Assessing Students’ Learning Achievement: An Evaluation

Dr. Fadly Azhar

Faculty of Education, University of Riau, Pekanbaru, Indonesia, Fadly Azhar, Faculty of Education, University of Riau, Pekanbaru, Indonesia.

Email: fadlyazhar57@gmail.com

Doi:10.5901/mjss.2015.v6n2p535

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 served as a research design; and two sets of questionnaires served as data collection. The input factor contained knowledge on alternative assessment and components of testing while the process was concerned with the frequency of implementation and supplementary assessment. The research findings showed that the input factor was at a high level. In the process factor, the aspects of written, performance, self/peer, portfolio, rubric, validity, reliability, table of specification, test sources, and item analysis were at a moderate level; project, product, diligence, kinship, request, honesty, and try-out were at a low level; but attitude, reference, domain, participation, and attendance were at a 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 by having higher knowledge on alternative assessment and components of testing, the lecturers were encouraged to vary their types of assessment, increase the frequency of the implementation of testing components, and elaborate the factors of supplementary assessment.

Keywords: alternative assessment, components of testing, supplementary assessment

1. Introduction

In order to have a better quality of education, students learning achievement is the main target of every educational institution including state and private universities in Pekanbaru, Indonesia. One of the reasons for this is that learning achievement reflects the quality of education of those educational institutions where lecturers are involved in achieving it. In this regard, Quality Assurance Unit (QAU) of each university along with the QAU of Faculty of Education (FKIP) University of Riau, Indonesia needs to conduct such activities as observation, supervision, evaluation holistically towards all lecturers’ activities covering lesson plan, lecture implementation, implementation management, planning improvement, evaluation tasks, quiz, mid-term test, full-term test, and final scores (QAU FKIP UR, 2013). However, QAU FKIP UR has not yet fully evaluated to what extent all variants related to learning achievement have met qualified criteria.

The pilot studies conducted towards 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%). In addition to this, without using table of specification (98%), without using project assessment (82%), without using performance assessment (63%), and without using product assessment (73%) (Fadly Azhar, 2013a). These results provided the evidence that lecturers had not yet applied various types of assessments, even many of them still focused on written assessment and only covered cognitive domain.

Furthermore, literature review written by the following experts including Yustisia, (2008); O’Malley & Pierce, (1996); Darling & Hammond, (2000); Andrade & Ying, (2005); Angelo & Cross, (1993); Brown, (2004); Arikunto, (2012); Chase, (1974); Mehrrens, (1998); Zunairi, (2008); Shoharny, (1985) stated that to construct quality testing or quality assessment, two criteria must be met. Firstly, testing or assessment known as alternative assessment, classroom-based assessment, or authentic assessment should be varied and analyzed through various components known as components of testing. Secondy, testing or assessment needs to be supplemented with other external norms known as supplementary assessment (Fadly Azhar, 2013b).

Alternative assessments in this context which are also called classroom-based assessment or authentic assessment may 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. However, the types of assessments including attitudes assessment and self/peer assessment gain information from students’ affective domain, whereas performance assessment, project assessment, and product assessment pursue information through students’ skills. Written assessment evaluates students’ cognitive domain; and portfolio assessment collects and keeps all information concerning students’ knowledge, attitudes, and skills towards certain courses and functions as the evidence of teaching-learning process (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; Johar & Ariffin, 2001; Weir, 1993; and Monamara, 1996). In addition, supplementary assessment includes such factors as participation, attendance, request, diligence, kinship, and honesty in this respect deals with factors that can be considered influencing the explicitly or implicitly of the final scores of students’ learning achievement, (Fadly Azhar, 2013b).

2. Research 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 are used to assess the students’ learning achievement. The research questions are (a) How good is the cognitive level of the lecturers on alternative assessment and the components of testing? (b) What is the frequency of 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? (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?

3. Research Objectives

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. (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.

4. Methodology

This study used the CIPP evaluation model (Stufflebeam et al., 1971). CIPP stands for Context, Input, Process, and Product. In this study, only two aspects were investigated, namely: input and process. A variety of educational contexts including the assessment of students learning achievement has used CIPP evaluation model (Fritz, 1996; Stufflebeam & Shinkled, 1988). The reason for choosing CIPP evaluation model is due to fact that the major purpose of this research is to evaluate the factors of input and Process. Stufflebeam & Shinkled (1985), Rossi et al. (2004) and Gredler (1996) defined the term of CIPP evaluation model as a methodology focusing on improving rather than on proving the strengths and the weaknesses of a program in terms of objectives, design, implementation, and impacts descriptively. Therefore, CIPP evaluation model is very effective to identify the background knowledge of the sample respondents about assessment (input factor) and the way to implement the assessment (process factor).

The aspect of input covers 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.

Two sets of questionnaires were used to collect the data for both input and process factors. The constructs and items for the two factors were adopted from the previous studies as well as theoretical concepts written by Ali (2005),
Ariev, (2005), Baghetto, (2004), Crooks, (2011), Dickens & Germaine, (1992), Forgette & Marelle, (2000), Hughes, (2003), and Pusat Penilaian Depdiknas, (2003). However, the items and constructs of several factors of supplementary assessment were constructed through focus group discussion in collaboration with the lecturers of Learning Psychology Department of Faculty of Education (FKIP), University of Riau, Pekanbaru, Indonesia (Fadly Azhar, 2013b).

5. Findings and Discussions

The research findings proved 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). The range of the scores was 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) (Fadly Azhar, 2013c). This condition was contradictory to the results of pilot studies which discovered 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 multiple choice and essays (Fadly 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. Yustisia (2008), Tola (2006), Zunairi, (2008), Zakaria (2006), and Baghetto, (2004) even emphasized that it was also important to use various types of alternative assessment and components of testing not only by primary and secondary school teachers but also by lecturers in assessing students’ learning achievement at a university level.

The research findings also identified that the factor of process was at a moderate level. The aspects included in terms of written assessment are true-false, matching, sentence completion, paragraph completion and multiple choices. The aspects included in terms of performance assessment are speech, role-play, quiz, interview, monologue, brainstorming, drawing sketches based on an order, and demonstrating a certain process. The aspects included in terms of product assessment are prototype, miniature, blueprint, sketch, design, and graphic/diagram. The aspects included in terms of self/peer assessment are suggesting, inputting, criticizing, and proposing; meanwhile attitude scale in terms of attitude assessment; and working portfolio, documentary portfolio, and showcase portfolio in terms of portfolio assessment.

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

Comparatively, Fadly 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 were at a the moderate level (2.76 – 3.63) particularly for the implementation of various types of alternative assessment. This was due to the differences in terms of academic qualification whereby most of the lecturers already held masters degree as well as doctorate degree.

Furthermore, the findings of the study in terms of the frequency of the implementation of the components of testing were as follows. The aspects of true-false, matching, sentence completion, and paragraph completion were at a low level (1.83 – 2.15) but multiple choice was at a moderate level (3.07) in terms of try-out.

However, such aspects as analytic, holistic, and mixture of analytic and holistic were at a moderate level (2.92 – 3.01) in term of rubric. Similarly, such aspects as face, concurrent, construct, and content validity were at a moderate level (2.61 – 3.51) in terms of validity; equivalent, test-retest, and split-half method were at a moderate level (2.67 – 278) in terms of reliability. It also happened to mono skill, multi skill, and skill-oriented were at a moderate level (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 were at a moderate level.

In addition, the aspects of 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 aspects of supplementary assessment, participation-based assessment and attendance-based assessment (3.75 – 3.81) were at a high level; diligence-based assessment and honesty-based assessment (3.49 – 3.51) were at a moderate level; while kinship-based assessment and order/request-based assessment (1.62 – 1.81) were at a low level.

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 Fadly Azhar (2013c). Their research findings only discussed the advantages, the weaknesses, and the needs for training whenever teachers were encouraged to use the a fore-mentioned types of classroom-based assessment; even, they did not discuss at all some factors 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 factor), frequency of implementation (process factor) and supplementary assessment viewed from the aspects of teaching experience and academic qualification except project assessment in terms of academic qualification. This is in line with the Fadly Azhar’s (2013c) research findings 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 qualification.

6. Implications and Recommendations

This study has implications for the lecturers of state and private universities in Pekanbaru, Indonesia. In the first place, they could gain a high level of knowledge on various types of alternative assessment, components of testing, and supplementary assessment. It could happen since they got master and doctorate degrees and supported with teaching experience including related training on assessments. However, they have a low level of frequency of the implementation of components of testing in try-out particularly such aspects as true-false, matching, sentence completion, and paragraph completion. Probably, this is due to the time consuming on the aspects of teaching load, research activities, seminars, workshops, and so forth.

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 in terms of written assessment; speech and role-play in terms of performance assessment; prototype, miniature, and blueprint in terms of product assessment. In terms of written assessments, this is due to the fact that majority of lecturers (85%) tended to use multiple-choice and essay (90%); however, only 37% out of them used performance assessment and only 27% chose product assessment (Fadly Azhar, 2013a).

Therefore, it is recommended that lecturers 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) viewed 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 one another and even they can share ideas on the implementation of project assessment, performance assessment, and product assessment. Besides, they are also encouraged to use focus group discussion, in which they can learn, watch, share ideas, and imitate one another (Krueger, 1994). In terms of the implementation of try-out on objective testing, they are encouraged to review the advantages and weaknesses of the 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 (in 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.

7. Acknowledgements

My special thanks are given to firstly, Prof. Dr. H. M. Nur Mustafa, M.Pd, Dean of the Faculty of Education (FKIP) University of Riau, Indonesia for his support of the research activities. Secondly, colleagues of Learning Psychology Department of Faculty of Education (FKIP) University of Riau, Indonesia for their roles as partners in focus group discussion. Thirdly, lecturers who were taken as the sample of this research for their time to complete the questionnaires. Finally, my grateful thanks also go to those who have given invaluable contributions for the accomplishment of this research.
References


Minoto &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 portfolio 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).
