Abstract

Authenticity is one of effective test criteria and principles that are important in language testing as it is in language teaching. The crucial aspects of task authenticity are whether real communication takes place and the language has been used for genuine purpose. Large-scale multiple-choice tests that were used as summative assessment cannot offer much authenticity, nor can portfolios and such alternatives achieve much practicality or reliability. The challenge that was faced by teachers is to transform the inauthenticity produced tests into more pedagogically fulfillment. Hence, this study focuses on analysis of authenticity in summative assessment and a number of approaches. The data collected was summative test in one of vocational school in Bandung. The analysis used five questions to evaluate the extent to which a test is authentic. The study found out that accomplishing the inauthenticity was extra job for teacher. It demanded effort in fulfilling one of important aspect in language testing.

Keywords: *Authenticity, Language testing, Summative assessment*

INTRODUCTION

Assessment poses a challenge to teachers and test designers. Formal standardized tests are almost by definition highly practical, reliable instruments. They are designed to minimize time and money on the part of test designer and test-taker, and to be painstakingly accurate in their scoring. Alternatives such as portfolios or conferencing with students on drafts of written work, or observations of learners over time all require considerable time and effort on the part of the teacher and the student (Brown: 2004). Even more time must be spent if the teacher hopes to offer a reliable evaluation within students across time, as well as across students. But the alternative techniques also offer markedly greater washback, are superior formative measures, and, because of their authenticity, usually carry greater face validity.

This relationship shows practicality/reliability on one axis and washback, authenticity on the other. Notice the implied negative correlation: as a technique increases in its washback and authenticity, its practicality and reliability tend to be lower. Conversely, the greater the practicality and reliability, the less likely you are to achieve beneficial wash back and authenticity (Brown: 2004). Large-scale multiple-choice tests cannot offer much washback or authenticity, nor do can portfolios and such alternatives achieve much practicality or reliability.

However, teachers and test designers obviously try to design the assessment as effective and efficient as possible in term of time and money. Everyone knows teachers secretly hate to grade tests. The two principles that stand out in support of multiple choice formats are practicality and reability. Those are some of reason why they choose multiple-choice as the tool to assess student's capability. The problem is assessments that require behaviors or cognitive operations that are not intrinsically meaningful, (e.g.

responding to multiple-choice questions on an externally produced standardized test) are not authentic.

It should not sit idly by accepting the presumably inescapable conclusion that all standardized tests will be devoid of washback and authenticity. With some creativity and effort, it can transform otherwise inauthentic and negative-wash back-producing tests into more pedagogically fulfilling learning experiences. A number of approaches to accomplishing this end are possible, there are; building as much authenticity as possible into multiple-choice task types and items, designing classroom tests that have both objective-scoring sections and open-ended response sections, varying the performance tasks, turning multiple-choice test results into diagnostic feedback on areas of needed improvement, maximizing the preparation period before a test to elicit performance relevant to the ultimate criteria of the test, teaching test-taking strategies, helping students to see beyond the test: don't "teach to the test", triangulating information on a student before making, a formal assessment of competence.

The purpose of this paper is to analyze the authenticity in summative assessment; the challenge that faces conscientious teachers and assessors in teaching profession is to transform the inauthenticity producing tests into more pedagogically fulfillment.

Learning and Assessment

Black has argued that 'researchers are faced with the difficult task of changing understanding of assessment issues, both amongst the general public and amongst policy makers' (2000, p. 407). A great deal is known about skills learning and learning for understanding, rather less about encouraging the incremental self-theories that go with persistence and strategic thinking (Dweck, 1999 as cited in Knight,

Peter. T, 2002), and there is considerable dispute about the extent to which learning is contextualized and hence transferable. However, there is good evidence that student achievement is related, first and foremost, to engagement (Astin, 1997 as cited in Knight, Peter. T, 2002). Engagement does not simply equate to the amount of involvement in and time on task, important though that is. It extends to learners' engagement in communities of practice, to their involvement in a variety of networks and to the amount and quality of interchanges with others.

This is an endorsement of the neo-Piagetian position that cognitive engagements with others are powerful stimuli for learning and of Vygotsky's analysis of learning as social acts. According to Brown & Duguid (2000), participation in communities and networks regularly sustains learning that is not easily specified in advance, cannot necessarily be measured and is often unpredictable. Important things are learned in vibrant communities that lie out with the formal curriculum and complement it.

Notice that these assessments have at least three conditions to meet if they are to be consistent with the account of learning that has just been sketched.

1. They have to be faithful to the curriculum (charged with developing understandings, skills, self-theories and reactivity).
2. They must align with the notion that education is concerned with some degree of abstraction, generalization or transfer.
3. They should not impede student engagement in communities of practice, but should encourage behaviors associated with good learning.

Authentic Assessment

The idea of using authentic language material in teaching a foreign language is generally approved by the vast majority of language teachers, especially in teaching ESP, where it is hoped to be used for achieving a 'real-life communicative purpose' (Lee, 1995: 324 as cited in Blagojevic: 2013). Toward the end of the 1970s, there was an inevitable propensity towards communicative use testing and scholars felt that authentic stimulus material was a necessary component of any test of communicative ability (Shomossi & Tavakoli, 2010: 4). Therefore, authentic tests are often regarded as synonymous with 'communicative' tests, 'direct' tests, and 'performance' tests, etc. (Jian-lan, 2007). Lewkowicz (2000) believes that despite the importance of authenticity, there hasn't been a marked body of research to demonstrate the characteristic.

Authentic assessment is a process-oriented means of evaluating communicative competence, cognitive abilities and affective learning (Kohonen, 1999, Hart, 1994, O'Malley & Pierce, 1996 as cited in Finch, 2002), using reflective forms of assessment in instructionally-relevant classroom activities and focusing on curriculum goals, enhancement of

individual competence and integration of instruction and assessment. "The essentially interactive nature of learning is extended to the process of assessment" (Williams & Burden, 1997), examining what learners can do with their language, through real-life language use tasks (Weir, 1998). So that, authentic assessment means collecting information about learner progress and the social learning environment in the class, along with a re-assessment of classroom roles and responsibilities.

Evaluate the extent to which a test is authentic by asking the following questions (Brown: 2004, 35):

1. Is the language in the test as natural as possible?

The question will be examined in relation to the views on the nature of test tasks, test takers and the analysis of Target Language Use (TLU), as well as its practical limitations. This item will have implications for both achievement and proficiency testing, with a focus on the real-life language use environment.

2. Are items as contextualized as possible rather than isolated?

This item will assess the authenticity of each item in summative assessment in which the conversation is one that might occur in the real world, even if with a little less formality. Doye (1986) provides an example of such tests by referring to the ways that a language learner can demonstrate performing the speech act of "Asking the way in an English speaking environment".

3. Are topics and situations interesting, enjoyable, and/or humorous (relevant for the learner)?

Bachman suggested that there was a need to distinguish between two types of authenticity: situational and interactional authenticity. Situational authenticity - that is, the perceived match between the characteristics of test tasks to target language use (TLU) tasks and interactional authenticity - that is, the interaction between the test taker and the test task (Bachman, 1991 as cited in Lewkowicz, 2000: 48). The item evaluates the topic in the test items that perceived match between the characteristics of test tasks to (TLU) tasks and the test takers.

4. Is some thematic organization provided, such as through a story line or episode?

This item aim is to assess different types of abilities that underpin literacy in contexts that are similar to actual situations in which those abilities are used. Story line provides thematic organization in designing test item. The thematic organization leads to achieve the authenticity in language testing.

5. Do tasks represent, or closely approximate, real-world tasks?

Authenticity as the extent to which test tasks replicate real-life language use tasks (Shomoossi & Tavakoli, 2010). Authentic tasks were considered to be those mirroring real-life tasks, but they did not give rise to genuine interaction because

they were simulations and not real tasks. This question evaluates the authenticity of expression that possibly use in the real-life situation.

Summative Assessment

Summative assessment is usually conducted in the last few weeks of term to see how good students have learned what they were supposed to have learned. The results from these assessments are aggregated and used to determine whether a student has fulfilled the specified learning outcomes and may achieve some kind of accreditation. This usually causes a degree of anxiety since the grades received in summative assessments are final and can affect their future prospects. In summative assessments, therefore, students are less keen to experiment with ideas and concepts (Biggs & Tang, 2007 as cited in Surgenor. P, 2010), preferring to ‘play it safe’, giving the answers they believe are expected of them, and banking the marks provided.

There are some reasons behind, according to Knight, Peter. T, (2002), the first is higher education institutions are generally expected to have learning goals that are far more extensive and complex than mastery of subject matter alone, and that are being held to account for student achievement in terms of those goals. Secondly, at the same time, a greater range of assessment techniques has come into currency, which has introduced substantial practical and theoretical problems, with the comparability and aggregation of performances judged by different assessment methods. Thirdly, public sector services are nowadays marked by low-trust management systems, when once there would have been a greater readiness to trust that good people engaged on worthwhile activities would learn the sorts of things that were intended. Assessment is supposed to supply evidence to bridge the trust gap. Fourthly, the eternal concern with value for money has taken a rationalist turn, with the belief that it is prudent to specify objectives, measure inputs, assess performance in terms of those objectives, allocate the next round of resources to efficient providers and apply sanctions to the less efficient.

RESEARCH METHOD

There are three sequence stages in investigating the data. They are collecting, analyzing and presenting the result of result (Sudaryanto, 1993). Descriptive qualitative method was used in analyzing the data. The data was collected from document and interview. Multiple choice tests that used in summative assessment were taken as document data. The researcher analyzed each of multiple-choice items to test its authenticity. Qualitative researchers use interviews to uncover the meaning structures that participants use to organize their experiences and make sense of their worlds (Hatch, J. Amos, 2002: 91). The interview held to gain more information how

teachers design test items and transform the inauthenticity producing tests into more authentic.

A teacher of vocational schools in Bandung participated as subject research. The subject research was selected because of his endeavor to be a professional teacher in pursuing his master degree in one of university in Bandung. Designing authentic summative assessment is one of his big efforts. It means that the subject research is selected through the purposive technique sampling.

FINDINGS AND DISCUSSION

Document Analysis

Table 1.1 Evaluations of authenticity of test items

Evaluation of Authenticity (5 items)	Summative test (30 items)
5	1
4	12
3	5
2	3
Total	21

Table 1.1 shows that twenty-one test items measure as a continuum authentic test items. There are consist of one item achieves all of authenticity’s evaluation, twelve items achieve four of authenticity’s evaluation, five items achieve three of authenticity’s evaluation, and three items achieve two of authenticity’s evaluation.

The one summative test item that achieves all of authenticity’s evaluation is question number ten. The question is “*When was he born?*”. This question is one of three questions from reading comprehension in the test item. Haladyna (2002) adds that “cognitive learning theory and the social/constructivist learning theory seem to favor teaching and testing that happen in a natural context where students see the merits of what they are learning” (p. 194). Nitko and Brookhart (2007), cited by Dickinson *et al*(2007) provide the following perspective on authenticity:

“... the ‘authentic’ in *authentic assessment* usually means presenting students with tasks that are directly meaningful to their education instead of indirectly meaningful. For example, reading several long works and using them to compare and contrast different social viewpoints is directly meaningful because it is the kind of thoughtful reading educated citizens do. Reading short paragraphs and answering questions about the ‘main idea’ or about what the characters in the passage did, on the other hand, is indirectly meaningful because it is only one fragment or component of the ultimate learning target of realistic reading. (p.253)”

The test item achieves the naturalness of language that might occur in the real-life situation. Because of its carefully designed, the test designer considers that item characteristics are the restricted or controlled versions of those contextual features. It determines

the nature of the language performance expected for the given test. The test items were designed based on the basis models of language ability which suit to the target situation as well as the characteristics of the test takers. It is relevant to the model of competence to be tested. Careful examination of the target situations with the help of studies from pragmatics, discourse analysis, interactive sociolinguists also take into account in designing this test item.

There are nine decontextualized test items and three non-thematics organization. Question number eight (Q8) is decontextualized. It represents eight other test items (Q5, Q6, Q7, Q24, Q 25, Q 26, Q27, and Q29) that achieve four of authenticity's evaluation (naturalness, topic situation, thematic, and real-life situation). This sequence is likely to be written or spoken in the real world, but this item was limited by including otherness (Jennifer and Billy in Q8). This item is contextualized without including otherness. It takes into account both the input provided in a test as well as the expected outcome arising from the input by characterizing not only test tasks but also test takers' interactions with these (Lewkowicz, 2000: 49). Three of non-thematic organization items are Q4, Q5 and Q13. The thematic format of the sections, the authentic language within each item, and the contextualization add face validity, interest, and some humor to what might otherwise be a mundane test (Brown, 2004: 54). The format of the items is contrived, thus lowering the level of authenticity because of non-thematic format.

Five items (Q9, Q12, Q20, Q21, and Q22) achieve three of authenticity's evaluation. Q20, Q21, and Q22 show lack of contextualize and non-thematic situation. The absent of naturalness and real world representation is represent by Q9. The Q12 lacks in term of contextual and thematic organization. Thus, they perceive the test to be inauthentic or less authentic.

The three items that achieve two of authenticity's evaluation are Q1, Q2 and Q23. The tasks would not necessarily be either authentic or inauthentic but would lie on a continuum which would be determined by the extent to which the assessment task related to the context in which it would be normally performed in real-life. Chavez (1998), cited by Shomossi&Ketabi (2008) argues that any text that has been taken out of its original context and away from its intended audience automatically becomes 'inauthentic'. Decontextualized of the text in the test items attempt to accept a nonreal-life criterion. However, what was lacking in the equation was a theoretical framework to provide a coherent rationale in identifying and defining the authenticity.

Interview Analysis

Based on the interview, the study found there were some reasons of using multiple choice test item in summative assessment and challenge that was

faced by teachers in transforming the inauthentic producing tests into more pedagogically fulfillment.

Table1.2 multiple-choice in summative assessment and endeavor transform inauthentic to be more authentic.

Code	Text Section	Source	Theme
Practicality and instutusional regulation	Well... Selainkepraktisand alamscoringnyadan karenabisadipakeb uattesapaaja... Jugaemangpermint aanpihakskolahnya gitu...	Q1	Practicality, efficiency and reliability
Economical issue	Praktisdalamhalwa ktuuntukmemeriks alembaranjawabans iswa.	Q2	Time and economical issue
Curriculum and policy maker's authority	Pihaksekolahjugabi samenghematbiaya untukpenilaian.	Q3	Stakeholder and Curriculum cosideratio n
Process of constructing test items	Iya... kitasmua guru nyadikasiedarangit u...	Q4	Teacher endeavor to transform the inauthentic into authentic assessment
Curriculum and syllabus goals/objectives	Bahansoalnya yang dah diajarkan.. Harustercoversemu a...	Q5	Assessment achieve the goals/ objectives of curriculum and syllabus
Explaining context and real life situation	Ada yang buatsendiriadajuga yang ngambildarisoaluji annasional..tapikeb anyakansendiri	Q6	favor testing that happen in a natural context
Engaging students' motivation	Kalokontenkandise suaikansamakuriku lumdansyllabi yang berlaku.	Q7	The use of of authentic assessment
Contextual reading	Karenabuku yang digunakanyahasil karya orang siniyakonteksnnya jadugadisesuaikandeng antempat kita. Tapikadangsaya juga memperkenalkan budaya orang asingitugimana, misalnya di InggrisatauAmerik aitugimanabudaya nyadankeadaannya ... danhabityajuga...	Q8	Naturalness and real life situation

	<p>Oohhyaa... tentusajadengansa ngatsengaja, dari 35 siswasebagianbesarsiswasayacowokhanya 10 orang cewek. Selainsayasendirijugasukasepakbola, saya rasa anak-anakakanlebihsenangdantermotivasi membacasesuatu yang merekasukai.</p> <p>Jelaas doong, sebagaiurang Bandung danwarganegara Indonesia yang baiksayaharusbangga. Yabukanhanyasekedarbanggajugasih.... sayajugaberusaha tukmembuatbutirsoaldalamtesitujaditerlihatseperti... yaaahsepertimereka, muridsayadangan gobrolsamatemannya, mencari tau tentangsosok yang dikaguminya. Jaditidaksepertisedangmengerjakantes. Dalam UAS itusayatidakbisamenilai <i>performance</i> atauh lain yang bersifatkomunikasi secaralangsung.</p>		
--	---	--	--

To determine why multiple choice was used in summative test, from the interview it found that:

1. Vocational school as the higher education institution is generally expected to achieve learning goals/ objectives that are far more extensive and complex that are being held to account for student achievement in terms of those goals. Validity, this feature enables the teacher using multiple-choice items to test a broader sample of course content in a given amount of testing time. Hence, they have to be faithful to the curriculum that charged with developing understandings, skills, self-theories and reactivity. (Q1, Q3, and Q5)
2. By considering curriculum and syllabus, a greater range of assessment techniques has come into currency, which has introduced substantial practical and theoretical problems, with the comparability and aggregation of performances judged by different assessment methods. Summative assessment is usually conducted in the last few weeks of term to see how good students have learned what they were supposed

to have learned. Multiple choice as a large-scale test was used to meet that goal. The results from these assessments are aggregated and used to determine whether a student has fulfilled the specified learning outcomes and may achieve some kind of accreditation from curriculum and syllabus. (Q3)

3. The eternal concern with economical issue has taken a rationalist turn, with the belief that it is prudent to specify objectives, measure inputs, assess performance in terms of those objectives, allocate the next round of resources to efficient providers and apply sanctions to the less efficient. (Q2)

To determine how to transforming the inauthentic producing tests into more pedagogically fulfillment, the study found that public sector services are nowadays marked by low-trust management systems, when once there would have been a greater readiness to trust that good people engaged on worthwhile activities would learn the sorts of things that were intended. Assessment is supposed to supply evidence to bridge the trust gap. So that, "cognitive learning theory and the social/constructivist learning theory seem to favor teaching and testing that happen in a natural context where students see the merits of what they are learning (Q6, Q7, and Q8)

CONCLUSION

This study concludes that summative assessment covers what students have learned and what they were supposed to have learned. Large-scale multiple-choice tests cannot offer much authenticity but expect to achieve learning goals/ objectives that are far more extensive and complex that are being held to account for student achievement in terms of those goals. This multiple choice achieves much practicality in term of four aspects; the first is efficiency. The efficiency in multiple-choice items are amenable to rapid scoring. The next is validity. This feature enables the teacher using multiple-choice items to test a broader sample of course content in a given amount of testing time. Consequently, the test scores will likely be more representative of the students' overall achievement in the course. The third is versatility. It means the adaptability to various levels of learning outcomes, from simple recall of knowledge to more complex levels, such as the student's ability to analyze phenomena, apply principles to new situations, comprehend concepts and principles, discriminate between fact and opinion, interpret cause-and-effect relationships, interpret charts and graphs, judge the relevance of information, make inferences from given data and solve problems. The last one is reliability. The well-written multiple-choice test items are less susceptible to guessing than are true-false test items, and therefore capable of producing more reliable scores.

The lack of authenticity on the multiple-choice test was noted as a disadvantage was higher, to deal with this test designer has big effort to transform the inauthentic produced tests into more pedagogically fulfillment. Although some contexts for assessment may be a step removed from daily classroom life, a teacher should integrate the task undertaken and use of knowledge and skills as they employed in practice.

REFERENCES

- Black, P. (2000) *Research and Development of Language Classroom*, Oxford Review of Education, Vol. 26, pp. 407-419.
- Blagojevic, S. (2013) *Original Text as Authentic ESP Teaching Material-the Case of Philosophy*. Journal of English for Spesific Purposes at Tertiary Level. Vol. 1(1)(2013).
- Breen, M. P. (1985). *Authenticity in the language classroom*. Applied Linguistics, Vol. 6, pp. 60-70.
- Brown, D. H. (2004). *Language Assessment: Principles and Classroom Practices*. New York: Pearson Education.
- Dickinson, et al. (2007). *Summative Assessment Strategies for Statistical Learning: Development, Administration, and Scoring of Authentic and Performance assessment*. IASE /ISI Satellite.
- Doye, P. (1986). *Authenticity in foreign language testing*. Online document retrieved January 15th, 2008 from http://www.eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED350819&ERICExtSearch_SearchType_0=no&acno=ED350819
- Finch, A. E. (2002). *Authentic assessment: Implications for EFL performance testing in Korea*. Secondary Education Research, Vol. 49, pp. 89 - 122.
- Fulcher, G & Davidson, F. (2007). *Language Testing and Assessment*. Canada: Routledge.
- Haladyna, T. M. (2002). *Essentials of Standardized Achievement Testing: Validity and Accountability*. Boston: Allyn and Bacon.
- Jian-lan, B. (2007). *An analysis of authenticity in CET-4 and TEM-8*. Sino-US English Teaching, Vol. 4(2), pp. 28- 33.
- Knight, P. T. (2002). *Summative Assesment in Higher Education: Practices in Disarray*. Studies in Higher Education Vol. 27, No. 3, 2001, Cartax Publishing.
- Kramsch, C. (2012). *Authenticity and Legitimacy in Multilingual SLA*. Critical Multilingualism Studies Vol. 1:1 (2012): pp. 107-128.
- Lewkowicz, J. A. (1997). *Authenticity for whom? Does authenticity really matter?* In: Huhta, A., Kohonen, V., Lurki-Suonio, L., Luoma, S. (Eds.), Current Developments and Alternatives in Language Assessment. Jyväskylä University, Finland, pp. 165±184.
- Lewkowicz, J. A. (2000). *Authenticity in language testing: some outstanding questions*. Language Testing Vol. 17 (1), pp. 43-64.
- Rogers, C & Medley, F. Jr. (1988). *Language with a purpose: using authentic materials in the foreign language class*. Foreign Language Annals, Vol. 21, pp. 567–478.
- Shomoossi, N & Ketabi, S. (2007). *A Critical Look at the Concept of Authenticity*. Electronic Journal of Foreign Language Teaching, Vol. 4(1), pp. 149-155 Retrieved October 25, 2009 from <http://www.eFLT.nus.edu.sg/main.htm>.
- Shomossi & Tavakoli. (2010). *Authenticity and Authentication in Language Testing: An Operational Perspective*. MJAL, Vol. 2(1).
- Sudaryanto. (1993). *Metod dan Aneka Teknik Analisis Bahasa*. Yogyakarta: Duta Wacana University Press.
- Surgenor, P. (2010). *Summative and Formative Assesment*. Teaching toolkit, UCD Teaching and Learning. www.ucd.ie/teaching.
- Taylor, D. (1994). *Inauthentic authenticity or authentic inauthenticity?* TESL-EJ, 1(2), 1–12. Retrieved October 1, 2006, from <http://www-writing.berkeley.edu/TESL-EJ/ej02/a.1.html>.
- Weir, C. J. (1998). *Communicative language testing*. Exeter: University of Exeter Press.
- Williams, M., & Burden, R. L. (1997). *Psychology for language teachers: A social constructivist approach*. Cambridge: Cambridge University Press.