

Malaysian Instructors' Assessment Beliefs in Tertiary ESL Classrooms

Niveen R. M. Elshawa

Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Malaysia

E-mail: niveenshawa@yahoo.com

Ain Nadzimah Abdullah (Corresponding author)

Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Malaysia

E-mail: ain@upm.edu.my

Sabariah Md Rashid

Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Malaysia

E-mail: smrashid@upm.edu.my

Received: 12-03-2017

Accepted: 16-04-2017

Published: 30-04-2017

doi:10.7575/aiac.ijels.v.5n.2p.29

URL: <http://dx.doi.org/10.7575/aiac.ijels.v.5n.2p.29>

Abstract

Language assessment can be a valuable tool for providing information regarding language teaching. Given the importance of assessment that has undergone much change, there are important issues that warrant investigation, particularly those related to language instructors. Understanding the assessment beliefs of ESL instructors, especially at the tertiary level, is important since it can help improve the quality of assessment practices as well. Therefore, this study investigated English language instructors' assessment beliefs in the Malaysian context. This study adopted a cross-sectional research design. The survey method was utilized to collect data from six Malaysian universities using a purposive sampling strategy. English language instructors (n=83) were selected via purposive sampling for the study. Findings of the study revealed that English language instructors believed that the purpose of assessment was to improve teaching and learning. Regarding the assessment beliefs that are related to the assessment purposes, analyses of data showed that the items that received the highest percentage of agreement were diagnosing strength and weaknesses in students, providing information about students' progress and providing feedback to students as they learn, respectively. Although they reported using both formal and informal assessment of their students' work, English language instructors relied heavily on paper and pencil assessment while giving more weightage on formative assessment. The majority of English language instructors reported employing marking schemes for the courses they taught, carrying out sample marking and providing feedback. Finally, English language instructors reported using different types of assessments for every language skill taught in their language unit/center. The findings highlight the fact that English instructors should be more empowered in their role as the assessors of students. Their knowledge about what, how, when to assess should be developed through long professional development courses; one-shot workshops or seminars would not be enough to improve instructors' assessment literacy.

Keywords: Assessment, Assessment beliefs, Tertiary ESL classroom

1. Introduction

In the Malaysian educational context, higher education institutions have developed various programs for their undergraduates and postgraduates to contribute to the nation. Inevitably, implementers of curriculum are the academic staff, who can be considered as an integral resource that is a corner stone for the achievements of an institution. Higher education institutions that aspire to achieve world-class ranking need to recruit and retain the best instructors (Norizan Md et al., 2010). In carrying out their duties, assessment knowledge and training are essential parts of their experience, especially when they take on the role of custodians of the quality performance of their students. Therefore, an instructor's assessment knowledge and competence can be influencing factors in undermining or encouraging students' learning in the classroom. With such a prominent role, assessment and testing issues have begun to witness increasing emphasis in the agenda of higher educational institutions around the world. In recent years, there has been an increasing interest in public accountability, standards and the imposition of more stringent reporting requirements to ensure quality and to meet the educational objectives in the Malaysian context. It has become increasingly difficult to ignore the fact that higher educational institutions have introduced a variety of testing and assessment procedures in order to make decisions on selection, clarification and achievement (Brindley, 2001). These procedures range from the use of

standardized proficiency tests to outcome-based learning systems, which require university instructors to report on learners' progress and achievement against predetermined attainment targets.

1.1 Problem Statement

During the last two decades, some important changes have been identified in the literature on how assessment has changed in higher education (White, 2009). Holroyd (2000) summarized the general patterns of change in seven key findings. The first is the growing concern of how to enhance assessment learning purposes instead of accountability and certification purposes. Next comes an increasing emphasis on formative aspects of assessment rather than end-of-course assessment. The third is the focus on a standardized model for assessment using criterion-referenced assessment and being less focused on a measurement that contains norm-referenced assessment. Another important point is the focus on giving constructive feedback rather than just awarding marks, grades and summary labels. Using multiple methods of assessment rather than depending on one main method— summative assessment- is also an evident change. In addition to this is the use of self-and-peer assessment rather than depending on assessment by teaching staff alone. Holyroyd's recommendation is to consider assessment as part of the teaching process rather than an activity taking place at the end of teaching (White, 2009). Thus, the current study was an attempt to investigate the English language instructors' assessment beliefs compliance with those current and improved trends in assessment in the Malaysian higher education context.

Empirical research on teachers' beliefs and perceptions aims to study beliefs in a wide variety of contexts and to discover the underlying factors that constrain or facilitate these beliefs to be translated into practice. In line with this, cross-cultural research suggest that teachers' conceptions of assessment differ across contexts and these differences reflect teachers' internalization of their society's cultural priorities and practices (Barnes, Fives & Dacey, 2015; Brown & Harris, 2009; Brown, Lake, & Matters, 2009, 2011). Thus, it appears that understanding assessment in the Malaysian tertiary context would provide further insights into cross-cultural differences in teachers' conceptions of assessment reported in the literature.

1.2 Research Questions

Understanding the assessment beliefs of ESL instructors, especially at the tertiary level, is important since it can help improve the quality of assessment practices. In this light, this study investigated English language instructors' assessment beliefs in the Malaysian context. The following research questions were formulated to address this objective:

1. What are the English language instructors' beliefs about the purposes of assessment?
2. What are the English language instructors' beliefs about methods and techniques of assessment?
3. What are the English language instructors' beliefs about feedback, grading and reporting of grades?
4. What are the English language instructors' beliefs about types of the four English language skills' assessment (Reading, Writing, Listening, and Speaking)?

2. Literature Review

Interest in assessment research in higher education contexts has significantly increased in recent years. Several studies were found in literature on faculty assessment beliefs/perceptions in the EFL/ESL classrooms and in different tertiary contexts from 2004 to 2016. In their study on the beliefs about the value of assessment and evaluation held by ESL/EFL instructors in three different contexts: Canada, Hong Kong, and Beijing, Rogers, Cheng and Hu (2007) reported that the beliefs expressed by the instructors in these three different contexts were somewhat mixed, uncertain, and, at times, contradictory. This was particularly evident about the use of paper-and-pencil and performance assessments, the time required for assessments and evaluations and their understanding of and preparation for assessment and evaluation. Moreover, this study revealed the differences in the training they had received and in their confidence in applying what they had learned about assessment and evaluation. Beyond that, judging and scoring student performance and reporting final course grades were reported in Cheng and Wang (2007). Seventy-four ESL/EFL university teachers were interviewed from seven universities in Canada, Hong Kong, and China. The researchers concluded that, in spite of contextual differences, most ESL/ EFL teachers employed self-designed marking criteria for the courses they taught. Further, they tended to design those marking criteria before they assessed their students. However, assessment seemed to be done on the students rather than with them. Differences exist across the three contexts in terms of grading practices and providing feedback.

In keeping with the growing interest in the assessment practices employed in ESL/EFL classrooms, Cheng, Rogers and Wang (2008) explored the same university ESL/EFL instructors' assessment practices but focused on the methods they used, why they used them, who developed them and when they were used. The findings revealed a considerable number and variety of assessments conducted by instructors in the three different contexts. The findings also revealed that a relationship exists between the instructional contexts of an ESL/EFL program and the assessment methods used. The differences among the three contexts are reflected somewhat in the assessment methods developed or chosen by the instructors. This, in part, was related to the variant assessment purposes of the university instructors in the three contexts, which determined their choice of assessment methods and when each method was used during instruction. On

the other hand, the assessment practices used appear to be influenced more by the nature of the instructional context and purposes of the assessment and less by the instructors' views of the relative advantages and disadvantages of the two types of assessment methods. In another context, Yang (2011) investigated the extent to which tertiary EFL teachers implemented a variety of assessment tasks. The findings indicated a variety of test techniques implemented, but varied in the frequency of each task used.

In another study, a survey was conducted in the USA, (Shohamy, Inbar-Lourie, & Poehner, 2008) to examine issues relating to ALP (Advanced Language Proficiency) and its assessment in classroom settings. The results showed that most teachers preferred using assessment for formative rather than summative purposes. Interestingly, "alternative" forms of assessment, including portfolios and performance-based assessments, were seen as invaluable when working with ALP learners. The teachers felt that alternative assessment was the best way for assessing their students. In terms of perceptions of assessment, they believed that assessing students through an on-going process with a formative dimension was the best.

Another study found in the literature that explored teachers' assessment knowledge and practice was one by Xu and Liu (2009). This study revealed the assessment experiences of one teacher participant, suggesting that his/her knowledge is not a static end product, but a highly complex, dynamic, and ongoing process. Moreover, this study confirmed that teachers' knowledge of assessment developed on a temporal continuum and that their assessment practice was by no means uniform, standardized and consistent.

In a study conducted on the Malaysian context, Zubairi, Sarudin and Nordin (2008) investigated the competency of faculty members on assessment based on the six categories of assessment competency in IIUM – International Islamic University of Malaysia. The study found that the use of alternative assessment and/or performance assessments was not common among faculty staff involved in the study. The findings highlighted the need to conduct training for alternative assessments to improve the faculties' staff practices to include more alternative assessments than traditional tests.

Although these studies have been carried out in different contexts of EFL/ESL tertiary contexts, studies which adequately cover the Malaysian ESL context are still lacking. Thus, this study includes a new context, Malaysia and examined the beliefs of ESL English language instructors in the tertiary context of six universities in the state of Selangor.

3. Method

This study adopted a cross-sectional research design. The survey methodology was utilized to collect the data of the study. A questionnaire was developed, validated, and administered to the English language instructors to respond to 67 items based on their beliefs about assessment. The data gathered from the respondents were analyzed quantitatively and descriptively.

3.1 Participants

The respondents of this study comprised 83 English language instructors teaching English proficiency courses at 6 Malaysian universities in the state of Selangor. These participants were selected based on the following 3 criteria. The first criterion was that the respondents had to come from universities with a language centre/unit for teaching English proficiency courses. The second was that the universities were all located in the state of Selangor, while the last criterion was that the participants were required to be English Proficiency course instructors. The study's pool of respondents comprised a breakdown of 42 junior instructors, who had one to ten years of ESL teaching experience (50.6 %) and 41 experienced instructors, who had eleven to twenty years of ESL teaching experience (49.4 %). As for the instructors' academic qualifications in TESL, 44 respondents (53%) had a bachelor's degree, a master, or a PhD degree in TESL, while 39 (47%) had either a bachelor's, master's or a PhD degree in general English studies or literature. Regardless of the title of the course/module they taught (the instructors taught English for proficiency courses), the number of courses that they were teaching differed. Forty-six (55.4%) taught one to two courses, while 37 (44.6%) taught more than two courses.

The picture of the context of teaching English proficiency courses was further illustrated by focusing specifically on the average number of students in the classes. There was no evidence of marked differences in the average number of students in each class. Forty-seven instructors (56.6%) taught classes averaging between 10 to 25 students, while 36 instructors (43.4%) taught classes with more than 25 students. Regarding the instructors' preparation for assessing their students, the majority of the instructors (85.5 %) reported having knowledge about assessment and evaluation. The topics of assessment and evaluation had been reported as either part of one full course, or part of a workshop for 52 instructors (62.7%), while 19 (22.8%) indicated that they had taken more than one assessment training. However, only a small number of respondents (12; 14.5%) reported not having training in assessment and evaluation.

3.2 Instrument

Two sets of questionnaires were developed for this study. To generate the initial item pool for the two questionnaires, the researchers conducted an extensive review of existing literature about teachers' assessment beliefs and classroom practices, specifically, regarding assessment in the tertiary EFL/ESL context. The items were either drawn or adapted from previous similar questionnaires and studies in the literature of ESL classroom assessment (Brindley, 2001; Cheng et al., 2004; Cheng et al. 2008; Cheng and Wang, 2007; Cizek et al., 1995; Roger et al. 2007; Shohamy et al., 2008; Xu

and Liu, 2009; Yang, 2011). An analysis of the information provided by previous research enabled the researcher to generate and adapt statements that could be used in the instrument.

In order to ensure the validity of the new instrument, six collaborative meetings with one professor and two associate professors who are experts in English language assessment and Applied Linguistics were conducted. All the three professors felt that the two questionnaires were appropriate and comprehensive for the purpose of this study. The experts provided guidance on “the wording of questions, the structure of questions, the response alternatives, the ordering of questions, instructions to interviews for administering the questionnaire, and the navigational rules of the questionnaire” (Groves et al., 2009, p. 260). After the validation process, the questionnaire was administered to 12 instructors and no other items were added. However, their feedback reduced the number of items to 67 as the instructors confirmed that the items were clear and readable, but the survey on a whole was long. The overall internal reliability of this questionnaire was 0.83. Johnson and Christensen (2012) stated, “a popular rule of thumb is that the size of coefficient alpha should be, at a minimum, greater than or equal to .70 for research purposes” (p. 142). Thus, the developed questionnaire (Appendix) has been established as an acceptable and a usable questionnaire for gathering the required quantitative data.

The final questionnaire consisted of 67 four-point Likert scale items. Responses ranged from 1 (referring to “Strongly Disagree”) to 4 (“Strongly Agree”). Thus, higher mean scores were later interpreted as high levels of English language instructor agreement with the statement/s reflected by each item mean score or subscale total score. On the other hand, lower mean scores indicated less English language instructor agreement with the statement/s. In other words, higher mean scores indicate a more positive belief or view about different aspects of assessment. The first part of the questionnaire had a section on the respondents’ demographic information. The first section of the questionnaire items was designed for eliciting instructors’ beliefs about assessment purposes (10 items). A total of (17 items) were designed to explore the instructors’ beliefs about methods and techniques of assessment. The questionnaire also covered information on instructors’ beliefs about feedback, grading and reporting of grades (14 items). Finally, the questionnaire ended with 26 more items eliciting information on the beliefs of ESL instructors about different types of assessment of English language skills.

3.3 Data Analysis Method

SPSS (Version 21) was used to analyze the data. To answer the research questions, descriptive statistics including frequencies, percentages, means and standard deviation (SD) were used to report descriptive data. Initially, the Assessment Beliefs questionnaire was divided into two sections: background information and the questionnaire items. Frequency distributions were reported using percentages to summarize respondents’ background information including: TESL qualification, years of teaching, number of courses taught, class size and finally, sources of assessment training (Background information page). For questionnaire data, frequency distributions, means and SDs were used to summarize overall assessment beliefs of English language instructors.

4. Results

This section presents and discusses the results based on the research questions.

4.1 ESL Instructors’ Beliefs about the Purposes of Assessment

The analysis of the data from the first section of the questionnaire revealed that the English language instructors believed that assessment should be used for different purposes, which are discussed in the following related subsections.

4.1.1 Informing Instruction

The overall mean of 3.36 and standard deviation of .49 shown in Table 1 indicate that the majority of the respondents seemed to have positive beliefs/views towards using assessment to inform instruction. The results also indicate that most respondents show high agreement with item 5 (95.2%, M=3.43, SD=.59), item 1 (95.2%, M=3.41, SD=.59), item 2 (94%, M=3.34, SD=.59) and item 4 (89.2%, M=3.25, SD=.64). Overall, among the eighty three participants, item 5 in the subscale of beliefs about the assessment purposes was ranked highest followed by items 1, 2 and 4, that is, the highest mean scores were the items which were relevant to providing information about students’ progress (item 5), helping to focus teaching (item 1), helping to group students for instructional purposes (item 2) and diagnosing strength and weaknesses in teaching (item 4). It seems that the English instructors believed that assessment should be used for informing instruction that would improve their students’ learning. Table 1 shows the frequencies, percentages, means, as well as standard deviations of English language instructors’ beliefs about the instructional purposes of assessment.

4.1.2 Improving Learning

The overall mean of 3.22 and standard deviation of .45 presented in Table 2 show that the majority of the respondents reported a high level of agreement with the statements in this subscale, which suggests that the instructors tended to have positive views towards using assessment to improve learning. The results also indicate that most participants show high agreement with item 3 (97.6%, M=3.43, SD=.55), item 9 (95.2%, M=3.39, SD=.58), item 8 (91.6%, M=3.24, SD=.6), item 10 (88%, M=3.23, SD=.65) and item 7 (81.9%, M=3.08, SD=.67). However, this percentage of agreement dropped to approximately two-thirds on item 6 (73.5%, M=2.94, SD=.69). Overall, among eighty-three participants,

Table 1. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about the Instructional purposes of Assessment (n = 83, Overall Mean = 3.36, SD = .49)

Ranked position	Item	D F (P)	A F (P)	Mean	SD
	Assessment ...				
1.	provides information about students' progress. (5)*	4 (4.8%)	79 (95.2%)	3.43	.59
2.	helps to focus teaching. (1) *	4 (4.8%)	79 (95.2%)	3.41	.59
3.	helps to group students for instructional purposes. (2)*	5 (6.0%)	78 (94.0%)	3.34	.59
4.	diagnoses strength and weaknesses in teaching.(4)*	9 (10.8%)	74 (89.2%)	3.25	.64

Note: The number in the brackets with the asterisk (*) represents the item number in the questionnaire. D= "Strongly Disagree" and "Disagree"; A= "Agree" and "Strongly Agree"; F = "Frequency"; P = "Percentage"

item 3 in the subscale of beliefs about improving learning was ranked highest followed by items 9, 8, 10, 7 and 6, that is, the highest mean scores were for the items which were relevant to diagnosing strengths and weaknesses in students (item 3), providing feedback to students as they learn (item 9), motivating students to learn (item 8), determining the student's mastery of their learning (item 10), and creating a valuable learning experience for students (item 7). However, only about two thirds of the instructors agreed that assessment created competition among students (item 6).

It seems that English instructors believed that assessment should be used to improve students' learning through several approaches. However, they did not support using assessment to create competition among students. This could be related to the nature of the context, which is tertiary education. Table 2 shows the frequencies, percentages, means, as well as standard deviations of English language instructors' beliefs about student-centered purposes of assessment.

Table 2. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about the student-centered purposes of Assessment (n = 83, Overall Mean = 3.22, SD = .45)

Ranked position	Item	D F (P)	A F (P)	Mean	SD
	Assessment ...				
1.	diagnoses strengths and weaknesses in students.(3)*	2 (2.4%)	81 (97.6%)	3.43	.55
2.	provides feedback to students as they learn.(9)*	4 (4.8%)	79 (95.2%)	3.39	.58
3.	motivates my students to learn.(8)*	7 (8.4%)	76 (91.6%)	3.24	.60
4.	determines the students' mastery of their learning.(10)*	10 (12.0%)	73 (88.0%)	3.23	.65
5.	creates a valuable learning experience for my students.(7)*	15 (18.1%)	68 (81.9%)	3.08	.67
6.	creates competition among students.(6)*	22 (26.5%)	61 (73.5%)	2.94	.69

Note: The number in the brackets with the asterisk (*) represents the item number in the questionnaire. D= "Strongly Disagree" and "Disagree"; A= "Agree" and "Strongly Agree"; F = "Frequency"; P = "Percentage"

4.2 ESL Instructors' Beliefs about Methods and Techniques of Assessment

This section presents the results that relate to the beliefs that the English language instructors held about the appropriate methods and techniques of assessment. The presentation and the discussion of the results are organized according to the subsections of Section B of the Assessment Beliefs Questionnaire.

4.2.1 Beliefs about Assessment Format

Table 3 shows the frequencies, percentages, means, as well as standard deviations of English language instructors' beliefs about assessment formats.

Table 3. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about Assessment Formats (n = 83, Overall Mean = 3.02, SD = .39)

Ranked position	Item	D F (P)	A F (P)	Mean	SD
1.	Formal assessment provides a good evaluation of students' work.(11)*	8 (9.6%)	75 (90.4%)	3.27	.63
2.	Assessment questions should reflect real life language use.(15)*	9 (10.8%)	74 (89.2%)	3.31	.66
3.	Informal assessment provides a good evaluation of students' work.(12)*	13 (15.7%)	70 (84.3%)	3.11	.64
4.	Paper and pencil assessment is the best method in evaluating students' work.(13)*	49 (59.0%)	34 (41.0%)	2.37	.68

Note: The number in the brackets with the asterisk (*) represents the item number in the questionnaire. D= "Strongly Disagree" and "Disagree"; A= "Agree" and "Strongly Agree"; F = "Frequency"; P = "Percentage"

As shown in Table 3, the purpose of items 11, 12, 13 and 15 was to address English language instructors' beliefs about the preferred assessment format or procedure. The mean scores of responses to these items ranged from 2.37 to 3.27, that is, the highest mean scores were for the items which were relevant to providing formal assessment as a good evaluation of students' work (item11), using assessment questions that reflect real life language use (item 15), and finally, providing informal language assessment (item 12). Interestingly, however, more than half of the participants believed that paper and pencil assessment was not the best method for evaluating their students' work. In terms of assessment format, items 11 and 15 were ranked highest followed by items 12 and 13. In other words, the highest mean scores were for the items which support formal assessment and real life situations, leaving the traditional paper based assessment to rank the lowest. This suggests that the participants seemed to be in favor of formal assessment, in that it provides better assessment than informal assessment.

4.2.2 Beliefs about Sources to Construct Assessment Items /Tasks

Table 4 shows the frequencies, percentages, means, as well as standard deviations of English language instructors' beliefs about sources to construct assessment items / tasks.

Table 4. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about the sources used to construct assessment items/tasks (n = 83, Overall Mean = 2.98, SD = .44)

Ranked position	Item	D F (P)	A F (P)	Mean	SD
1.	Assessment items are best prepared collaboratively.(18)*	9 (10.8%)	74 (89.2%)	3.30	.66
2.	The best assessment items are the ones developed by the language instructor.(17)*	12 (14.5%)	71 (85.5%)	3.20	.71
3.	Computer technology helps in assessing students' work.(14)*	11 (13.3%)	72 (86.7%)	3.07	.58
4.	Ready-made assessment items found on the internet are a good source for assessing language use.(19)*	38 (45.8%)	45 (54.2%)	2.67	.73
5.	Assessment items from published textbooks are a better source for assessing language use than those found on the internet.(20)*	36 (43.4%)	47 (56.6%)	2.65	.77

Note: The number in the brackets with the asterisk (*) represents the item number in the questionnaire. D= "Strongly Disagree" and "Disagree"; A= "Agree" and "Strongly Agree"; F = "Frequency"; P = "Percentage"

As can be seen in Table 4, the purpose of items 14 and 17-20 were to explore English language instructors' beliefs about the preferred sources to construct assessment items/tasks. The mean scores of the responses to the items ranged from 2.65 to 3.3 on a 4-point Likert scale. The results in the table above indicate that the majority of the participants agreed on item 18 (89.2%, M=3.3, SD=.66), item 17 (85.5%, M=3.2, SD=.71) and item 14 (86.7%, M=3.07, SD=.58).

The highest mean scores were for the items which were relevant to preparing assessment tasks collaboratively (item 18), developing assessment items by the language instructors (item 17), and finally, using computer technology in assessing students' work (item 14).

Interestingly, however, only about half of the participants believed that ready-made assessment items either found on the internet or extracted from textbooks were good sources for assessing students' language use (54.2%, $M=2.67$, $SD=.73$) and (56.6%, $M=2.65$, $SD=.77$). In other words, the English language instructors seemed to view preparing assessments collaboratively more positively than preparing them individually or utilizing ready-made assessments from other available sources (e.g. the internet or textbooks). This suggests that collaborative effort amongst colleagues in preparing assessment tasks is regarded beneficial in the construction of assessment items/tasks.

4.2.3 Beliefs about Types of Assessment

As seen in Table 5, descriptive statistics for items 16 and 24-27, which address different types of assessment, indicate that all English language instructors believe in the need to use a variety of assessment methods (100%, $M=3.63$, $SD=.49$). The best method of doing so was reported as subjective testing (94%, $M=3.29$, $SD=.57$), followed by objective testing (75.9%, $M=2.92$, $SD=.67$). Interestingly, however, the mean scores as well as the frequency percentages for self-and-peer assessments were identical, constituting about two thirds of the participants with agreeing responses.

Taken together, these results indicate that the participants were quite convinced that different types of assessment must be utilized when assessing students. When doing so, however, they believed that priority should be given to subjective testing. Nevertheless, the participants seemed to believe that using self-and-peer assessments to help assess objective tests rather than subjective should be adapted. The lowest mean scores were of the items which were relevant to considering self-assessment as a good method of assessment (item 26) and considering peer-assessment as a good method of assessment (item 27). Table 5 shows the frequencies, percentages, means, as well as standard deviations of English language instructors' beliefs about types of assessment.

Table 5. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about types of assessment (n = 83, Overall Mean = 3.11, SD = .4)

Ranked position	Item	D F (P)	A F (P)	Mean	SD
1.	Language instructors need to use a variety of assessment methods to assess students.(16)*	0 (0%)	83 (100%)	3.63	.49
2.	Subjective testing (e.g. journal entry, portfolio, short essay, sentence completion, reflective task) is a good method of assessment.(25)*	5 (6.0%)	78 (94.0%)	3.29	.57
3.	Objective testing (e.g. matching items, multiple-choice items, true - false items, cloze items) is a good method of assessment.(24)*	20 (24.1%)	63 (75.9%)	2.92	.67
4.	Self-assessment by the student is a good method of assessment.(26)*	20 (24.1%)	63 (75.9%)	2.87	.66
5.	Peer-assessment is a good method of assessment.(27)*	21 (25.3%)	62 (74.7%)	2.87	.71

Note: The number in the brackets with the asterisk (*) represents the item number in the questionnaire. D= "Strongly Disagree" and "Disagree"; A= "Agree" and "Strongly Agree"; F = "Frequency"; P = "Percentage"

4.2.4 Beliefs about Time of Preparing and Conducting of Assessment

As shown in Table 6, the purpose of items 21-23 was to ask about the beliefs of the English language instructors about the appropriate time for conducting assessments as well as the design of test specification forms. Interestingly, the results indicate that the majority of the instructors believed in the need to design a test specification before conducting any type of assessment (95.2%, $M=3.42$, $SD=.63$). However, when reporting about the time of conducting assessments their responses differed to some extent. More than two thirds of the participants (79.5%, $M=2.95$, $SD=.71$) viewed formative assessment as a better means of assessing students than the summative (62.7%, $M=2.67$, $SD=.68$). Table 6 shows the frequencies, percentages, means as well as standard deviations of English language instructors' beliefs about time of preparing and conducting assessment.

Table 6. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about time of preparing and conducting of assessment (n = 83, Overall Mean = 3.02, SD = .44)

Ranked position	Item	D F (P)	A F (P)	Mean	SD
1.	Designing test specifications before a test is constructed is important.(23)*	4 (4.8%)	79 (95.2%)	3.42	.63
2.	The best means of assessing language is through formative assessment.(21)*	17 (20.5%)	66 (79.5%)	2.95	.71
3.	The best means of assessing language is through summative assessment.(22)*	31 (37.3%)	52 (62.7%)	2.67	.68

Note: The number in the brackets with the asterisk (*) represents the item number in the questionnaire. D= “Strongly Disagree” and “Disagree”; A= “Agree” and “Strongly Agree”; F = “Frequency”; P = “Percentage”

Overall, the analysis of the data from Section B of the assessment belief questionnaire revealed that the English language instructors seemed to be in favor of formal assessment rather than informal. In addition, they believed that working collaboratively in preparing test specifications and developing different types of formative and summative assessment tasks to report the final grades of students is advantageous.

4.3 ESL Instructors' Beliefs about Feedback, Grading and Reporting of Grades

This section presents results that relate to the English language instructors' beliefs about feedback, grading and reporting of grades. The presentation and the discussion of the results is organized according to the subsections of Section C of the Assessment Beliefs Questionnaire.

4.3.1 Beliefs about Components of Final Grades

As shown in Table 7, the mean scores for items 28-30, which comprise this subsection, ranged from 2.00 to 3.42 on a 4-point Likert scale. The majority of the respondents reported high levels of agreement with item 30 (96.4%, M=3.42, SD=.61), i.e., they seemed to have positive beliefs/views towards using coursework, tests and examinations in reporting the final grades of the students. However, there is an obvious difference in the instructors' responses to items 28 and 29, i.e., about their belief concerning reliance on tests or coursework only and on test and exams only. The results indicate that only a minority of participants indicated that students' final grades should be based on either tests and exams or coursework only with resulting values of 14.5%, M=2.00, SD=.62 and 12.0%, M=2.00, SD=.58, respectively.

Taken together, these results provide evidence that the English language instructors in the Malaysian tertiary context tended to negatively view the use of only one source of assessment in reporting the final grades of students. Rather, they believed that it was better to use multiple sources to report the final grades of students. Table 7 below shows the frequencies, percentages, means, as well as standard deviations of English language instructors' beliefs about components of final grades.

Table 7. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about components of final grades (n = 83, Overall Mean = 2.47, SD = .44)

Ranked position	Item	D F (P)	A F (P)	Mean	SD
1.	Students' final grades should be based on coursework, tests and examinations.(30)*	3 (3.6%)	80 (96.4%)	3.42	0.61
2.	Students' final grades should be based on coursework only.(29)*	73 (88.0%)	10 (12.0%)	2.00	0.58
3.	Students' final grades should be based on tests and exams only.(28)*	71 (85.5%)	12 (14.5%)	2.00	0.62

Note: The number in the brackets with the asterisk (*) represents the item number in the questionnaire. D= “Strongly Disagree” and “Disagree”; A= “Agree” and “Strongly Agree”; F = “Frequency”; P = “Percentage”

4.3.2 Beliefs about Marking Schemes/Criteria

Based on the overall mean of 3.49 and standard deviation of 0.5 displayed in Table 8, the results showed that the majority of the respondents seem to have a positive belief/view towards using and preparing marking schemes. The results in the table indicate that almost all participants agreed on item 31 (98.8%, M=3.52, SD=.53). A majority of the

participants agreed or strongly agreed on item 33 (96.4%, $M=3.47$, $SD=.61$) and item 32 (94%, $M=3.39$, $SD=.6$). The overall response to the items in this subsection was very positive (viewing constructing and utilizing a marking scheme as a preliminary procedure for conducting of assessment). Table 8 shows the frequency, percentage, mean as well as standard deviations of English language instructors' beliefs about constructing a marking scheme clearly illustrates the findings.

Table 8. Frequency, Percentage, Mean and Standard Deviation of the Assessment Beliefs about marking scheme/criteria ($n = 83$; Overall Mean = 3.46, $SD = .50$)

Ranked position	Item	D F (P)	A F (P)	Mean	SD
1.	A marking scheme should be prepared before assessment is given.(31)*	1 (1.2%)	82 (98.8%)	3.52	.53
2.	Sample marking would help in improving marking criteria.(33)*	3 (3.6%)	80 (96.4%)	3.47	.61
3.	Sample marking should always be carried out.(32)*	5 (6.0%)	78 (94.0%)	3.39	.60

Note: The number in the brackets with the asterisk (*) represents the item number in the questionnaire. D= "Strongly Disagree" and "Disagree"; A= "Agree" and "Strongly Agree"; F = "Frequency"; P = "Percentage"

4.3.3 Beliefs about Giving Feedback and Reporting Final Grades

The purpose of items 34–37 and 40 in Table 9 was to explore English language instructors' beliefs about giving feedback and reporting final grades. The mean scores of the items ranged from 2.57 to 3.52 on a 4-point Likert scale. From the data in the table, it is apparent that the majority of the participants agreed on item 37 (97.6%, $M=3.52$, $SD=.59$), item 34 (94%, $M=3.33$, $SD=.59$) and item 35 (84.3%, $M=3.0$, $SD=.64$). The highest mean scores were for the items on giving feedback to students after assessment (item 37), and conferencing with students when giving feedback (item 34). About two-thirds of the participants, however, believed that students should be given back their results no later than a week after the assessment (71.1%, $M=2.86$, $SD=.78$).

Interestingly, however, when participants were asked the question regarding their beliefs about whether a letter grade is better than a percentage score as a performance indicator (item 36), more than half of the language instructors showed disagreement with the statement in item 36 (53%, $M=2.57$, $SD=.7$), thus, indicating greater support for percentage score as a better performance indicator than a letter grade.

Table 9. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about giving feedback and reporting final grade ($n = 83$, Overall Mean = 3.05, $SD = .39$)

Ranked position	Item	D F (P)	A F (P)	Mean	SD
1.	Students should be given feedback after assessment.(37)*	2 (2.4%)	81 (97.6%)	3.52	.59
2.	Conferencing with students is a good way of giving feedback during the course.(34)*	5 (6.0%)	78 (94.0%)	3.33	.59
3.	Criterion-referenced assessment is better than norm-referenced assessment.(35)*	13 (15.7%)	70 (84.3%)	3.00	.64
4.	Students should be given back their assessment results no later than a week after the test.(40)*	24 (28.9%)	59 (71.1%)	2.86	.78
5.	Giving a letter grade (e.g. A, B) is better than a percentage score as a performance indicator.(36)*	44 (53.0%)	39 (47%)	2.57	.70

Note: The number in the brackets with the asterisk (*) represents the item number in the questionnaire. D= "Strongly Disagree" and "Disagree"; A= "Agree" and "Strongly Agree"; F = "Frequency"; P = "Percentage"

4.3.4 Beliefs about Students' Role in the Marking Process

The purpose of items 38, 39 and 41 in Table 10 were to explore English language instructors' beliefs about students' roles in the marking process. The mean scores of the responses to the items ranged from 2.25 to 3.31 on a 4-point Likert

scale. From the data in the table, it is apparent that the majority of the participants agreed or strongly agreed on item 38 (92.8%, $M=3.31$, $SD=.64$) and item 41 (86.7%, $M=3.23$, $SD=.59$), which indicates the instructors' support for the need to inform students about the marking criteria or the mark allocation of any given test. Nonetheless, when participants were asked regarding their belief about involving students in preparing the marking criteria (item 39), more than two-third of the participants showed disagreement with the statement for that item (72.3%, $M=2.25$, $SD=.82$).

Table 10. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about students' role in the marking process ($n = 83$, Overall Mean = 2.93, $SD = .5$)

Ranked position	Item	D		A		Mean	SD
		F	(P)	F	(P)		
1.	Students should be informed about the marking criteria before being assessed.(38)*	6	(7.2%)	77	(92.8%)	3.31	0.64
2.	Marks allocation for each test question should be made known to students.(41)*	11	(13.3%)	72	(86.7%)	3.23	0.67
3.	Students should be involved in preparing the marking criteria.(39)*	60	(72.3%)	23	(27.7%)	2.25	0.82

Note: The number in the brackets with the asterisk (*) represents the item number in the questionnaire.

D= "Strongly Disagree" and "Disagree"; A= "Agree" and "Strongly Agree"; F = "Frequency"; P = "Percentage"

4.4 ESL Instructors' Beliefs about Types of English Language Skills Assessment

This section presents results that relate to the English language instructors' beliefs about different types of English language skills assessment. For three language skills (reading, writing and listening), the assessment types were basically divided into two groups, traditional and alternative, except for speaking skill which was dealt with, according to the literature, by using one type of assessment item. The presentation of the results is organized according to the various aspects related to the Assessment Beliefs Questionnaire (Section D).

4.4.1 Beliefs about Types of Reading Skill Assessment

Descriptive statistics including frequencies, percentages, means and standard deviations were computed to explore the rank order of English language instructors' reports for their beliefs regarding traditional as well as alternative assessment techniques when assessing reading skill (Table 11).

It is apparent from the table that the most favorable type of assessment in the given list of traditional types of reading skill assessment reported by the respondents were multiple-choice items and true-false items (89.2%, $M=3.20$, $SD=.88$). This was followed closely by matching items (86.7%, $M=3.1$, $SD=.84$) and cloze items (84.3%, $M=3.07$, $SD=.88$). The smallest percentage of agreement was reported on sentence completion items (74.7%, $M = 2.80$, $SD = 1.04$).

Table 11. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about Traditional Types of Reading Skill Assessment ($n = 83$; Overall Mean = 3.06, $SD = .71$)

Ranked position	The language skill can be assessed through ...	D		A		Mean	SD
		F	(P)	F	(P)		
1.	multiple-choice items.(49)*	9	(10.8%)	74	(89.2%)	3.2	.88
2.	true-false items.(47)*	9	(10.8%)	74	(89.2%)	3.11	.75
3.	matching items.(48)*	11	(13.3%)	72	(86.7%)	3.10	.84
4.	cloze items.(45)*	13	(15.7%)	70	(84.3%)	3.07	.88
5.	sentence completion items.(46)*	21	(25.3%)	62	(74.7%)	2.80	1.04

Note: The number in the brackets with the asterisk (*) represents the item number in the questionnaire.

Referring to the alternative types of reading skill assessment, Table 12 below shows that the most agreed upon alternative type of reading skill assessment reported by the respondents was reading aloud (78.3%, $M=2.96$, $SD=.89$). This was followed closely by self-assessment (73.5%, $M= 2.78$, $SD=1.07$). However, only about half of the respondents reported agreement on peer-assessment, taking notes and student portfolios. The least favorable type of assessment was role-playing (37.3%, $M=1.99$, $SD=1.08$).

Table 12. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about Alternative Types of Reading Skill Assessment ($n = 83$; Overall Mean = 2.44, $SD = .72$)

Ranked position	The language skill can be assessed through ...	D F (P)	A F (P)	Mean	SD
1.	read aloud.(42)*	18 (21.7%)	65 (78.3%)	2.96	.89
2.	self-assessment.(67)*	22 (26.5%)	61 (73.5%)	2.78	1.07
3.	peer assessment.(66)*	34 (41.0%)	49 (59.0%)	2.53	1.14
4.	taking notes.(65)*	40 (48.2%)	43 (51.8%)	2.34	1.13
4.	student portfolios.(53)*	40 (48.2%)	43 (51.8%)	2.30	1.09
6.	role play.(59)*	52 (62.7%)	31 (37.3%)	1.99	1.08

Note: The number in the brackets with the asterisk (*) represents the item number in the questionnaire.

D= "Strongly Disagree" and "Disagree"; A= "Agree" and "Strongly Agree"; F = "Frequency"; P = "Percentage"

4.4.2 Beliefs about Types of Writing Skill Assessment

Descriptive statistics including frequencies, percentages, means and standard deviations are presented in Table 13 to show the rank order of English language instructors' reports for their beliefs regarding traditional as well as alternative assessment techniques when assessing writing skills.

The table shows that the majority of respondents had a high level of agreement on summary writing (92.8%, $M=3.39$, $SD=.73$). Following closely were editing tasks, error recognition and sentence completion items. Almost half of the respondents, however, reported agreement with transfer of information, cloze items, and description tasks. At the bottom of the list, and with less than half of the participants' agreement, were multiple-choice items and true-false items with identical percentages of agreement of 44.6%, $M=2.29$, $SD=1.05$ and 44.6%, $M=2.28$, $SD=1.02$, respectively.

Table 13. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about Traditional Types of Writing Skill Assessment ($n = 83$; Overall Mean = 2.85, $SD = .53$)

Ranked position	The language skill can be assessed through ...	D F (P)	A F (P)	Mean	SD
1.	summary writing.(50)*	6 (7.2%)	77 (92.8%)	3.39	.73
2.	editing tasks.(54)*	10 (12.0%)	73 (88.0%)	3.23	.87
3.	error recognition.(55)*	12 (14.5%)	71 (85.5%)	3.13	.81
4.	sentence completion items.(46)*	14 (16.9%)	69 (83.1%)	3.04	.96
5.	transfer of information (from nonlinear to linear text). (56)*	20 (24.1%)	63 (75.9%)	2.89	1.05
6.	cloze items.(45)*	29 (34.9%)	54 (65.1%)	2.58	1.05
7.	description tasks.(62)*	32 (38.6%)	51 (61.4%)	2.54	1.09
8.	dictation.(43)*	34 (41.0%)	49 (59.0%)	2.45	1.04
9.	multiple-choice items.(49)*	46 (55.4%)	37 (44.6%)	2.29	1.05
10.	true-false items.(47)*	46 (55.4%)	37 (44.6%)	2.28	1.02

For alternative types of writing assessment, Table 14 shows that the most agreed on alternative types of writing skill assessments reported by the respondents were essay writing and reflective writing (95.2%, $M=3.51$, $SD=.67$). These were followed by student portfolio (89.2%, $M= 3.23$, $SD=.83$). However, more than two thirds of the respondents reported agreement on peer-assessment, self-assessment, and taking notes.

4.4.3 Beliefs about Types of Listening Skills Assessment

Descriptive statistics, including frequencies, percentages, means, and standard deviations are presented in Table 15 to illustrate the rank order of English language instructors' reports for their beliefs regarding traditional as well as alternative assessment techniques when assessing listening skills.

Based on the overall mean of 2.43 and standard deviation of .79 presented in the table, the results show that the respondents reported low level of agreement to different traditional types of listening skills assessment. Hence, English language teachers seemed to be uncertain about the types of listening skills assessment. Almost two thirds of the respondents agreed on multiple-choice items, true-false items and matching items, respectively. However, this percentage of agreement dropped to less than half of the respondents on cloze items and error recognition with resulting values of 45.8%, $M=2.25$, $SD=1.09$ and 34.9%, $M=2.01$, $SD=1.04$, respectively.

Table 14. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about Alternative Types of Writing Skill Assessment ($n = 83$, Overall Mean = 3.15, $SD = .58$)

Ranked position	The language skill can be assessed through ...	D F (P)	A F (P)	Mean	SD
1.	essay writing.(52)*	4 (4.8%)	79 (95.2%)	3.51	.67
2.	reflective writing.(51)*	4 (4.8%)	79 (95.2%)	3.41	.70
3.	student portfolios.(53)*	9 (10.8%)	74 (89.2%)	3.23	.83
4.	peer assessment.(66)*	17 (20.5%)	66 (79.5%)	2.99	.92
5.	self-assessment.(67)*	19 (22.9%)	64 (77.1%)	2.90	1.00
6.	taking notes.(65)*	17 (20.5%)	66 (79.5%)	2.88	1.03

Note: The number in the brackets with the asterisk (*) represents the item number in the questionnaire.

D= “Strongly Disagree” and “Disagree”; A= “Agree” and “Strongly Agree”; F = “Frequency”; P = “Percentage”

Table 15. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about Traditional Types of Listening Skills Assessment ($n = 83$, Overall Mean = 2.43, $SD = .79$)

Ranked position	The language skill can be assessed through ...	D F (P)	A F (P)	Mean	SD
1.	multiple-choice items.(49)*	27 (32.5%)	56 (67.5%)	2.72	1.09
2.	true-false items.(47)*	28 (33.7%)	55 (66.3%)	2.65	1.02
3.	matching items.(48)*	29 (34.9%)	54 (65.1%)	2.64	1.04
4.	sentence completion items.(46)*	43 (51.8%)	40 (48.2%)	2.29	1.07
5.	cloze items.(45)*	45 (54.2%)	38 (45.8%)	2.25	1.09
6.	Error recognition.(55)*	54 (65.1%)	29 (34.9%)	2.01	1.04

Note: The number in the brackets with the asterisk (*) represents the item number in the questionnaire. D= “Strongly Disagree” and “Disagree”; A= “Agree” and “Strongly Agree”; F = “Frequency”; P = “Percentage”

For alternative types of listening assessment, Table 16 shows that the most agreed on alternative type of listening skills assessment reported by the respondents was taking notes in item 65 (79.5%, $M=2.99$, $SD=.94$). On the other extreme, the least agreed on were oral interview and oral presentation with resulting values of 55.4%, $M= 2.42$, $SD=1.16$ and 49.4%, $M= 2.30$, $SD=1.11$, respectively.

Table 16. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about Alternative Types of Listening Skill Assessment ($n = 83$; Overall Mean = 2.65, $SD = .73$)

Ranked position	The language skill can be assessed through ...	D F (P)	A F (P)	Mean	SD
1.	taking notes.(65)*	17 (20.5%)	66(79.5%)	2.99	.94
2.	retelling a story after listening to a passage.(64)*	23 (27.7%)	60(72.3%)	2.89	1.13
3.	oral summaries following a listening passage.(63)*	22 (26.5%)	61 (73.5%)	2.87	1.08
4.	oral interview.(44)*	23 (27.7%)	60 (72.3%)	2.80	1.13
5.	self-assessment.(67)*	29 (34.9%)	54 (65.1%)	2.69	1.15
6.	role play.(59)*	30 (36.1%)	53 (63.9%)	2.67	1.15
7.	peer assessment.(66)*	35 (42.2%)	48 (57.8%)	2.61	1.14
8.	oral discussion.(60)*	35 (42.2%)	48 (57.8%)	2.58	1.17
9.	dictation.(43)*	32 (38.6%)	51 (61.4%)	2.52	1.11
10.	oral interview.(58)*	37 (44.6%)	46 (55.4%)	2.42	1.16
11.	oral presentation.(57)*	42 (50.6%)	41 (49.4%)	2.30	1.11

4.4.4 Beliefs about Types of Speaking Skills Assessment

As shown in Table 17, almost all the respondents reported high level of agreement on oral discussion (item 60), role play (item 59), oral interview (item 44), public speaking (item 61), and oral presentation (item 57). Almost two thirds of the respondents agreed on oral summaries (item 63), description tasks (item 62), and peer-assessment (item 66). Interestingly, however, this percentage of agreement dropped on self-assessment (65.1%, $M=2.66$, $SD=1.15$).

Table 17. Frequencies, Percentages, Means and Standard Deviations of the Assessment Beliefs about Types of Speaking Skill Assessment ($n = 83$; Overall Mean = 3.23, $SD = .53$)

Ranked position	The language skill can be assessed through ...	D F (P)	A F (P)	Mean	SD
1.	oral discussion.(60)*	2 (2.4%)	81 (97.6%)	3.61	.58
2.	role play.(59)*	1 (1.2%)	82 (98.8%)	3.59	.56
3.	oral interview.(44)*	5 (6.0%)	78 (94%)	3.47	.82
4.	public speaking.(61)*	6 (7.2%)	77 (92.8%)	3.46	.80
5.	oral presentation.(57)*	6 (7.2%)	77 (92.8%)	3.41	.80
6.	retelling a story after listening to a passage.(64)*	10 (12.0%)	73 (88.0%)	3.30	.89
7.	oral summaries following a listening passage.(63)*	21 (25.3%)	62 (74.7%)	2.92	1.05
8.	description tasks.(62)*	23 (27.7%)	60 (72.3%)	2.92	1.12
9.	peer assessment.(66)*	22 (26.5%)	61 (73.5%)	2.87	1.10
10.	self-assessment.(67)*	29 (34.9%)	54 (65.1%)	2.66	1.15

Overall, the results showed that the respondents reported high levels of agreement for different types of speaking assessment.

5. Discussion

Regarding the assessment beliefs and practices that are related to the assessment purposes, the results provided evidence that the English language instructors in the Malaysian tertiary context tend to view the important role of assessment in improving teaching and learning. Mukundan and Ahour (2009) arrived at similar finding in their study in the same Malaysian context. They found that the main reason for assessing students' writing was for identifying their strength and weaknesses, indicating that the improvement of students' learning was the target. Interestingly, however, the above findings are contrary to a previous study conducted in a government funded Malaysian university by Zubairi, Sarudin and Nordin (2008). Although this university's assessment policy issued in 2005 stated that assessment should serve as a powerful tool to enhance teaching and learning, their study found that the use of alternative assessment was not a common practice among faculty members and that giving unannounced assessment involving other paper-pen activities was not evidently practiced by the academic staff in this university.

On the other hand, however, the finding of the present study is in line with previous models of assessment conceptions developed by earlier researchers, such as Brown (2002). According to his model, teachers who agreed that the purpose of assessment is to improve teaching and learning were identified as holding the improvement conceptions about assessment. Such is the case, too, in the study of Muñoz, Palacio and Escobar (2012) who investigated teachers' beliefs about assessment systems applied at a language center of a private university in Colombia. Using surveys, written reports, and interviews, the researchers concluded that the majority of teachers agreed that assessment helps in improving their students learning and their own instruction. Overall, this sample of tertiary English language instructors conceived assessment primarily as an active agent in regulating teaching and learning process and thus were making efforts to focus on the formative assessment purposes rather than on the summative purposes. Such finding is in line with Shohamy, Inbar-Lourie and Poehner's (2008) who found that teachers in the advanced foreign language classroom in the USA were more interested in diagnosing students' abilities in order to decide on areas in need of more support rather than in assigning certain grades. Likewise, Harris and Brown (2009) contend the conception emphasize that assessment is for the joint use of teachers and learners to facilitate learning. Similarly, recent research (Brown & Remesal, 2012; Remesal & Brown, 2015; Muñoz et al., 2012), too, has shown that teachers mostly agree that the main purpose of assessment in the foreign language classroom is to serve the improvement of learning and teaching.

Moreover, results of this study revealed one localized belief of assessment that differs from those beliefs reported in the published literature. Our results indicated that using assessment to create competition among students was the least reported purpose of assessment. However, interestingly, this is contrary to a study conducted by Cheng et al. (2008). They found that grading, testing and competition shared among students and communities are the best indicators of success. Research findings by Remesal (2011) and Azis (2012, 2014, 2015) also indicated that teachers and students

alike were motivated by grading practices. The tertiary context in which this study was conducted may have contributed to the fundamental differences reported in comparison to previous studies of teachers' conception.

Regarding the methods and techniques of assessment, results of this study revealed that English language instructors tended to use a variety of assessment methods to assess students' language ability in their classrooms and that they relied heavily on paper and pencil assessment. Such findings are consistent with that of Graham (2005) who found that teachers are more likely to rely on traditional paper and pencil assessments and attributed this to the fact that these are the types of assessments they experienced when they were students. Along with this, the majority of the Finish academics (Postareff, Virtanen, Katajavuori, and Lindblom-Ylänne, (2012) study used only one type of assessment: the traditional paper and pencil exam at the end of the study module; thus emphasizing the summative assessment purpose. On the other hand, in their study the least frequent methods of assessment used were peer and self-assessment. Moreover, they found out that the faculty teachers rarely used peer assessment and that self-assessment was not used at all. In terms of feedback, grading and reporting of grades, the results were in line with sound practices reported in previous studies although conducted in different contexts -Canada, Hong Kong and China. Cheng and Wang (2007) revealed that most of ESL/EFL university teachers in these three contexts tended to design their own marking criteria before assessing their students while informing them about it ahead. Inevitably, the transparency of the learning expectations and the assessment criteria would be increased. In addition, they found that students had no role in preparing the scoring criteria. They concluded that "assessment seems to be done to the students rather than with them (Cheng and Wang, 2007:101).

Results of the current study show that English language instructors reported using different types of assessments for every language skill taught in their language unit/center. As for the reading skill assessment, results show that instructors tend to use traditional types of assessment more than alternative ones since the highest percentages were detected for multiple choice items, true-false, and matching items, respectively. This finding is in line with Cheng, Rogers and Wang study (2008) study. They reported using selection methods items in assessing ESL students by more than half of the instructors in Canada and China denoting the superiority of traditional methods over alternative ones when assessing this skill.

Regarding the writing skills, the results of this study highlighted the fact that instructors valued essay writing and practiced the most when evaluating their students writing. Such findings are consistent with one past research of Shohamy, Inbar-Lourie and Poehner (2008) where teachers in Advanced Language Proficiency classes reported essay and composition writing to have the highest contributions to the calculation of their students' final grade. As for listening and speaking assessment methods, oral presentation, oral discussion, role-play and public speaking were the most used methods. Similarly, the Chinese participants in Cheng, Rogers and Wang's (2008) study reported using oral discussions and public speaking the most when assessing speaking skill in their classes. The assessment practices in that study was highly structured; the researchers attributed this to having standardized testing program as well as to large class size in the Chinese context.

6. Conclusion

Several implications for the current status of English language classroom assessment could be considered from the results of the present study. The development, validation, and application of the assessment beliefs questionnaire could yield applicability in international contexts across a broad spectrum of language teaching in the field of assessment. Such studies when conducted would provide comparable data to assist with analyzing the effects of different assessment systems in dissimilar contexts. Nonetheless, if the classroom assessment is the real focus of assessment reform in English language centers in the Malaysian universities, instructors should be more empowered in their role as the assessors of students. Their knowledge about what, how, when to assess should be developed through long profession development courses; one-shot workshops or seminars would not be enough to improve instructors' assessment literacy. Instead, supporting university instructors by providing them with materials and other resources to practically encourage them to apply assessment for learning is the way to go.

The main contribution of this study is the newly developed questionnaire used to investigate the English language instructors' assessment beliefs (Appendix). The development of this questionnaire is an important outcome for investigating the English language instructors' assessment beliefs in the tertiary context. It can be applicable in other contexts to help provide comparable studies to assist understanding the issue of assessment in participating countries. As for the context of the study, this study presents a first step towards investigating English language instructors' assessment beliefs in the Malaysian tertiary context; it provides a starting point for complementary research. The study focused on English language instructors who teach English proficiency courses in the state of Selangor. Replications with a larger population of instructors teaching specific English language skills or English content courses and who are located in other states of the country may allow a deeper comprehension of this issue, thus, allowing wider comparisons in the field of English language assessment. Finally, as a recommendation, if English language units/centers in the Malaysian universities plan to move instructors into preferred practices of assessment, it is crucial to take account of their pre-existing beliefs and conceptions. Studying instructors' beliefs about assessment can allow researchers and policy makers to delve into the factors that may contribute to improve assessment practices to use it as a means of improving teaching and learning.

References

- Barnes, N., Fives, H. & Dacey, C.M. (2015). Teachers' beliefs about assessment. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teachers' beliefs* (pp. 284-300). London: Routledge.
- Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (Vol. 11, pp. 137-143). Cambridge, United Kingdom, Cambridge University Press.
- Brown, G.T.L., & Harris, L.R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources heighten conceptions of assessment as school accountability. *Journal of Multidisciplinary Evaluation*, 6 (12), 68-91.
- Brown, G.T.L., Lake, R., & Matters, G. (2009). Assessment policy and practice effects on New Zealand and Queensland teachers' conceptions of teaching. *Journal of Education for Teaching*, 35, 61-75.
- Brown, G.T.L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, 27, 210-220.
- Brown, G. T., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish journal of psychology*, 15(01), 75-89.
- Cheng, L., & Wang, X. (2007). Grading, Feedback, and Reporting in ESL/EFL Classrooms. *Language Assessment Quarterly*, 4(1), 85-107. doi: 10.1080/15434300701348409.
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21(3), 360-389.
- Cheng, L., Rogers, T., & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education*, 33(1), 9-32.
- Cizek, G., Fitzgerald, R., & Rachor, R. (1995). Teacher's assessment practices: Preparation, isolation and the kitchen sink. *Educational Assessment*, 3(2), 159-179.
- Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, 21, 607-621.
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.
- Harris, L. R. & Brown, G. T. L., (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, 16, 365-381.
- Holroyd, C. (2000). Are assessors professional? *Active Learning in Higher Education*, 1(1), 28-44.
- Johnson, B., & Christensen, L. (2012). *Educational Research: Quantitative, qualitative, and mixed approaches*. Los Angeles, CA: Sage Publications, Inc.
- Mukundan, J., & Ahour, T. (2009). Perceptions of Malaysian school and university ESL instructors on writing assessment. *Journal Sastra Inggris*, 9 (1), 1-21.
- Muñoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *Profile Issues in Teachers' Professional Development*, 14(1), 143-158.
- Norizan Md Nor, Noriah Mohamed, Hasuria Che Omar & Ruslan Rainis. (2010). The Direction and Future of Higher Education in Malaysia. In Z. Moris (Ed.), *50 Years of Higher Education Development in Malaysia (1957-2007)*. (National Higher Education Research Institute, Universiti Science Malaysia, Trans.).
- Postareff, L., Virtanen, V., Katajavuori, N., & Lindblom-Ylänne, S. (2012). Academics' conceptions of assessment and their assessment practices. *Studies in Educational Evaluation*, 38(3), 84-92.
- Remesal, A., & Brown, G. T. (2015). Conceptions of assessment when the teaching context and learner population matter: compulsory school versus non-compulsory adult education contexts. *European Journal of Psychology of Education*, 30(3), 331-347.
- Rogers, T., Cheng, L., & Hu, H. (2007). ESL/EFL instructors' beliefs about assessment and evaluation. *Canadian and International Education/Education canadienne et internationale*, 36(1), 39-61.
- Shohamy, E., Inbar-Lourie, O., & Poehner, M. (2008). Investigating assessment perceptions and practices in the advanced foreign language classroom. *Report n, 1108*.
- White, E. (2009). Are you assessment literate? Some Fundamental Questions Regarding Effective Classroom-based Assessment. *OnCUE Journal*, 3(1), 3-25.
- Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *Tesol Quarterly*, 43(3), 493-513.
- Yang, T. (2011). Tertiary EFL Teachers' Classroom Assessment Practices. *International Journal of Learning*, 18(3), 1-8.
- Zubairi, A. M., Sarudin, I. H., & Nordin, M. S. (2008). Faculty competency in assessment. In *Annual Conference of the Association for Educational Assessment (IAEA)*, UK: Cambridge.

Appendix**Assessment Beliefs Questionnaire**

This four-section questionnaire aims to explore your assessment beliefs as an English language instructor.

Please tick (✓) your response to the following statements by using the scale below.

1 = Strongly Agree (SA)

3 = Disagree (DA)

2 = Agree (A)

4 = Strongly Disagree (SD)

Section A: Beliefs about Assessment Purposes

Assessment _____.		SA 1	A 2	DA 3	SD 4
1	helps to focus teaching				
2	helps to group students for instructional purposes				
3	diagnoses strengths and weaknesses in students				
4	diagnoses strengths and weaknesses in teaching				
5	provides information about students' progress				
6	creates competition among students				
7	creates a valuable learning experience for my students				
8	motivates my students to learn				
9	provides feedback to students as they learn				
10	determines the students' mastery of their learning				

Section B: Beliefs about Methods and Techniques of Assessment

		SA 1	A 2	DA 3	SD 4
11.	Formal assessment provides a good evaluation of students' work.				
12.	Informal assessment provides a good evaluation of students' work.				
13.	Paper and pencil assessment is the best method in evaluating students' work.				
14.	Computer technology helps in assessing students' work.				
15.	Assessment questions should reflect real life language use.				
16.	Language instructors need to use a variety of assessment methods to assess students.				
17.	The best assessment items are the ones developed by the language instructor.				
18.	Assessment items are best prepared collaboratively.				
19.	Ready-made assessment items found on the internet are a good source for assessing language use.				
20.	Assessment items from published textbooks are a better source for assessing language use than those found on the internet.				
21.	The best means of assessing language is through formative assessment.				
22.	The best means of assessing language is through summative assessment.				
23.	Designing test specifications before a test is constructed is important.				
24.	Objective testing (e.g. matching items, multiple-choice items, true - false items, cloze items) is a good method of assessment.				
25.	Subjective testing (e.g. journal entry, portfolio, short essay, sentence completion, reflective task) is a good method of assessment.				
26.	Self-assessment by the student is a good method of assessment.				
27.	Peer-assessment is a good method of assessment.				

Section C: Beliefs about Feedback, Grading and Reporting of Grades

		SA 1	A 2	DA 3	SD 4
28.	Students' final grades should be based on tests and exams only.				
29.	Students' final grades should be based on coursework only.				
30.	Students' final grades should be based on coursework, tests and examinations.				
31.	A marking scheme should be prepared before assessment is given.				
32.	Sample marking should always be carried out.				
33.	Sample marking would help in improving marking criteria.				
34.	Conferencing with students is a good way of giving feedback during the course.				
35.	Criterion-referenced assessment is better than norm-referenced assessment.				
36.	Giving a letter grade (e.g. A, B) is better than a percentage score as a performance indicator.				
37.	Students should be given feedback after assessment.				
38.	Students should be informed about the marking criteria before being assessed.				
39.	Students should be involved in preparing the marking criteria.				
40.	Students should be given back their assessment results no later than a week after the test.				
41.	Marks allocation for each test question should be made known to students.				

