

Assessing Testing Practices with Reference to Communicative Competence in Essay Writing at Undergraduate level in Pakistan

Dr Khurram Shahzad

Assistant Professor, Department of English Studies
National University of Modern Languages

Dr Fauzia Janjua

Associate Professor, Department of English Studies
International Islamic University

Dr Jamil Asghar

Assistant Professor, Department of English Studies
National University of Modern Languages

Abstract

Assessment in Second Language (SL) teaching is of vital importance in the broader praxis of applied linguistics. Even within the framework of assessment in SL teaching, the assessment of testing practices for essay writing is particularly significant as it not only indicates/measures the overall efficacy of the undergraduate programme, but also holds the promise of engaging teachers as well as students in a critical and self-reflective dialogue. The study was conducted to investigate how teachers have been assessing essay writing of the learners at the undergraduate level in two of the well-known universities of Pakistan. For this purpose, a semi-structured questionnaire was constructed keeping in mind the theoretical framework provided for assessment in the Common European Framework of Reference (2001). The questionnaire was pilot-tested and then administered to 134 teachers in eight cities of Pakistan. The data demonstrated that there were many lacunas in the testing practices of the teachers, and they did not have any agreed-upon criterion to evaluate the learners' communicative competence in essay writing. The study offers recommendations to ameliorate this state of affairs and make the essay-writing assessment practices more meaningful and relevant.

Keywords: assessment, criterion, essay writing, communicative competence

Introduction

Essay writing in second language (SL) is a multidimensional concept, which is a significant skill because learners' assignments, intelligence and scholarships are primarily evaluated through it in the academic institutions and at the workplace. Effective essay writing skill is a doorway to better scholastic grades and innumerable educational endeavors. It is also central to pragmatic goals and professional mobility. It is modes of research, expressiveness, thinking, publication, role negotiation, identity formation and social integrations (Elbow, 2000). Moreover, it constitutes the core of assessment in language education and a fundamental part of undergraduate courses in Pakistan and all over the world (McWhorter, 2012; Shahzad, 2018).

Essay writing underscores linguistic and communicative features with an optimum level of accurateness, formality and appropriateness. Thus, it is not just restricted to spellings, grammar, and punctuation; rather, special attention is paid to the overall development of message, organization of ideas, cohesion, coherence, style and tone (Bailey, 2011).

Moreover, assessing writing is very vigorous and indispensable in university education in Pakistan. It helps develop learners' capacities, hence setting up their credentials. It has also a formative role, which helps students integrate into various segments of society to execute diverse roles. SL assessment is given utmost substance, making it important to comprehend its minutiae (Cummings & Wyatt-smith, 2009). Therefore, teachers who often collect data from language tests need to make educated decisions about learners (Luoma, 2004).

Statement of the Problem

Assessment constitutes one of the most essential components in the wider framework of applied linguistics; and within assessment, it is essay writing, which poses unique challenges to examiners and teachers. Given its academic and practical utility, essay writing necessitates special kinds of assessment protocols marked by validity, reliability and authenticity. However, the way essay writing is routinely assessed in Pakistan is far from being satisfactory as there is wide range of issues, which are at once procedural, academic, affective and linguistic. The testing practices usually employed by the teachers are dated, inadequate and lopsided and they tend to focus more on the motor skills than the creative engagement of the students with the interface of text and thought. Long established culture of rote learning and its institutionalized encouragement over decades has placed a very minimal premium on the comprehensible and meaningful learning. Moreover, such a pedagogy drives a wedge between the discursive classroom experience and the real life challenges/necessities. An overemphasis upon linguistic aspects such as syntax and lexis, in fact, tends to take the focus considerably away from the socio-pragmatic competence of the Pakistani SL learners. This has serious implications for language pedagogy currently in vogue in Pakistan as language, by definition, is a contextual, pragmatic and social phenomenon.

Research Questions

1. How do English teachers measure the SL learners' communicative competence in essay writing in Pakistan?
2. How closely are their evaluative practices in measuring essay writing aligned with the CEFR (2001)?

Delimitation of the Study

This study was delimited to the department of English at two of the universities of Pakistan, namely: the National University of Modern Languages (NUML), Islamabad and its regional campuses, and the University of Gujrat (UoG), Gujrat. The regional campuses of the NUML included: Peshawar, Lahore, Karachi, Hyderabad, Multan and Faisalabad. These two universities have been selected since they are running BS (Honors) and MA English courses in their English departments, and are known for teaching all the four skills of English. Bachelor of Science (BS Honors) in English is a four year program, whereas Master of Arts (MA) English is a two year program. Higher Education Commission of Pakistan (HEC) has recognized MA English course as undergraduate course, for it makes up sixteen years of education, which is a pre-requisite for the MS/M. Phil English program.

Literature Review

In SL educational contexts, expertise in essay writing is an important indicator of language ability so far as language testing is concerned. There are different types of language tests such as entrance, placement and proficiency test etc. in which essay type writing tasks are manipulated to assess essay writing ability in the Pakistani universities as well as the universities all over the world (Hamp-Lyons, 2003; Shahzad, 2011).

Since communicative tests evaluate linguistic as well as sociolinguistic, discourse, pragmatic and strategic competencies of the learners (Brown, 2004), they, therefore, are used to measure SL learners' development and progress. These tests are not end in themselves, rather means to actualize a more effective teaching and learning process. They also offer invaluable feedback to learners; they help develop classroom tutoring, ambiance and scholarship (Airasian, 2005). Moreover, they help collect data to assist learners, curriculum developers and teachers in classroom work and materials development (O'Neill, Moore & Huot, 2009).

Reliability, validity, interactive-ness, authenticity, practicality and impact are other key features of communicative language tests. So while developing essay writing tests, different techniques are used from diverse perspectives. Different appraisal methods have been framed that claim to have test reliability. The most important and relevant of these assessment methods are: *analytical* grading scales, *holistic* grading scales, *primary* trait scales and *multiple* trait scales (Fulcher & Davidson, 2007; Weigle, 2002).

These grading scales are different from one another on two elementary grounds: (1) they offer assessors an alternate to grant candidates a single mark or multiple marks for the substance of a timed impromptu essay, and (2) they suggest an alternative to score contenders on overall tools or precisely built instruments for a specific test task. Table 1 adapted from Weigle (2002, p. 109) below summarizes these different kinds of grading scales.

Table 1

	<i>Generalizable to a class of writing test tasks</i>	<i>Specific to a particular test task</i>
Multiple score	Analytic scoring	Multiple trait
Lone score	Holistic scoring	Primary trait

Another major aspect of essay writing assessment is the criteria against, which it is assessed. A short sum-up of criteria for its assessment with descriptors is given in Table 2 below.

Table 2

<i>Scoring Criteria</i>	<i>Descriptor(s)</i>
Content	Is the writing test task produced by the assessee skillfully answer the question/prompt? Is the subject matter of the written discourse relevant to the prompt and test task?
Rhetorical Organization (discourse features)	Is the answer well-formed? Is it cohesive and coherent? Is it organized according to the conventions of essay writing? Is the core notion well-developed?
Communicability	Have student stated their thoughts well? Can they be understood without difficulty?
Syntactic accuracy	Does writing form suitable structures? Does there exist any grammatical mistakes? Do they obstruct message? Are meanings unclear?
Vocabulary	Does writing employ appropriate vocabulary items? Are there any slips of word selection? Do they hamper communication? Does there exist an extensive vocabulary? Is the word selection suitable? Are meanings imprecise?
Mechanics	Does writing adhere to the rubrics of capitalization, spelling and punctuation conventions? Are there several errors? Do they interfere in communication? Are meanings vague?

For the purposes of essay writing assessment, grades are dispersed on the scoring key, permitting various features that evaluators need to measure. Furthermore, the constructs elucidated on the scale should be proportional with the objectives of the course (Alderson, 2000). The constructs of essay writing that is to be assessed on the CEFR (2001) are as follows: *accuracy, content, coherence, range, rhetorical organization and mechanics*.

Methodology and Theoretical Framework

The researchers used quantitative approach as their research paradigm and evaluative research method for the study. CEFR (2001) was taken as a model for the research inquiry as it describes communicative language teaching and testing comprehensively. This model is based on Bachman & Palmer (1996), Celce-Murcia, Dörnyei & Thurrell (1995), Canale & Swain (1980), Hymes (1972), which were specifically developed for language teaching and testing. It “describes in a comprehensive way what language learners have to learn to do in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act efficiently” (CEFR, 2001, p. 1).

The model offers an action-oriented approach trailing collaborative essay writing strategies like reflective discussion, brainstorming, peer practice and writing for meaning (Green, 2011). Besides, this framework provides a complete model for assessing essay writing and suggests rubrics that integrate all the necessary requisites for its evaluation. The researchers used this framework to evaluate the testing practices concerning essay writing in Pakistan.

Population and Setting

The representation of the population and setting is given below in Table 3.

	<i>Location</i>		
	<i>Frequency</i>	<i>Percent</i>	<i>Valid Percent</i>
Islamabad	61	45.5	45.5
Lahore	11	8.2	8.2
Faisalabad	8	6	6
Multan	8	6	6
Peshawar	13	9.7	9.7
Karachi	10	7.5	7.5
Hyderabad	5	3.7	3.7
Gujrat	18	13.4	13.4
Total	134	100	100

Sampling

Since there was not a huge number of population at both the universities, the researchers administered the questionnaire to all of them.

Data Collection Tool

A semi-structured questionnaire, which was constructed using the CEFR (2001), was administered. Its validity was pilot-tested, and reliability was confirmed, using *Cronbach Alpha* on SPSS (version, 21). The questionnaire comprised both open-ended and closed-ended questions and carried twenty questions in all.

Data Analysis

This section of the article puts forth the analysis of the data conducted analytically and submits the findings of the study. It was primarily a quantitative research study; the data, collected through a questionnaire, are analyzed below.

Analysis of the Teachers' Questionnaires

The questionnaire had twenty questions related to assessing essay writing. It also had some questions about demographic information. Most questions had three options for the responders, but some of them had four. They were coded in this way: *yes* as '1', *no* as '2', and *no idea* as '3'. Whenever there was a fourth option, it was coded as '4'. The data were input in the SPSS and were analyzed through descriptive statistics.

Sixty nine male and sixty five female teachers participated in the study. Their ‘age’, ‘designation’ and ‘teaching experience’ is given below in Table 4.

Table: 4 Summary of the respondents’ age, designation and teaching experience

	<i>Age</i>		<i>Designation</i>		<i>Teaching Exp.</i>			
	<i>Frequency</i>	<i>%</i>	<i>Frequency</i>	<i>%</i>	<i>Frequency</i>	<i>%</i>		
under 30 y	36	26.9	Lecturer	113	84.3	Less than 5 y	34	25.4
31-40 y	76	56.7	Assist. Prof	14	10.4	5-10 y	57	42.5
41-50 y	15	11.2	Prof	7	5.2	11-15y	31	13.1
over 50 y	7	5.2				16-20y	6	4.5
						Over 20 y	6	4.5
Total	134	100		134	100		134	100

Table 4 shows the *age*, *designation* and *teaching experience* of the responders. It ranged from under 30 years to over 50 years. They were of different age groups; nonetheless, 31-40 year age group dominates that shows most of them were fairly experienced. Table 4 also indicates the *designation* that reveals that the largest number of responders were *lecturers* followed respectively by *assistant professors* and *professors*. They also had diverse teaching experience, and most of them had between 10 to 15 years.

In the questionnaire, the first question was about *situation* or *setting* in which the *test task* comes about. Situation, in fact, communicates purposes of the essay test and requires from examinees to bring background knowledge to the fore to comprehend it and then develop a strategy to deal with it (Bachman & Palmer, 1996). Its analysis is demonstrated below.

Table: 5 Summary of Q1.

<i>Do you explain the examinees the situation/setting in which the test task occurs?</i>			
	<i>Frequency</i>	<i>Percent</i>	<i>Valid Percent</i>
Yes	61	45.5	45.5
No	37	27.6	27.6
no idea	36	26.9	26.9
Total	134	100	100

Table 5 displays that overall 73 responders either did not elucidate the setting/situation or they had no idea about it in which the test task happens. This implies that this feature of testing which is a vibrant one and is essential to attempt the essay writing test task has been overlooked in the assessment of essay writing.

The next item was about the *number of words* within the test task. As there always exists a time restraint during examination, assessors should put a word limit to the test tasks (CEFR, 2001). This process gives validity to the process of testing. The analysis of this item is elaborated below.

Table: 6 *Summary of Q2.*

<i>Do you ask the test taker the number of words in the task s/he is supposed to write?</i>			
	Frequency	Percent	Valid Percent
Yes	80	59.7	59.7
No	47	35.1	35.1
no idea	7	5.2	5.2
Total	134	100	100

Table 6 demonstrates that 80 responders ask their examinees the number of words in the test task which they are to attempt. Contrarily, 47 responders do not ask the number of words and 7 responders do not have any idea about it. On the whole, almost half the responders either do not ask their examinees about word limit in essay writing test task or they have no idea about it, which is one of the important components of the test of time-impromptu essay writing.

The next item was about the *range of vocabulary*. Range stands for variety or mixture of low as well as high frequency words. In fact, importance of vocabulary cannot be denied in essay writing test task, for it not only assists in understanding others, but also assists authors in expressing their opinions fittingly in written communication (CEFR, 2001).

The analysis of the item is presented below.

Table: 7 *Summary of Q3.*

<i>Do you ask the student about the range of vocabulary s/he is supposed to write?</i>			
	Frequency	Percent	Valid Percent
Yes	45	33.6	33.6
No	74	55.2	55.2
no idea	15	11.2	11.2
Total	134	100	100

Table 7 expresses that 45 responders ask their examinees to employ a wide range of vocabulary in their essay writing tasks; whereas, 74 responders do not ask them about it, and 15 have no idea about it. On the whole, 66.4% of the responders either did not give any significance to different types of vocabulary, including its form, meaning and use, in the test task, which is one of the constructs of essay writing assessment or they have no idea about it.

The next item was about the *number of test tasks* in the final essay writing question paper. It is thought suitable to provide diverse essay writing test tasks to measure various competencies of examinees such as vocabulary items, linguistic structures and discourse aspects, and thereby demanding a communicative competence in order to find out how much competent learners have grown over a period of time. The analysis of the item is given below.

Table: 8 *Summary of Q4.*

<i>How many test tasks do you give in a final test to measure the essay writing performance of the student?</i>			
	Frequency	Percent	Valid Percent
1	4	03	03
2	30	22.4	22.4
3	26	19.4	19.4
4 and more	74	55.2	55.2
Total	134	100	100

Table 8 points out that responders were given four option in this item. The responders have given different responses. Overall half the responders opted for *two* or *three* test tasks which is very less so far as the international testing practice is concerned for three hours paper. If more test tasks are given, there are chances that more of essay writing features such as organization, theme/rhyme constructions, discourse markers, varied syntactic structures, etc., students will be able to display them in their essays.

The next item was about the *choice* that examiners give to testees *within the test tasks*. Indeed, choice is presented in the number of test tasks or within the test tasks, but it must be of identical stress in terms of meta-cognitive, cognitive and procedural requirements (Fulcher & Davidson, 2007). The analysis of this item is presented below.

Table: 9 *Summary of Q5.*

<i>Do you give choice to students within the test tasks (essays) they are supposed to attempt?</i>			
	Frequency	Percent	Valid Percent
yes	80	59.7	59.7
no	48	35.8	35.8
no idea	6	4.5	4.5
Total	134	100	100

Table 9 exhibits that 80 responders *give choice* within the test tasks, while 48 responders stated that they do not give choice within the test tasks; however, six have no idea about it. In one of the studies, Shahzad (2017) has measured the cognitive load of the test tasks, which shows that many paper setters have no idea as how to maintain cognitive load of the questions given to learners in essay writing. For example, look at:

Define descriptive writing. Compare and contrast it with argumentative writing. OR What is argumentative writing? Think about a topic for an argumentative essay and form a thesis statement for it.

The example above shows that one test task deals with memory of the learners, whereas the second task offers open choice, saying *think about any topic*. Thus there is a great discrepancy in both the test tasks. The first one appears quite easy as compared to the second one, giving less load to learners.

The next item was about the choice in the *number of test tasks* examinees were required to attempt. Its analysis is presented below.

Table: 10 *Summary of Q6.*

<i>Do you give choice to the students in the number of test tasks (essays) they are supposed to attempt?</i>			
	Frequency	Percent	Valid Percent
yes	96	71.6	71.6
no	31	23.1	23.1
no idea	7	5.2	5.2
Total	134	100	100

Table 10 indicates that 96 responders offer *choice* in the *number* of test tasks examinees are required to attempt, whereas 31 responders do not give *choice* in the *number* of test tasks examinees are supposed to attempt, and 7 responders have no idea about it.

The findings of questions five and six establish that no homogeneity exists in the practices of the examiners, asking students to attempt essay writing papers. For example, some give *choice within* the test task and some in the *number* of test task. Moreover, the findings confirmed that both types of choices sometimes were provided to examinees that is against the principles and norms set by the language experts.

The next item was about *time given* to examinees for *measuring their essay writing performance*. The analysis of the item is presented below.

Table: 11 *Summary of Q7.*

<i>How much time do you give to the students to test their written performance in essays?</i>			
	Frequency	Percent	Valid Percent
1 hour	10	7.5	7.5
2 hours	39	29.1	29.1
3 hours	85	63.4	63.4
Total	134	100	100

Table 11 specifies that 10 responders give one hour of time to examinees in the final exam; contrarily, 39 responders give two hours, and 85 responders three hours for measuring essay writing performance. It establishes that the final essay writing question paper is three hours at both the universities, but there is no homogeneity in the practices related to choice given in the test tasks in the summative assessment.

The next item was the *explanation* of the *criteria* to examinees before marking their answer sheets. Marking scheme is of highest importance in our democratic world today, endorsing scholarship, accuracy, fairness and student independence, and also recognizing examiners' anticipations. It also certifies exactitude and impartiality (Fulcher & Davidson, 2007). The analysis of this item is presented below.

Table: 12 *Summary of Q8.*

Do you explain the criteria you use to mark the students' answer sheets before testing them in essay writing?

	Frequency	Percent	Valid Percent
yes	94	70.1	70.1
no	31	23.1	23.1
no idea	9	6.7	6.7
Total	134	100	100

Table 12 displays that 94 responders explain the *criteria* to students before marking their final essay writing papers, whereas many don't and some have no idea about it. In fact, whatever criteria they explain to students, it relates to only linguistic competence and do not correspond with the criteria indicated in the CEFR (2001), and this is also one of the gaps in the assessment procedures of the universities.

The next question was about the *number of evaluators* expected to judge the same answer script. To warrant fairness and objectivity in assessment, at least two assessors should evaluate the same answer script. Hence, this item was supplemented in the questionnaire, and its analysis is presented below.

Table 13 *Summary of Q9.*

How many evaluators test the essay writing of the students?

	Frequency	Percent	Valid Percent
01	118	88.1	88.1
02	12	9	9
03	4	3	3
Total	134	100	100

Table 13 shows that most responders stated that there is only *one* evaluator for the assessment of the students' answer scripts. Conversely, a few responders stated *two* evaluators do it and some had no idea about it. The next question related to this item was about departmental *committee* or *authority* who could arbitrate in case of any incongruity in the scoring of the two evaluators. Its analysis is given below.

Table 14 *Summary of Q10.*

<i>In case of any discrepancy in the marking of the two evaluators, does your department have any committee or authority to have a final word?</i>			
	Frequency	Percent	Valid Percent
Yes	13	9.7	9.7
No	60	44.8	44.8
no idea	61	45.5	45.5
Total	134	100	100

Table 14 displays that most of the responders stated that there exists no any committee or authority to decide in case of any discrepancy if two evaluators do not agree on the scores to be awarded to a learner, and most of them had no idea about it. Hence, the data establish that, in reality, there is only one evaluator who marks the answer scripts of the learners and this is against the norms set by the testing experts (CEFR, 2001).

The next item dealt with *rating scales* with *descriptors* for judging the learners' answer sheets. Since rating scales determine the validity of assessment, they explicate the criteria against which the excellence of judgement is made. They also offer a shared understanding for the scoring and interpretation of the written essay (Bachman, 1990). The result of this item is detailed below.

Table 15 *Summary of Q11.*

<i>Do you use any rating scales or bands with descriptors while marking the students' answer sheets?</i>			
	Frequency	Percent	Valid Percent
yes	41	30.6	30.6
no	62	46.3	46.3
no idea	31	23.1	23.1
Total	134	100	100

Table 15 demonstrates that 41 responders stated to have used *rating scales* with descriptors for measuring learners' performance in essay writing. Nonetheless, most responders stated they do not use or they have no idea about it. Further, the data show that they do not use them in their evaluative practices. The evaluators actually depend on their own assumed criteria. This way, they make the whole process of assessing essay writing whimsical, unfair and subjective.

To probe further, the next item related to the previous one was about the *kind of rating scales* assessors were using for assessment. The significance of rating scales cannot be overlooked because they warrant all-encompassing sort of evaluation. They, without a doubt, offer a realistic and illustrative comprehension of the way essay writing assessment is theorized and practiced. These scales are – *analytic, holistic* and *primary* trait scales. The analysis of this item is discussed below.

Table 16 *Summary of Q12.*

<i>What kind of scales or bands do you use to test the written performance of the students?</i>			
	Frequency	Percent	Valid Percent
holistic scales	25	18.7	18.7
analytical scales	22	16.4	16.4
primary trait scale	6	4.5	4.5
none of them	81	60.4	60.4
Total	134	100	100

Table 16 indicates that different responders have given different answers because they do not have any prescribed or agreed-upon principles for the measurement of communicative competence in essay writing. Moreover, the data illustrate 81 respondents stated that they do not use any of the *scales* for essay writing assessment, for both the universities have not provided any of them to teachers, nor have they been trained for their use. Consequently, the data exhibit that their method of essay writing assessment is invalid, unfair and subjective.

The next item relates to formative assessment and asks how teachers keep *record of the learners' classroom essay writing performance*. Keeping the record of learners' essay writing can take various forms such as diaries, folders and portfolios and lies at the heart of SL teaching and learning process. It is an effective tool to track single student's input in the classroom (Airasian, 2005). The analysis of this item is presented below.

Table 17 *Summary of Q13.*

<i>How do you keep the record of the classroom essay writing performance of the students?</i>			
	Frequency	Percent	Valid Percent
portfolio	12	9	9
diaries	17	12.7	12.7
observations	63	47	47
none of them	42	31.3	31.3
Total	134	100	100

Table 17 reveals that different respondents have given different answers against the entries. However, most of them say they use *observations* as a tool to keep record of the learners' performance and 42 respondents stated that they use none of these techniques for record keeping of an individual learner's performance. The data reveals that teacher observed students, but there was no formal way to keep record of their essay writing performance. They, in fact, made "mental notes". This also shows that there exists a gap in their essay writing testing practices.

The next item was how many times a teacher *scores or marks a student's classroom work*. In formative assessment, scoring students' classroom work also signifies providing them corrective feedback, which is usually given in the form of codes, symbols or grades. Sometimes, comments are also given in the margin so that students may mend their work, and hence improve their essays (Truscott, 2007). The analysis of this item is given below.

Table 18 *Summary of Q14.*

<i>How many times do you test students' classroom performance in essay writing in a semester?</i>			
	Frequency	Percent	Valid Percent
1 time	21	15.7	15.7
2 times	47	35.1	35.1
3 times	32	23.9	23.9
4 times and more	34	25.4	25.4
Total	134	100	100

Table 18 displays that most teachers score their students' work *once* or *twice* in a semester, and only some of them mark their students' work *four* or *more* times in a semester. The data, in fact, show that more than half the teachers judge their students' work just once or twice in a semester, which is carried out after the mid-term exam where their mistakes are shown to them and usually they do not check their learners' work.

The next item was about the *use of criteria* to assess the learners' classroom essay writing performance because criteria detail what skills or abilities examiners should look for. The analysis of this item is presented below.

Table 19 *Summary of Q15.*

<i>Do you use any criteria to test the classroom writing performance of the students?</i>			
	Frequency	Percent	Valid Percent
Yes	63	47.1	47.1
No	13	9.7	9.7
no idea	58	43.2	43.2
Total	134	100	100

Table 19 displays that more than half the teachers either do not use any rating scales or they do not have any idea about them. This is also corroborated that while answering about summative assessment most teachers stated they have no idea about scales with descriptors to be employed to measure the essay writing performance of the learners. It makes the whole process of assessment invalid and unreliable.

The last closed-ended item was about *rating scales* they were using to test the students' classroom essay writing performance. Without doubt, rating scales expound the criteria against, which judgement is given about the quality of discourse learners have produced (Bachman & Palmer, 1996). The analysis of this item is given below.

Table 20 *Summary of Q16.*

<i>What criteria do you use to test the classroom essay writing performance of the students?</i>			
	Frequency	Percent	Valid Percent
holistic scales	26	19.4	19.4
analytical scales	37	27.6	27.6
primary trait	5	3.7	3.7
no idea	66	49.3	49.3
Total	134	100	100

Table 20 shows that different responders have given different responses against the item. However, most of them have said that they have no idea about scales to be used to measure the essay writing performance. The data, in fact, reveal that most of the teachers have no homogenous practices concerning the assessment of communicative competence in essay writing. Neither do they have valid, consistent and objective criteria, nor do they have any scales to measure the classroom as well as summative performance of the learners.

The four open-ended questions related to formative and summative assessment were:

1. What do you specifically test in the essay writing performance of the students?
2. What specific rating scale do you use to test the written performance of the students? Please name it.
3. Through observations, how do you keep the record of the classroom performance of the students?
4. If you use any other criteria to evaluate the classroom performance of the students, please explain.

They have been analyzed through *thematic analysis*. The verbal data were read and re-read many times and the patterns and themes, which emerged there-by have been analyzed. The data reveal that most of the teachers particularly focused on the language of the students. They marked the essay by checking it for grammatical mistakes such as syntax, lexis, spelling and punctuation marks. They did not pay attention to the communicative competence of the learners in essay writing, though some responders did write “content, ideas and organization” of the essay as well.

Concerning linguistic competence, responses included: “I specifically test topic, content, grammar of the essay written by students”, “language should be formal and up-to-the mark”, “clarity, ideas”, “vocabulary of the students”, “I test whether they know which type of essay they are to write or not”, “sentence formation, neatness, organization of ideas”, “I have no comments”, “formulation and organization of ideas and grammar”, “performances are measured in terms of attendance, discipline, classroom participation and presentation skills”, etc. If one looks at *discipline, presentation skills*, etc., one finds that such criteria have never been discussed in any of the rating scales for the measurement of essay writing in any of the documents.

Regarding record-keeping of the classroom performance, many teachers did not answer this open-ended question, but those who answered, their responses included: “mental notes”, “I keep the record of the students’ performance in mid-term exams”, “through their discussion in the

class on a specific topic”, “we keep students’ test record”, etc. All this shows that their evaluative practices related to the assessment of essay writing are faulty, invalid, subjective, inconsistent and unfair.

Findings

- Teachers mostly do not describe the situation/setting in which the test task occurs.
- Teachers mostly do not mark the assignments of the learners.
- Teachers do not use scales with rubrics to grade and score the assignments as well as final paper answer sheets of the learners.
- Teachers do not have valid, objective and well-established criteria to measure the communicative performance of the learners in essay writing.

While designing essay writing tests, teachers should learn clearly to link every single question to a learning objective, and should describe the prompt/topic in detail.

Conclusion and Suggestions

Teachers should prepare students in essay writing for real challenges to perform in real life situations. Keeping in mind the context of Pakistani classrooms, the researchers, while employing the CEFR as a theoretical framework, have evaluated the testing practices concerning essay writing and found considerable gaps and deficiencies. To overcome these problems, the researchers have offered the following recommendations:

- Teachers should determine the objectives of the proposed essay writing test;
- They should recognize and outline the previously planned learning outcomes;
- They should formulate the test specifications (to articulate clearly what one intends to assess);
- Test items should be pertinent and appropriate.
- There has to be some marking key as to measure the communicative competence of the learners.
- There should be some laid down criterion for teachers to interpret the scores awarded to the students.

It is hoped that these recommendations, if followed, will standardize the existing testing practices and make them align with the larger curricular objectives.

References

- Airasian, W. P. (2005). *Classroom Assessment: concepts and applications* (5th ed.). Boston: McGraw Hill Companies.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bailey, S. (2011). *Academic writing: A handbook for international students* (3rd ed.). London & New York: Routledge.
- Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practice*. White Plains, N.Y.: Pearson Education, Inc.
- Celce-Murcia, M., Dornyei, Z. & Thurell, S. (1995). *Communicative competence: a pedagogically motivated model with content specifications*. Issues in Applied Linguistics Regents of the University of California. Accessed on 4/2/2014 at 6: 30 pm
- Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- CEFR. (2001). *Common European framework of reference*. Cambridge: Cambridge University Press.
- Cumming, J & Wyatt-Smith, C. (2009). *Educational assessment in the 21st century: Connecting theory and practice*. Springer. Dordrecht Heidelberg.
- Elbow, P. (2000). *Everyone can write: Essays toward a hopeful theory of writing and teaching writing*. London & New York: Oxford University Press.
- Fulcher, G. & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. USA & Canada: Routledge.

- Green, A. (2011). *From common European framework to classroom application: the English profile solution*. Proceedings of the 16th Conference of Pan-Pacific Association of Applied Linguistics. Accessed from, paaljapan.org/conference2011/ProcNewest2011/pdf/keynote/TG.pdf at 7: 40pm. On 1/6/2014
- Hamp-Lyons, L. (2003). *Writing teachers as assessors of writing*. In Kroll, B (ed.). *Exploring the dynamics of second language writing* (pp. 162-190). Cambridge: Cambridge University Press.
- Hymes, D. (1972). *On communicative competence*. In J. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings*. Harmondsworth, England: Penguin Books.
- Kachru, Y., & Nelson, L. C. (2011). *World Englishes in Asian contexts*. Hong Kong University Press. Aberdeen, Hong Kong.
- Khan, I. H. (2012). English teachers' perceptions about creativity and teaching creative writing in Pakistan. *American International Journal of Contemporary Research*. V, (3), retrieved from [www. Aijcnet.com/journals/vol_2_No_March_2012/6.pdf](http://www.Aijcnet.com/journals/vol_2_No_March_2012/6.pdf)
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- McWhorter, T. K. (2012). *Successful college writing* (5th ed). New York: Bedford/St. Martin' S.
- O'Neill, P, Moore, C. & Huot, B. (2009). *A guide to college writing assessment*. Logan: Utah State University Press.
- Savignon, J. S. (1997). *Communicative competence: Theory and classroom practice: Texts and contexts in second language learning*. UK: Open University Press, McGraw-Hill Education.
- Shahzad, K. (2011). *An evaluative study of spoken language test performed at NUML: A case study*. An unpublished thesis. International Islamic University, Islamabad.
- Shahzad, K. (2017). Designing a question paper: A critique of essay type questions constructed to assess written discourse of English language students. *Pakistan Journal of Languages and Translation Studies*. 1. 24 – 39.

Shahzad, K. (2018). *Analyzing English language teaching and testing practices in developing discourse competence in essay writing*. An unpublished thesis. International Islamic University, Islamabad.

Truscott, J. (2007). The effect of error correction on learner's ability to write accurately. *Journal of Second Language Writing*. 16. 255-272.

Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.