



Role of Question Types in Assessing Content Knowledge & Language Competence

Lina Mukhopadhyay

School of English Language Education, The English and Foreign Languages University

Pin 500605, Hyderabad, India

E-mail: linamukhopadhyay@gmail.com

Received: 21-05-2013

Accepted: 07-07-2013

Published: 01-09-2013

doi:10.7575/aiac.ijalel.v.2n.5p.185

URL: <http://dx.doi.org/10.7575/aiac.ijalel.v.2n.5p.185>

Abstract

This study sets out to examine whether ESL teachers design evaluation criteria according to the demands of different question types assessing literary competence. Four students enrolled in an MA TESL programme were asked to respond to three types of questions – a summary, a content based question, and a reader response question - based on one literary text. Five ESL teachers were asked to select appropriate evaluation criteria for each question type and assess student performance. The results show that all the teachers differentiate between content knowledge and language competence in identifying evaluation criteria for the three questions. But most of them do not select criteria to reflect variations in demands made by each question type. A pedagogical implication of the study is that ESL teachers need training to design evaluation criteria that can match the demands of different question types. This would make content based assessment fair, ensure high inter-rater reliability and produce beneficial washback for ESL students.

Keywords: ESL writing, content based assessment, evaluation criteria, inter-rater reliability, washback

1. Introduction

In language education, there exists a thin dividing line between teaching and assessing ‘content’ and ‘language competences’. In most ESL curriculum, starting from the secondary level and right up to the university level, English literature is taught to serve dual purposes - to develop content knowledge and language abilities, especially reading and writing skills (Lukmani, 1996). Hence, questions that constantly worry us while assessing knowledge of literature are - *What should a language teacher concentrate on – a student’s knowledge of content or language competence or both? Should different question types such as describing the plot of a story or summarizing an author’s work or evaluating a literary text have different evaluation criteria or can all questions be assessed using one single criteria?*

1.1 Role of evaluation criteria in writing assessment

It is a well-known fact that performance on disciplines like literature or social sciences involves both content knowledge and language competence. However, in assessing content based questions, teachers often fail to give independent credits to these two knowledge bases (Crandall, 1987; Gottlieb, 2006). To ensure fair assessment practices, ESL teachers need to be sensitized about the importance of designing evaluation criteria that would assess a student’s performance on both of these abilities independently.

Teachers also need to consider the varying demands in writing abilities made by different question types. For instance describing an event in a novel would require a writing ability that is quite different from critically evaluating an author’s work. The first would require a student to recall details of the event and describe it; in contrast, for the second one, the student would need to evaluate some literary features in a piece of work. So, in the second one, the student would need to express his/ her opinion and justify it. Answering the second question, therefore, would involve a much higher level of writing skill than mere description of an event involved in the first one.

The primary aim of this paper is to explore whether teachers distinguish between content knowledge and language competence in assessing performance on English literature within an ESL curriculum. So, we address two research questions –

1. *Do language teachers distinguish between content knowledge and language competence in content based assessment?*
2. *Is teachers’ choice of evaluation criteria based on the requirements of the questions they assess?*

To examine the validity of these questions, we made a short enquiry in a group of university teachers in India about their views and practices in assessing knowledge of English literature. In the next section, we present the details of the study.

2. The Study

2.1 Subjects

Four students (2 male and 2 female), aged between 19 and 23 years, enrolled in an M.A TESL programme in an Indian university served as subjects. On an average, the students had thirteen to fifteen years of exposure to English as a SL both through classroom and naturalistic contexts. According to the CEFR norms, they were at B1 to B2 levels of proficiency in English. No prior assessment of linguistic or literary competence was made before their selection.

Five professors teaching in the same university and in the same programme also participated in this study. At the time of the study, all the teachers had at least five years of experience in teaching and assessing knowledge of English literature.

2.2 Materials and procedure

The students had to read an article by Chinua Achebe, *The Truth of Fiction* (1989), which elaborates upon the different kinds of fiction written in various parts of the world. This text was chosen because it could provide scope for free and critical responses. The students had to answer three questions on the text and submit the answers as an assignment. The following questions were asked:

1. Summarize the main arguments in *The Truth of Fiction*. Once you have summarized the text, indicate which parts of the text you have left out and why.
2. How do the different types of fiction differ from each other? Explain with reference to Chinua Achebe's text.
3. *Fiction in today's world is more political than imaginative. This is reflected in Afro-American fiction and any kind of postcolonial literature.* What are your personal views regarding the role of fiction and its relationship to imagination. Explain with help of Chinua Achebe's text and your personal examples.

In answering these questions, the students would have to concentrate on *what* the text means and *how* it can be interpreted. Furthermore, each of these three questions places different demands on students' linguistic and cognitive abilities (Bloom, 1956). Language-wise each question requires different writing and text organization abilities. They also involve the use of several components of cognitive abilities as specified in the 'Revised Bloom's Taxonomy' (RBT) (Krathwohl, 2002) like – *remember, comprehend, analyse, apply, evaluate* and *create*. Below we list the range of writing sub-skills and cognitive abilities required in answering each of the three questions:

- a) The first question is a summary question and deals with 'what' the text means. Students would have to present the main ideas coherently. In doing so they would need to employ cognitive abilities like *comprehending* the gist of the text and *applying* summarizing skills.
- b) The second question is a thematic question. For this response, the students would have to write about the distinctive features of each type of fiction as specified in the Achebe text. This would require them to employ cognitive abilities like *comprehension* of text and *analysis* to compare and contrast the distinctive features of each type of fiction.
- c) The third question is a reader response question as it examines whether the reader is aware of his/ her meaning making process. So, the students would use cognitive abilities like *comprehension* and critical *evaluation* of the text. In evaluating the text, they would have to use critical thinking abilities and argue in a logically convincing and coherent manner.

The five university teachers who participated in this study were asked to assess student performances. They were given a letter in which they were asked to do the following:

1. Indicate evaluation criteria with specific sub-features as necessary to assess responses on each of three questions. Each question carried 10 marks. The teachers were also asked to indicate the break up for each criterion out of 10.
2. Provide descriptions of expected performances on a rating scale with four bands - 80% and above, 60% and above, 40% and above and 20% and below.
3. Score three responses of each student by using their suggested evaluation criteria.

The study was conducted in two phases: In the first phase, the students were given one week to finish the assignment. In the second phase, the teachers were asked to design evaluation criteria and evaluate the student scripts. They were also given a week to do this task.

2.3 Method of data analysis

Teachers' statements on selecting evaluation criteria, their expected levels of performance in each band and their actual patterns of assessing student performance will be presented on the basis of frequency counts. A qualitative analysis of their perceptions on choice of evaluation criteria for the three question types will also be presented¹.

3. Findings

Here, we present the salient findings of our study according to the range of criteria selected by the teachers to assess the three question types followed by their actual assessment pattern.

3.1 Teacher expectations of student performance: Band descriptions

Evaluation criteria selected and the statements on expected responses revealed the kind of content knowledge and language competence the teachers expected from each question type. Two (T3 and T4) out of five teachers selected criteria but did not express any opinion about the expected level of literary and language competence of students on the four band rating scale. This shows that these two teachers did not have any preconceived notion about how a response that is awarded a 80% is likely to be different from one that is awarded a 40%. They were likely to follow a subjective pattern of assessment according to performance in each question. So, this hints at a kind of arbitrariness in assessment because it does not reflect a well thought out set of features to be checked for while awarding different scores to student responses.

The other three teachers T1, T2 and T5, who presented their statements, showed some regularity in their thinking. The differentiation of each criterion can be understood on a continuum with three bands as given in Table 1 below:

80% and above  20% and below

Table 1. *Criterion-Wise Band Descriptions*

CRITERION	Band A	Band B	Band C
<i>Content</i>	Shows originality of ideas.	Attempts only comprehension.	Shows rote learning from notes.
<i>Language</i>	Shows good expressive and grammatical use of language.	Shows mostly grammatical use of language.	Shows many errors in grammar.
<i>Argumentation</i>	Shows ability to argue ideas convincingly.	Does the needful to approximately fulfil task demands.	Shows inability to argue convincingly.

From the band descriptions in Table 1, it is clear that the three teachers distinguish between assessing content and language along with a third distinction, namely use of argumentation skills, a higher order cognitive ability. The descriptors indicate that these teachers have a clear estimate of how responses might vary across a rating scale with three levels of abilities reflected in the three bands. This is quite unlike T3 and T4 who show almost no awareness of varying levels of performance across each criterion.

3.2 Relationship between evaluation criteria and question types

Two (T3 and T4) out of five teachers, who did not give any statement about expected responses in different bands on the rating scale, did not change their evaluation criteria for each question type. Their responses show an interesting similarity with respect to three criteria given in Table 2 below:

Table 2. *Evaluation Criteria of T3 and T4*

Criterion	Teacher 3 (% given)	Teacher 4 (% given)
Language (grammar and style)	50	50
Content and Organization	45	40
Handwriting	5	----
Total	100	90

Note that these criteria pertain to a kind of general writing, rather than answering specific question types to show literary competence. For these teachers, writing is 'a general skill' that is evaluated through all the three questions and should be assessed according to one evaluation criteria with a tripartite structure – *content – organization – language*. But in answering questions based on literary texts, both literary and language competences are tested. Furthermore, literary competence requires knowledge of text at several levels starting from text comprehension and description of events to critical evaluation of characters, plot or works of an author, and so on. So, an understanding on the distinctions between content and language knowledge demands in different question types was found to be lacking in T3 and T4. The other three teachers, T1, T2, and T5, varied their evaluation criteria according to the three question types. They also followed a similar tripartite division – *content-organization-language* – in choosing evaluation criteria. But they made a distinction in the criterion of **content** to reflect *what* and *how* one is expected to read/ understand and respond to a literary text (Freedman, 1979; Sperling, 1998). This distinction is presented in Table 3 below:

Table 3. *Question-Wise Criterion in Reading and Critical Thinking Abilities*

	Question 1 (Q1)	Question 2 (Q2)	Question 3 (Q3)
--	-----------------	-----------------	-----------------

	<i>Summary question</i>	<i>Classification of fiction question (attending to details)</i>	<i>Reader response question</i>
Different types of reading abilities	1. Overall comprehension	1. Reading for detail	1. Engagement with the text
	2. Distinction between arguments & illustrations	2. Analytical reading	2. Reaction to the stance in the question
		3. Distinction between points	

The list in Table 3 reveals that T1, T2, and T5 make a clear distinction between what kind of information and/ or understanding they expect the students to get from the Achebe text. Once teachers show awareness that different question types have specific demands on content knowledge and language competence, then their assessment process is likely to become more objective and fair.

3.2.1 Reader-response question and variety in choice of evaluation criteria

It was found that in fixing the criteria for Q3 the teachers showed greater variability than for Q1 and Q2. This can be explained by the differences in reading demands of the three questions. Q1 would require summarizing skills and Q2 a detailed attention to content; so the teachers were unanimous in their selection of evaluation criteria for these two questions.

In contrast, Q3 is a reader-response type question and it requires a writer's personal interpretation to be backed with justifications. The teachers felt that this question has a quasi-persuasive function. Hence in answering Q3, aspects like - reaction to the stance in the question, originality in ideas and views, scholarship, and argumentation - would form a significant part of student response. Therefore, this question requires a much higher level of writing skill in contrast to Q1 and Q2. Naturally the teachers differ in assigning levels of reading and writing abilities required to answer Q3 as opposed to text comprehension and use of good language as in Q1 and Q2 (Elliot, 1990).

3.3 Rank ordered performance of students: Issues of inter-rater reliability

A rank ordered analysis of student performance based on a total score of 30 is presented in Table 4 below:

Table 4. *Teacher-Wise Rank Ordered Performance of Students*

Rank orders	S1	S2	S3	S4
T1	IV	I	III	II
T2	I	II	IV	III
T3	I	II	IV	III
T4	I	II	IV	III
T5	IV	I	III	II

Table 4 reveals variety in rank orders of each student. This implies that that the teachers are not unanimous in assessing student performance. So there is low inter rater reliability in this assessment procedure.

The rank ordering pattern, however, reveals an interesting trend: T1 and T5's ratings of students are similar while the ratings of T2, T3 and T4 are comparable. The similarity in assessments within these two sub-groups of teachers can be explained if we look at their respective choices of evaluation criteria: T3 and T4 have chosen the same evaluation criteria across three questions as they do not differentiate between writing in general and writing about literature (refer to Table 2 above). Although T2 has evaluation criteria, which shows some variation according to question types, it is quite similar to T3 and T4's tripartite division – content -organization - language – for all the three questions. In contrast, T1 and T5's rank ordered rating is similar because they have identified the kind of reading required to respond to each question type. For example

§ Q1 requires gleaning the gist for main ideas.

§ Q2 needs a close reading of ideas in the text to identify distinctive features of each type of fiction.

§ Q3 involves taking a stance about the issues raised in the question and justifying it.

As a result, T1 and T5 look for an ability that goes beyond the *content-organization-language* structure. Answering Q3 requires a combination of content knowledge and argumentation skills. These skills, according to them, have been best addressed by S2.

4. Discussion

Overall the results of our study show that the teachers do not practice impressionistic scoring pattern in assessing open-ended responses for literary texts. A reason for identifying evaluation criteria to score student performances could have been because the teachers were specifically asked to do so. We do not know if they practice this technique while assessing student performances in their daily contexts. The two main findings of our study - (i) all the teachers distinguish between assessing content knowledge and language competence (Gottlieb, 2006) and (ii) at least three out of five teachers are able to give expected response descriptions for three bands on a rating scale - validate the first research question.

The second research question is proved to be partially true in that only three out of five teachers make a distinction in identifying evaluation criteria which reflect the varying demands of each question type. But all the teachers are not in agreement while selecting criterion for each question type. This is reason why the teachers show variations in rank

ordering the performances of the students. Quite naturally then such variations lead to low inter-rater reliability in assessment as indicated in the rank ordering of student performances in Table 4 (Shohamy et al., 1992).

4.1 Inter-rater reliability and scope for beneficial washback

One way to tackle the issue of inter-rater reliability would be to train teachers to design evaluation criteria according to the requirements of a question type (Nimehchisalem & Mukundan, 2011). For the design, they would need to agree on the 'construct' of each criterion and list similar sub features under each criterion. Below we present a set of prototypical evaluation criteriaⁱⁱ that could be designed to assess the three question types used in this study:

Table 5. Evaluation Criteria According to Question Types

Band descriptions	Exceeds expectations A grade (80% and above)	Meets expectations B grade (79-40%)	Needs improvement C grade (Below 40%)
	Q1: Writing a summary		
Content (40%)	All the main ideas are presented.	Most of the main ideas presented.	Most of the main ideas are missing.
	The ideas are well elaborated.	The ideas are explained to some extent.	The ideas are not explained.
Organization (30%)	All links between main ideas well drawn to achieve overall coherence.	Main ideas are mostly connected; coherence is attempted.	Ideas lack connection; overall coherence missing.
	Appropriate use of cohesive devices.	Cohesive devices used, some errors present.	Frequent errors in use of cohesive devices.
Language and mechanics (30%)	Use of grammar accurate; a range of structures used.	Some grammatical mistakes do not impede communication.	Frequent grammar errors impedes communication
	Vocabulary use appropriate and has variety.	Vocabulary use is mostly appropriate; has some variety.	Vocabulary use mostly not appropriate; lacks variety.
	All conventions of punctuations met; references well cited.	All conventions of punctuation met; some references present.	Conventions of punctuation followed; references are missing.
	Q2: Content based (detailed)		
Content (40%)	Features of each type of fiction well presented; classifications are well drawn.	Features of fiction presented; classification is not very neat.	Some features of fiction presented; no attempts to classify.
	Adequate examples provided to support ideas.	Some examples provided to support ideas.	No examples provided.
Organization (30%)	Ideas are well connected to achieve overall coherence; contrastive ideas presented in separate paragraphs.	Ideas are mostly connected; coherence is attempted.	Links between ideas not clear; overall coherence missing.
	Appropriate use of cohesive devices.	Cohesive devices used but some errors are present.	Frequent errors in use of cohesive devices.
Language (30%)	Same as in Q1		
	Q3: Reader response		
Content (40%)	A clear stance regarding use of fiction presented.	Stance regarding use of fiction presented.	No stance presented; a general discussion on text provided.
	Appropriate elaboration of arguments done; appropriate use of examples to justify stance.	Some attempts to elaborate arguments; some examples to justify stance.	No attempts at elaborating arguments; examples are either absent or inappropriate.
Organization (30%)	Similarities and contrasts between different types of fiction presented coherently.	Some attempts at presenting similarities and contrasts between different types of fiction.	Some features of different types of fiction listed; ideas are not coherent.
	Text has a tripartite structure – introduction, body and conclusion; text coherence well developed.	Some attempts at building a tripartite text structure.	Text lacks overall structure.
	Appropriate use of cohesive devices.	Use of cohesive devices mostly appropriate.	Frequent errors in use of cohesive devices.
Language (30%)	Same as in Q1		

Note that Table 5 has an overall tripartite structure of evaluation for all the three questions. In it separate weight is given to each criterion - *content*, *organization*, and *language*. Since these are content based questions, so *content* has been given maximum weight at 40% followed by *organization* and *language* (and mechanics of writing) with 30% weight for each. The features – *content* and *organization* – have specific sub-features according to the demands of question types while the language demands are kept common across the questions. Furthermore, the evaluation criteria specifies expected response types for three different bands/ grades that can be awarded for each criterion. The descriptions of abilities represent a continuum of increasing levels of writing from the lowest to the highest levels. This is an example of analytical evaluation criteria.

Raters have always been a source of variability in assessment of writing, especially in open-ended questions. This is because there are variations in perceptions and attitudes of those who read the essays (Hamp-Lyons, 1990) and hence low inter-rater reliability. We cannot expect every assessor to have a uniform manner of scoring free responses. But a consensus on choice of criteria for different question types and the weight to be given to each criterion as given in Table 5 is likely to make the assessment process transparent, fair, reliable and valid (Cho, 2008). Teachers/ assessors need to be trained to use such an evaluation criteria so that their ratings do not exhibit glaring differences. Such an analytic evaluation criteria could also be used to give valuable feedback to ESL students about their performance in each sub criterion fixed for a question type. This would produce beneficial washback as the students would know the features on which they have performed well versus the features they need to improve upon.

4.2 Role of grammar in content assessment

Another issue, which emerges from our study, is teachers' perceptions on using grammatically correct language in content based questions. Three teachers T2, T3 and T4 were found to give almost 50% weight to grammar and expected students to use error free language. We agree that without language one cannot express thoughts and argue out ideas. But while evaluating questions on literary texts, grammatical accuracy cannot subsume the importance of expressing content knowledge. A student might have a few errors in sentence construction but might be able to communicate ideas perfectly well and as teachers we need to acknowledge this ability. After all the function of language is more to communicate effectively rather than aim for native like accuracy. Hence, if teachers give undue importance to grammatical accuracy, then they might overlook abilities like originality in thinking and argumentation skills, which are integral to literary competence.

4.3 Knowledge of discourse structure specific to text types and writing assessment

Another question that can be raised from the present study is: *Do different question types require text-specific discourse competences?* None of the teachers distinguished between discourse feature requirements of each question type in their choice of evaluation criteria. But this is an important segment of written discourse and should not be ignored. In this study, the students had to attempt three types of written discourses – a summary (Q1), a description (Q2), and an argumentative text (Q3). The differences in discourse structures across the three questions are presented in Table 5 under the criterion of *organization*. Such a distinction will help us study in future whether discourse competence is a stable feature in ESL writing or does it vary according to various text types.

5. Conclusion

Though the number of participants of this study was small and no generalizations can be drawn, the aim of the study was to demonstrate that many a times language teachers are not conscious of the thin dividing line between assessing content and language knowledge. They often confuse the presence of language knowledge (or the lack of it) for content knowledge. A pitfall of this confusion is that they do not appreciate content knowledge if a student's writing shows some irregularity in use of grammar. Teachers also tend to readily accept answers that display good knowledge of language at the cost of poor content knowledge. Hence, both teachers and students need to be sensitized on the separate features that could be used to assess content knowledge and language competence simultaneously in a variety of disciplines like social sciences, pure sciences and literature. This would integrate language teaching and assessment across the curriculum for ESL/ EFL students (Snow & Brinton, 1997) and create scope for further academic development.

References

- Achebe, C. (1989). The truth of fiction. In *Hopes and Impediments: Selected Essays*. New York: Doubleday.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York: David McKay Co Inc.
- Brown, H. D., & Abeywickrama, P. (2010). Principles of language assessment. In *Language Assessment: Principles and classroom practices* (2nd ed.) (pp. 25-51). White Plains, NY: Pearson Education.
- Cho, D. (2008). A study on ESL writing assessment: Inter-rater reliability of ESL compositions. *Melbourne Papers in Language Testing*, 1-24.
- Council of Europe. (2011). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Council of Europe.
- Crandall, J. (Ed.). (1987). *ESL through content area instruction*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Elliot, R. (1990). Encouraging reader-response to literature in ESL situations. *ELT Journal*, 44(3), 191-198.
- Freedman, S. (1979). How characteristics of student essays influence teachers' evaluation. *Journal of Educational Psychology*, 71, 328-338.
- Gottlieb, M. (2006). *Assessing English language learners: Bridges from language proficiency to academic achievement*. Thousand Oaks, California: Corwin Press.
- Hamp-Lyons, L. (1990). Second language writing: Assessment issues. In Barbara Kroll (Ed.) *Second Language Writing*. Cambridge: Cambridge University Press.

- Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into Practice*, 41 (4), 212-218.
- Kroll, B. (1998). Assessing writing abilities. *Annual Review of Applied Linguistics*, 18, 219-240.
- Lukmani, Y. (1996). Linguistic accuracy versus coherence in assessing examination answers in content subjects. In Michael Milanovic and Nick Saville (Eds.). *Performance Testing, Cognition and Assessment* (pp.130-150). Cambridge: Cambridge University Press.
- Nimehchisalem, V., & Mukundan, J. (2011). Determining the evaluative criteria of an argumentative writing scale. *English Language Teaching*, 4(1), 58-69.
- Shohamy, E.C., Gordon M., and Kraemer, R. (1992). The effect on rater's background and training on the reliability of direct writing tests. *Modern Language Journal*. 76(1).
- Snow, M.A., & Brinton, D. M. (Eds.). (1997). *The content based classroom: Perspectives on integrating language and content*. White Plains, NY: Addison-Wesley.
- Sperling, M. (1998). Teachers as readers of Students' Writing. In Nancy Nelson and Robert C Calfee (Eds.) *The Reading-Writing Connection* (pp. 131-157). Chicago, Illinois: University of Chicago Press.

Notes

ⁱ Henceforth in the study whenever there is a need to make individual references, the questions will be referred to as Q1, Q2, and Q3; the students will be referred to as S1, S2, S3, and S4 while the teachers will be referred to as T1, T2, T3, T4, and T5.

ⁱⁱ This is analytical evaluation criteria. Teachers can also design holistic evaluation criteria for the same purpose. An advantage of using analytical criteria is that it can be used to give students specific feedback on aspects they have done well and aspects they need to improve upon.