

ASSESSING STUDENTS' LEARNING ACHIEVEMENT: AN EVALUATION

Fadly Azhar

English Study Program of FKIP Riau University

Abstract: This research aimed to evaluate the use of alternative assessment, components of testing, and supplementary assessment by 127 lecturers (purposive sampling) of state and private universities in Pekanbaru, Indonesia in assessing students' learning achievement. The CIPP Evaluation Model focusing on input and process was used as a research design; and two sets of questionnaires. The input factor contained knowledge on alternative assessment and components of testing; the process concerned with the frequency of implementation and supplementary assessment. Research findings found that input factor was at a high level. In process factor, *written, performance, self/peer, portfolio, rubric, validity, reliability, table of specification, test sources, and item analysis* (moderate level); *project, product, diligence, kinship, request, honesty, and try-out* (low level); but *attitude, referenced, domain, participation, and attendance* (high level). However, there was no significant difference on factors of input and process viewed from teaching experience except *project assessment* in terms of academic qualification. The implication of this research was that having higher knowledge, lecturers were encouraged to vary their types of assessment, increase the frequency of implementation of components of testing, and elaborate the factors of supplementary assessment.

Key-words: *alternative assessment, components of testing, supplementary assessment*

INTRODUCTION

In order to have a better quality of education, particularly in students' learning achievement, seems to be the major target of every educational institution including state and private universities in Pekanbaru, Indonesia. One of the reasons for this is that learning achievement will reflect the quality of education provided by the lecturers of those universities. Therefore, the claim on quality testing or quality assessment must be met. To construct quality testing or quality assessment, the following things must be fulfilled: testing or assessment should be varied (known as alternative assessment, classroom-based assessment, or authentic assessment), and should be analyzed through various components (known as components of testing) (Yustisia, 2008; O'Malley & Pierce, 1996; Darling & Hammond, 2000; Andrade & Ying, 2005; Angelo & Cross, 1993; Brown, 2004; Arikunto, 2012; Chase, 1974; Mehrens, 1998; Zunairi, 2008; Shohamy, 1985). In addition, testing or assessment is supposed to be supplemented with other external norms (known as supplementary assessment) (Azhar, 2013a).

Alternative assessment (also called classroom-based assessment, and authentic assessment), in this context can be considered as many kinds of assessments or procedures that can be used to evaluate students' learning achievement. These types of assessments should cover three educational domains: cognitive, affective, and psychomotor. Such assessments as *attitudes assessment* and *self/peer assessment* gain information on students' affective domain; *performance assessment*, *project assessment*, and *product assessment* pursue information on students' skills; *written assessment* evaluates students' cognitive domain; and *portfolio assessment* collects and keeps all information concerning with students' knowledge, attitudes, and skills towards certain courses and functions as the evidence of teaching-learning process (Forgette & Marielle, 2000; Leahy, 2005; Zakaria, 2006; Tola, 2006; Tillema, 2011; Tierney & Marielle, 2004; Popham, 1995).

Parallel to this, components of testing consist of the aspects of *rubric*, *domain*, *referenced*, *item analysis*, *validity*, *reliability*, *try-out*, *table of specification*, and *test-sources* (Hughes, 2003; Depdiknas, 2005; Dickins & Germaine, 1992; Arikunto, 2012; Johar & Ariffin, 2001; Weir, 1993; and Mcnamara, 1996); and *Supplementary assessment*, in addition to this, are such factors as *participation*, *attendance*, *request*, *diligence*, *kinship*, and *honesty* in this context, concerns with factors that can be considered influencing explicitly or implicitly the final scores of students' learning achievement, (Azhar, 2013b).

Research Problems

Lecturing and assessing students' learning achievement legibly and accurately are two of important jobs that must be done by each lecturer including the lecturers of state and private universities in Pekanbaru, Indonesia. Concerning with this, Quality Assurance Unit (QAU) of each university conducts such activities as observation, supervision, evaluation holistically towards all lecturers' activities in lesson plan, lecture implementation, implementation management, planning improvement, evaluation tasks, quiz, mid-term test, full-term test, and final scores (QAU, 2013). However, QAU has not fully evaluated yet to what extent all variants related to learning achievement have met qualified criteria.

Parallel to this, the result of pilot studies on the assessment of learning achievement to fifty students of state and private universities in Pekanbaru, Indonesia showed that lecturers tended to use multiple choice (85%), essays (90%), learning material as test sources (76%), and focus on cognitive domain (87%); moreover, without using table of specification (98%), without using project assessment (82%), without using performance

assessment (63%), and without using product assessment (37%) (Azhar, 2013a). In summary, lecturers have not yet applied various types of assessments, even many of them focusing on written assessment; and only covering cognitive domain.

Refer to those problems, this study aims to evaluate the alternative assessment, the components of testing, and several factors of supplementary assessment that the lecturers have already been familiar with and used in assessing students' learning achievement. Accordingly, the research problems of this study can be formulated as in the following: (a) How good is the cognitive level of the lecturers on alternative assessment and the components of testing?; (b) How often is the implementation of alternative assessment, components of testing, and factors of supplementary assessment conducted by the lecturers?; (c) Is there any significant difference on the cognitive level of the lecturers on alternative assessment and components of testing viewed from the aspects of teaching experience and academic qualification?; and (d) Is there any significant difference on the frequency of implementation of alternative assessment, components of testing, and factors in supplementary assessment viewed from the aspects of teaching experience and academic qualification?

Research Objectives

Refer to the research problems, the objectives of this research are as in the following (a) to identify the knowledge of the lecturers of state and private universities in Pekanbaru, Indonesia on alternative assessment and the components of testing; (b) to prove the frequency of the implementation of alternative assessment, components of testing, and factors of supplementary assessment used by the lecturers of state and private universities in Pekanbaru, Indonesia; (c) to find out whether or not there is a significant difference of the knowledge of the lecturers of state and private universities in Pekanbaru, Indonesia on alternative assessment and components of testing viewed from the aspects of teaching experience and academic qualification; and (d) to find out whether or not there is a significant difference of the frequency of implementation on alternative assessment, components of testing, and factors in supplementary assessment used by the lecturers of state and private universities in Pekanbaru, Indonesia viewed from the aspects of teaching experience and academic qualification.

METHODOLOGY

This study used the CIPP evaluation model (Stufflebeam *et al.*, 1971). CIPP is a Context, Input, Process, and Product. However, two aspects only were investigated, namely input and process. CIPP evaluation model has been used in a variety of educational contexts, including in the assessment of students' learning achievement (Fritz, 1996, Stufflebeam & Shinkled, 1988; Rossi *et al.*, 2004). The aspect of input contains *gender, field of study, academic qualification, teaching experience, attended courses on assessment, types of alternative assessment, and the components of testing*. On the other hand, the aspect of process includes the frequency of implementation of the types of alternative assessment, the components of testing, and several factors of supplementary assessment.

There were two sets of questionnaires used to collect the data for the aspects of input and process. Constructs and items for the two aspects were adapted from past studies as well as theoretical concepts by Ali (2005), Arief, (2005), Baghetto, (2004), Birgin & Baki, (2009), Crooks, (2011), Arikunto, (2012), Depdiknas, (2005), Dickens & Germaine, (1992), Forgette & Marelle, (2000), Hughes, (2003), Johar & Ariffin, (2001), Klenowski, (2011), Leahy *et al.*, (2005), Mcnamara, (1996), Munoto & Meini Sondang, (2006), Petkovskaa, *et al.*, (2010), Popham, (1995), Pusat Penilaian Depdiknas, (2003), Tierney & Marielle, (2004), Tillema, *et al.*, (2011), Tola, (2006), Western & Northern Canadian Protocol for Collaboration in Education, (2006), Yustisia, (2008), Zakaria, (2006), Shohamy, (1985), and Zunairi, (2008). However, the items and constructs of several factors of supplementary assessment were constructed through focus group discussion in cooperation with lecturers of Learning Psychology of Faculty of Education, The University of Riau, Pekanbaru, Indonesia (Azhar 2013b).

FINDINGS AND DISCUSSIONS

The research findings found that the knowledge of the lecturers of state and private universities in Pekanbaru, Indonesia on alternative assessment was at a high level (3.92 - 4.34). The same things also happened to the components of testing (4.04 – 4.15). This score is much higher than that of the cognitive level of junior high school English teachers within Riau Province, Indonesia who got moderate level (3.29 – 3.55) (in Azhar, 2013c). Moreover, this condition seemed to be the opposite of the result of pilot studies which

found that lecturers had not yet applied various types of alternative assessment as well as components of testing in assessing students' learning achievement except in *written assessment* particularly on *multiple choice* and *essays* (Azhar, 2013a).

Parallel to this, Depdiknas, (2005) also stated that in assessing students' learning achievement, lecturers were supposed to use various types of alternative assessment and analyze the quality of that assessment through each component of testing. Even, Yustisia (2008), Tola (2006), Zunairi, (2008), Zakaria (2006), and Baghetto, (2004) emphasized that various types of alternative assessment and components of testing were also supposed to be used not only by primary and secondary school teachers but also lecturers in assessing students' learning achievement at a university level.

The findings in the aspect of process showed that *true-false*, *matching*, *sentence completion*, and *paragraph completion* (written assessment); *speech* and *role-play* (performance assessment); *prototype*, *miniature*, and *blueprint* (product assessment) were at a low level frequency of implementation. Meanwhile, *multiple choice* (written assessment), *quiz*, *interview*, *monologue*, *brainstorming*, *drawing sketches based on an order*, and *demonstrating a scientific process* (performance assessment); *sketch*, *design*, and *graphic/diagram* (product assessment); *suggesting*, *inputting*, *criticizing*, and *proposing* (self/peer assessment); *attitude scale* (attitude assessment); and *working portfolio*, *documentary portfolio*, and *showcase portfolio* (portfolio assessment) were at a moderate level.

However, *open-ended question*, *closed-ended question*, and *essay* (written assessment); *scientific presentation* and *discussion* (performance assessment); *writing scientific articles* (project assessment); '*interaction*, *participation*, and *active contribution*,' '*creative and appreciation*,' '*logical*, *critical*, and *lateral thinking*,' '*learning motivation*, *self-confidence*, and *work-in group*' (attitude assessment) were at a high level.

Refer to this, Azhar (2013c) found that this result was really much more comprehensive compared to the frequency of implementation of classroom-based assessment conducted by junior high school English teachers within Riau Province, Indonesia in which the mean scores showed the moderate level (2.76 – 3.63) particularly for the implementation of various types of alternative assessment. This is due to the differences in terms of academic qualification in which most of lecturers have already got masters degree as well as doctorate degree.

Next, the findings in terms of the frequency of implementation of the components of testing showed that firstly, *true-false*, *matching*, *sentence completion*, and *paragraph completion* were at a low level (1.83 – 2.15) in terms of *try-out*. Secondly, the following aspects such as *analytic*, *holistic*, and *mixture of analytic and holistic* (2.92 – 3.01) in term of rubric; *face*, *concurrent*, *construct*, and *content validity* (2.61 – 3.51) in terms of validity; *equivalent*, *test-retest*, and *split-half method* (2.67 – 2.78) in terms of reliability; *mono skill*, *multi skill*, and *skill-oriented* (3.04 – 3.58) in terms of table of specification; *standardized* (2.73) in terms of test-sources; *item differences*, and *quality of distracters* ((2.83 – 3.29) in terms of analysis; and *multiple choice* (3.07) in terms of try-out were at moderate level.

Thirdly, *cognitive*, *affective*, *psychomotor*, and *mixture of the cognitive, affective, psychomotor* (3.73 – 4.27) in terms of domain; *criterion-referenced* and *norm-referenced* (3.22 – 3.93) in terms of referenced; *content-oriented* (3.72) in terms of table of specification; *lecturer-made test* (4.23) in terms of test-sources; *item-difficulties* (3.67) in terms of analysis were at a high level. Finally, in the aspect of supplementary assessment, *participation-based assessment* and *attendance-based assessment* were at a high level (3.75 – 3.81); *diligence-based assessment* and *honesty-based assessment* were at a moderate level (3.49 – 3.51); while *kinship-based assessment* and *order/request-based assessment* were at a low level (1.62 – 1.81).

This research evidence seemed to be more comprehensive than that of the research activities done by Birgin & Baki (2009), Gansle *et al.* (2006), Munoto & Meini Sondang (2006), Crooks (2011), Segers & Tillema (2011), Klenowski (2011), Petkovskaa, *et al.* (2010), Western and Northern Canadian Protocol for collaboration in Education (2006), and Azhar (2013c). In their research activities, the findings only discussed the advantages, the weaknesses, and the needs for training whenever teachers were encouraged to use these types of classroom-based assessment; even, they did not discuss at all some factors in terms of supplementary assessments either explicitly or implicitly influencing the final scores of students' learning achievement as well as components of testing.

On the other hand, in terms of hypothesis testing, it was found that there was no significant difference in the aspects of knowledge (input), frequency of implementation (process) and supplementary assessment viewed from the aspects of teaching experience and academic qualification; except *project assessment* in terms of academic qualification even though there was a little bit difference in means scores. This is in line with the research findings done by Azhar (2013c) on teachers' teaching experience; but not on academic qualification

in which teachers with bachelor degrees were better in terms of knowledge, attitude, and skill in the implementation of classroom-based assessment, alternative assessment, or authentic assessment than that of teachers with diploma.

CONCLUSION AND SUGGESTION

This study has implications for the lecturers of state and private universities in Pekanbaru, Indonesia. In the first place, they have a high level of knowledge on various types of alternative assessment, components of testing, and supplementary assessment even though they have different academic qualification and teaching experience. However, they have a low level of frequency of implementation of components of testing in *try-out* particularly on *true-false*, *matching*, *sentence completion*, and *paragraph completion*. In terms of various types of alternative assessment, they also have a low level of frequency of implementation particularly on *true-false*, *matching*, *sentence completion*, and *paragraph completion* (written assessment); *speech* and *role-play* (performance assessment); and *prototype*, *miniature*, and *blueprint* (product assessment).

Therefore, it is recommended that they have to take into account the following actions. In terms of various types of alternative assessment particularly on *project assessment*, the lecturers are encouraged to attend peer-teaching activities either in a similar or different academic qualification. Golanaki & Vassilopoulou (2007), Stipeck (2006), and Beyazkurk & Kesner (2005) concluded that through peer-teaching activities, both groups (teacher-students) and (among colleagues) obtained “*adequate internal consistency and low standard error of measurement on conflict, closeness, and dependency*.” So, the lecturers of state and private universities in Pekanbaru, Indonesia, will be able to learn from one to another even sharing ideas in terms of the implementation of *project assessment*, *performance assessment*, and *product assessment*.

Parallel to this, they are also encouraged to use focus group discussion, in which, they can learn, watch, share ideas, and imitate from one to another (Krueger, 1994). In terms of the implementation of *try-out* namely on objective testing, they are encouraged to review the advantages and weaknesses of *try-out* (Hughes, 2003; Dickens & Germaine, 1992; Crooks, 2011; Mcnamara, 1996; Popham, 1995; Shohamy, 1985). In terms of a low level of frequency of implementation particularly on *true-false*, *matching*, *sentence completion*, and *paragraph completion* (written assessment), it can be concluded that this

is due to the fact that probably these types of assessment are rarely used in this era at a university level.

Acknowledgements

Special thanks are given to firstly, Prof. Dr. H. M. Nur Mustafa, M.Pd the Dean of the Faculty of Education and Teacher Training University of Riau, Indonesia who has supported the implementation of this research activity. Secondly, colleagues of Learning Psychology Department of University of Riau, Indonesia for their roles as partners in focus group discussion. Thirdly, lecturers as the sample of this research who have spent much time in completing the questionnaires; and finally, to those who have valuable contribution for the accomplishment of this research.

References

- Ali, Y. S. (2005). An introduction to electronic portfolios in the language classroom. *The Internet TESL Journal*, XI (8). Retrieved August 18, 2012, from <http://iteslj.org>
- Ariev, R. P. (2005). A theoretical model for the authentic assessment of teaching. *PAREonline.net*, 10(2). Retrieved June 28, 2012, from <http://PAREonline.net>.
- Arikunto, S. (2012). *Basics skills on educational evaluation* (Dasar-Dasar Evaluasi Pendidikan). Edisi 2. Jakarta: Bumi Aksara).
- Azhar, F. (2013a). *Students' perception in various types of assessments in Teaching and Learning: a pilot study*. (Persepsi mahasiswa tentang berbagai penilaian dalam pengajaran dan pembelajaran): suatu Kajian Awal. Pekanbaru: FKIP-UR
- Azhar, F. (2013b). *Some factors of supplementary assessment: Focus group discussion with Learning Psychology Lecturers of Faculty of Education, University of Riau, Pekanbaru, Indonesia*. (Faktor-faktor pelengkap penilaian: hasil diskusi kelompok dengan dosen Psikologi Pembelajaran): Pekanbaru: FKIP-UR
- Azhar, F. (2013c). Class-Based Performance Evaluation: An Evaluation. *Asian Social Science*. Vol. 9, No. 12. ISSN 1911-2017; E-ISSN 1911-2025. Canadian Center of Science and Education.
- Baghetto, A. R. (2004). Toward a more complete picture of student learning: Assessing students' motivational beliefs. University of Oregon. Retrieved May 19, 2012, from <http://PAREonline.net>
- Beyazkurk, D. & Kesner, E.J. (2005). Teacher – child relationships in Turkish and United States schools: A cross-cultural study. Retrieved March 12, 2014, from *International Education Journal*, 2005, 6 (5), 547 – 554. <http://iej.cjb.net>.
- Birgin, O. & Baki, A. (2009). An investigation of primary school teachers' proficiency perceptions about measurement and assessment methods: the case of Turkey. *Procedia Social and Behavioral Sciences*, 1: 681-685. Retrieved August 4, 2012, from www.sciencedirect.com
- Chase, I. (1974). *Measurement for Educational Evaluation*. London: Addison-Wesley.
- Crooks, T. (2011). Assessment for learning in the accountability era: New Zealand. *Studies in Educational Evaluation*, 37 (2011): 71-77. ScienceDirect. Retrieved July 20, 2012, from www.elsevier.com/stueduc

- Depdiknas, (2005). *Learning evaluation. Module: self-learning material D-II PGSD program*. Jakarta: Educational Information and communication Technology center.(Evaluasi Pembelajaran. Modul: bahan Belajar Mandiri Program D-II PGSD. Jakarta: Pusat Teknologi Komunikasi dan Informasi Pendidikan.)
- Dickens, R. P. & Germaine, K. (1992). *Evaluation*. New York: Oxford University Press.
- Forgette, G.R. & Marelle, S. (2000). *Organizational Issues Related to Portfolio Assessment Implementation in the classroom*. Retrieved June 21, 2012, from <http://PAREonline.net>
- Fritz, S. (1996). Assessing undergraduate student needs utilizing the CIPP model of evaluation. Dissertation Ph. D. University of Idaho.
- Gansle, A. K. (2006). Elementary school teachers' perceptions of curriculum-based measures of written expression. Retrieved August 17, 2012, from <http://PAREonline.net>
- Golanaki, P.E. & Vassilopoulou, D.H (2007). The student-teacher relationship scale in a Greek sample of preadolescents: reliability and validity data. *Psychology*, 2007, 14 (3). 292 – 310.
- Hughes, A. (2003). *Testing for Language Teachers*. Second Edition. Cambridge: Cambridge University Press.
- Johar, R. A. & Ariffin, R. S. (2001). *Issues on educational measurement and evaluation*. Bangi: Faculty of Education, National University of Malaysia. (Isu Pengukuran dan Penilaian Pendidikan. Bangi: Fakulti Pendidikan, University Kebangsaan Malaysia.)
- Klenowski, V. (2011). Assessment for learning in the accountability era: Queensland, Australia. *Studies in Educational Evaluation* 37 (2011) 78 -83 retrieved July 22, 2012, from ScienceDirect www.elsevier.com/stueduc
- Krueger, R. A. (1994). *Focus Group discussion: A practical guide for applied research*. Newbury Park, CA: Sage Publication.
- Leahy et al. (2005). Classroom assessment minute by minute, day by day. *Educational Leadership*, 63(3). Retrieved June 26, 2012, from <http://search.ebscohost>.
- Mcnamara, T. 1996. *Measuring second language performance*. London: Longman.
- Mehrens, W.A. & Ebel, R.L. 1998. *Principles of Educational and Psychological Measurement*. Chicago: Rand McNally.
- Munoto & Meini Sondang. 2006. *Development of portfolio assessment tools to improve students' achievement motivation in electric circuits I course in the Department of electrical engineering UNESA* (Pengembangan perangkat penilaian portofolio untuk meningkatkan motivasi berprestasi mahasiswa pada mata kuliah Rangkaian Listrik I di Jurusan Teknik Elektro UNESA). *Electrical Engineering Education program, Department of electrical engineering, Faculty of engineering, state University of Surabaya* (Program Studi Pendidikan Teknik Elektro, Jurusan Elektro, Fakultas Teknik, Universitas Negeri Surabaya).
- Petkovskaa, B. et al. 2010. Primary school education standards for student's assessment in primary school. *Procedia Social and Behavioral Sciences* 2: 2366-2370. Retrieved May 18, 2012, from www.sciencedirect.com
- Popham, W. J. (1995). *Classroom Assessment: What teachers need to know*. Boston: Pearson Education Inc.
- Pusat Penilaian Depdiknas. (2003). *Evaluation guidelines in classroom. (Pedoman penilaian di kelas)*. Jakarta: Development and research Board, Department of National Education. (Jakarta: Badan Penelitian dan Pengembangan, Departemen Pendidikan Nasional).

- Quality Assurance Unit, 2013. *Lecturers' performance evaluation sheet*. Pekanbaru: Faculty of Education, University of Riau (Lembar penilaian Kinerja dosen).). Pekanbaru: Fakultas Pendidikan, Universitas Riau.
- Rossi, H.P., Lipsey, W.M., & Freeman, E.H. (2004). *Evaluation. A systematic approach* (7 ed.). London: Sage Publications.
- Segers, M. & Tillema, H. (2011). How do Dutch secondary teachers and students conceive the purposes of assessment? *Studies in Educational Evaluation*, 37: 78-83. Retrieved July, 21, 2012, from ScienceDirect www.elsevier.com/stueduc
- Shohamy, E. (1985). *A Practical handbook in language testing for the second language teacher*. Israel: Tel-Aviv University.
- Stipek, D. (2006). Relationships matter. *Educational Leadership*. September 2006, volume 64, number 1 (46 -49).
- Stufflebeam, D.L. (1971). The relevance of CIPP evaluation model for educational accountability. *Journal of Research and Development in Education Fall*: 19-25.
- Stufflebeam, D. L. & Shinkled, A. (1988). *Systematic Evaluation*. Norwell: kluwer-Nijhof Publishing.
- Tierney, R. & Marielle, S (2004). *What's wrong with rubrics: focusing on the consistency of performance criteria across scale levels*. Retrieved July 6, 2012, from <http://PAREonline.net>
- Tillema, et al. (2011). Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning: A review of research studies. *Studies in Educational Evaluation*, 37, 25-34.
- Tola, B. (2006). *Self-assessment: Module assessment guidelines in the classroom*. Jakarta: Educational Assessment Research and Development, Ministry of National Education (Penilaian diri: modul pedoman penilaian di kelas. Jakarta: pusat Penilaian Pendidikan Badan Penelitian dan Pengembangan, Departemen Pendidikan Nasional.)
- Weir, C. J. (1993). *Understanding and Developing Language tests*. Hemel Hemstead: Prentice Hall.
- Western & Northern Canadian Protocol for Collaboration in Education.(2006). Rethinking classroom assessment with purpose in mind. Retrieved July, 8, 2012, from www.wncp.ca
- Yustisia, T.P. (2008). *Complete guide to curriculum level education units* (Panduan lengkap kurikulum tingkat satuan pendidikan). (Yogyakarta: Pustaka Yustisia).
- Zakaria, R.T. (2006). *Attitude assessment guidelines: Module assessment guidelines in the classroom*. Jakarta: Educational Assessment Research and Development, Ministry of National Education (Pedoman penilaian sikap: Modul pedoman penilaian di kelas. Jakarta: Pusat Penilaian Pendidikan Badan Penelitian dan Pengembangan, Departemen Pendidikan Nasional.)
- Zunairi,(2008). Internal assessment as classroom assessment model (Internal assessment sebagai model penilaian kelas). *Journal Paradigma*, XIII (25).