Vol. V, No. IV (Fall 2020)
p- ISSN: 2520-0348
e-ISSN: 2616-793X
ISSN-L: 2520-0348

Global Social Sciences Review (GSSR)
URL: http://dx.doi.org/10.31703/gssr.2020(V-IV).14
DOI: 10.31703/gssr.2020(V-IV).14
Pages: 120 – 136

Mahwish Farooq* | Khalid Ahmed† | Sahirish Farooq‡

Introducing Self-Assessment for Evaluating Learners in Pakistan

Abstract *'This study deals with the positive backwash effect of self-assessment on Second Language Learners in Pakistan. The hypothesis of this study is self-assessment will positively improve learners' performance. Therefore, small scale experimental research was conducted with fifty undergraduates of a public sector university. They were asked to write essays; those were also assessed by students themselves and also cross-checked by their language teacher. The 'Assessment Rubric of Punjab University for BS-level' was a research tool. Afterwards, all results were statistically analyzed in SPSS, i.e., dependent t-test. The results conclude the improvement in self-assessment to highlight the needs and demands for taking bold steps in traditional evaluation criteria that would ultimately become the cause for students' active participation and self-learning.*

Key Words: Self-assessment, BS-rubric, Assessment

Introduction

Writing is a crucially important skill; therefore, it is being taught with great effort (Fahimi and Rahimi 2015). A written manuscript may be assessed differently by considering the demand for a learner, subject, topic, and situation. In Pakistan, the formal assessment and evaluation are done by teachers; therefore, learners are kept uninformed about the used assessment criteria, demands of evaluation, and their weaknesses. Consequently, they would not be able to assess their performance and mostly left behind the assessment demands of a teacher. It's an understood reality that assessment is an essential part of the education system, used for the evaluation of learners' achievements after teaching specific material. It can be done either by following teacher-centred pedagogy or learner-centred pedagogy but, teachers are active in teacher-centred pedagogy, which is equally being used in Pakistan. Contrary to this, learners can actively participate in evaluating their learning which is called self-assessment (Shahrakipour 2014), (Alibakhshi and Sharakipour 2014). Past literature has confirmed that assessment is not only the responsibility of teachers but an equal and mutual responsibility for teachers and students. Such type of collaboration would prove beneficial for the whole educational cycle, i.e., teacher, learner, and institution.

Formally, self-assessment is a successful way to influence learners' performance, but it is the most neglected area in language classrooms. Bailey has defined self-assessment as a series of procedures that allow the learners to evaluate themselves after assessing their language knowledge and skills (as cited in Shahrakipour, 2014). Active participation of the learners is clear and obvious, which motivates them to take more responsibility and brings ultimate confidence in learners. It makes them able to judge their learning, weakness, and strength. This is an essential element of assessment but still has been ignored in Pakistan. Therefore, this work will prove an initiative to study the effect of self-assessment on the writings of second language learners in the public sector universities of Pakistan. For that reason, undergraduates of the 3rd semester have been selected as a research sample for English essay writing. The written essays are used as a research corpus. The essay writing has been

* Department of English, University of Central Punjab, Lahore, Punjab, Pakistan. Email: mahwish.farooq@ucp.edu.pk
† Assistant Professor, Department of English, University of Central Punjab, Lahore, Punjab, Pakistan.
‡ Department of Education, Minhaj University Lahore, Lahore, Punjab, Pakistan.

Citation: Farooq, M., Ahmed, K., & Farooq, S. (2020). Introducing Self-Assessment for Evaluating Learners in Pakistan. *Global Social Sciences Review, V*(IV), 120-136. https://doi.org/10.31703/gssr.2020(V-IV).14

done in two rounds; in the first round, a few instructions have been given to participants, but in the second round, the assessment rubric has been practised for fifteen days for a better understanding of the learning assessment. Afterwards, the results of both rounds have been analyzed and compared using a dependent t-test for knowing the backwash effect of self-assessment.

Hypothesis

Self-assessment of L2 learners has a positive backwash effect on their performance and learning. Therefore, this research will be able to answer the subsequent questions.

1. Can students evaluate their written performances formally?
2. Do you think teacher-based assessment and self-assessment have a significant relationship?
3. Does the self-assessment of L2 learners improve over time?
4. In the last cycle of assessment, do learners assess different components of their writing in the same way as the teacher did?
5. Can the self-assessment be used as a tool to improve the writing errors of language learners?

Research Significance

The focal point of the current study is a promotion and implementation of self-assessment for making improvements in learning. As "learners may come out of the bad effects of teaching but never come out of the bad effects of assessment", that's why there is a big need to force learners and teachers to use innovative and modern assessment techniques in the education system.

Literature Review and Background

Self-assessment may be an innovation in language teaching, but unfortunately, being used in very limited areas of education. Firstly, George Jardine (1774) has described a 'pedagogical plan' that includes peer and self-assessment methods. Then, Hounsell and McCulloch (1999) reported in a survey that almost a quarter of evaluations involved self-assessment. Later, different reviews have been done by several researchers (Boud & Falchikov, 1989; Brown & Dove 1991; Boud, 1995; Topping, 1998; William & Blake, 1998; Dochy, Segers, & Sluijsmans, 1999; Falchikov & Goldfinch, 2000). The reason is teachers, educators, and researchers are interested in developing new strategies by reducing the cost. The other reason is to achieve long-lasting results in the learning outputs of second language learners. Because the purpose and nature of assessment affect different facets of student's performance, including goal orientation and anxiety (Topping, 2003). A stepwise introduction of self-assessment has been discussed in subsequent sections.

Communication Skills

The mutual interaction between human beings is called communication. In other words, communication is a sharing process to create thoughtful ideas (Willkommen and Team 2010). Therefore, it is required to identify; (i) a routine problem, (ii) key components of communication, (iii) different strategies for handling various contexts, and (iv) making awareness about the social behaviours. As communication is a cyclical process that starts and ends with the speaker where a message is transmitted to the listener, and the listener sends feedback to the speaker, which ensures communication (Dixon and Hara 2016). According to Willkommen and Team (2010), it has multiple parts and stages i.e.

(i) Context
(ii) Message
(iii) Source
(iv) Encoding
(v) Receiver
(vi) Decoding
(vii) Channel, and
(viii) Feedback.

© 2020 The Author(s). This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

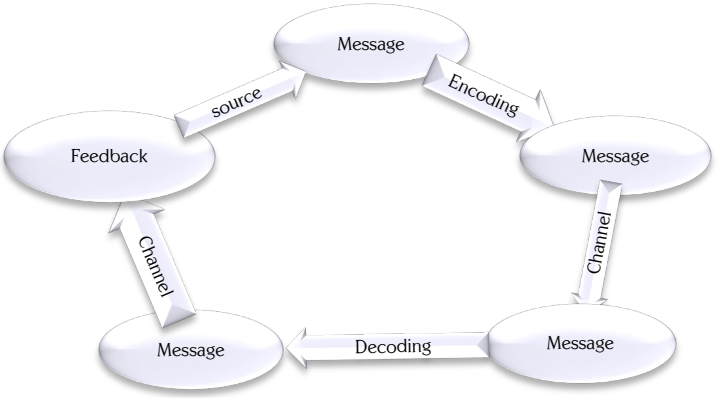

Figure 1: Communication Cycle

Communication may be verbal or non-verbal; verbal communication is more essential and clear, while (ii) non-verbal communication may be ambiguous. According to an estimation, we spent almost 70-80 per cent of the time (including 9% writing, 16% reading, 30% speaking, 45% listening) in communication using one to one interaction, attending the conference, etc. Therefore, we can say that listening, reading, writing, speaking skills, and essentially required for conveying and convincing people; otherwise, communication will be unproductive and flopped. Therefore, persuasively effective communication skills are demanded career advancement in any field of life (Worth 2004). According to the delimitation of research, this study will only highlight the writing skills.

Writing Skills

Writing is the most important element in expressing ideas. It is demanded to be brief, comprehensive, and clear. Therefore, there is a formula for writing well "express like a common man but think wisely" , which means that one must define a purpose to write with a clear objective which appeals quickly. It can be achieved by writing a problem statement, summary sentence, or subject line (as cited in Worth, 2004; Corbis). If someone fails to write a "focused" sentence may face failure. Therefore, good writing must be reader-friendly, persuasive, explanatory, but has correct short sentences, avoid repetition by doing proofreading. In short, a writer must keep in mind 4C's formula before starting to write, i.e. to be concise, correct, compelling, and clear. If a writer tries to follow this strategy, then it will be called "pyramid" writing style, which starts with the most important point and ends with three or four less focused or least important points (Worth, 2004).

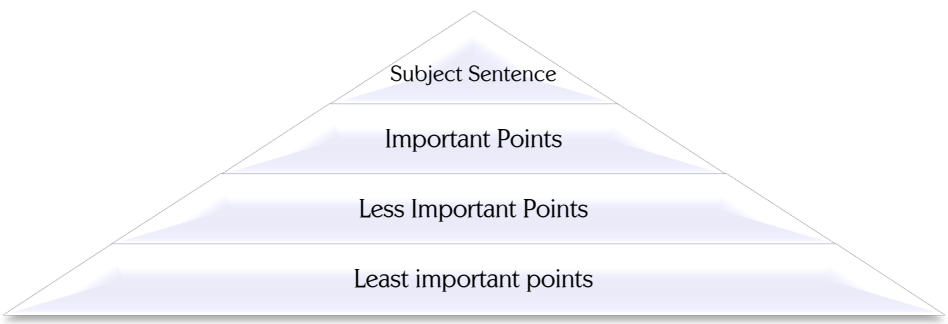

Figure 2: Pyramid Writing Style

Assessment

Writing has been considered a complicated job, and its assessment has been considered a challenge for evaluators. Teachers adopt different assessment methods for evaluating language learners to deal with the assessment criterion problem. Among them, the use of an assessment rubric is a systematic way to evaluate the linguistic and discourse features of a written paper (Razi 2015).

Rubrics and Significance of Rubrics in Assessment

An assessment could be automated electronically or manually (James 2009), but an authentic assessment is essential for maintaining validity and reliability (Jonsson and Svingby 2007). Teachers are human beings, so there is a possibility of committing an error in assessing monotonous rhetorical performances of different students; therefore, there is a need to follow some fixed criterion by using assessment rubrics (Petruzzi 2008) as are being considered reliable for two decades (Silva 2014). It is also known as a marking guide and marking scheme for learning assessment (Razi, 2015). Andrade (2000) focuses raters for using carefully designed instructional rubrics; they are the scoring tools that enlist a criterion for evaluating written work. Evaluators use rubrics due to five reasons; (i) tool for assessment, (ii) helpful to judge, (iii) saves time, (iv) accommodates heterogeneous students, and (v) easy to use (Andrade 2000). A rubric may be prepared by each evaluator individually, or a rater may use a "readymade" rubric. It is also true that an already available rubric has common assessing and evaluating features (Hyland 2009), but the best rubric is only that which is developed by every individual teacher by considering teaching objectives (Comer 2009). Because according to different works [(Cumming 1997), (East and Young 2007)], there are three types of assessments; (i) analytical, (ii) holistic, and (iii) primary.

Analytical Assessment

An analytical assessment demands deep analysis of written components by checking unity of thoughts, fluency of ideas, coherence, the level of formality, etc. (Becker 2011).

Holistic Assessment

In a holistic assessment method, the evaluator quickly assesses the writing skills of learners despite their weaknesses. In focused holistic assessment, learners' scores have been compared with the expected performance of the learners at different proficiency levels. Several problems have been reported against the holistic assessment. But it is largely used by the raters due to the practicality factor for saving time. Primary assessment is also considered a part of the focused holistic assessment and least common among raters. It focuses on individual writing tasks, e.g. finding differences among different kinds of writing essays (Razi, 2015).

Students' Self-Assessment

Self-assessment means the evaluation of learners by themselves. It has two benchmark standards that are used for language assessment. One is the perceived language proficiency of peers, and the other one is their difficulty in routine tasks of communication. It is also considered as one condition for second language learning and the author's construct for the 'locus of control'. Communication locus functions as an interface between second language acquisition and assessment research (Peirce, Swain and Hart 1993).

Many teachers have collaboratively involved students in learning and evaluation during classroom activities. Some teachers have included self-assessment as a part of the summative assessment, which precisely indicates weaknesses and strengths by constructive individual feedback. In self-assessment, learners are involved in the assessment process, which becomes the cause of positive self-improvement and learning. According to them, triangulation of self, peer, and teacher assessment plays the role to incur the hidden threats of learning (Topping 2003). The main purpose of the research is the knowledge about learners' strengths gained after self-assessment. It provides information about the achievement of students. Although it is widely in use, teachers have doubts about the accuracy and value of this technique. It gives inconsistent results across items and time (Ross 2006). Likewise,

few types of research are against self-assessment because, according to them, by using this methodology, neither performance nor proficiency level improved of learners (Dieten 1989).

Although it has been widespread in the field of sociology, psychology, business, its usage is quite rare in second language learning. Whether this situation stems from skepticism and disbelieves about students' ability to provide proper information about their capabilities to use language, the other reason may be inappropriate knowledge about how to use self-assessment, but still considered a valuable tool along with other instruments (LeBlanc and Painchaud 1985). Different self-assessment studies are contradictory to some extent, and these differences have supported the 'Monitor Model and Theory' presented by Krashen. Therefore, the teacher/researcher should be aware of the variant degree of influence on the self-assessment of foreign language learners (Blanche and Merino 1989).

Self-regulated learning emphasizes the role of self-assessment, and the reason is the conscious reflections of the performances of a learner increase the frequency rate of accuracy. It is further divided into many sub-processes such as self-instruction, self-monitoring, self-correction, self-reinforcement, self-evaluation (as cited in Vanderveen, 2006; Mace et al., 2001), the self-judgment, the self-observation, and the self-reaction (as cited in Vanderveen, 2006; Zimmerman, 1989), but the separating line is unclear between these assessments (as cited in Vanderveen, 2006; Benson, 2001) because they are interdependent. Self-monitoring is defined as checking comprehension while reading, writing, listening, or speaking. Self-monitoring contrasts with self-assessment in checking the outcomes of a learner's performances but also have fixed and standard criteria (as cited in Vanderveen, 2006; O'Malley & Chamot, 1990).

Self-monitoring and self-assessment have been considered the single construct referred to as conscious evaluations which are recorded usually for achieving learning tasks. Self-assessment is explored regarding different aspects of learning for the modifications. Even though, it proved unsuccessful for increasing learning and productivity (as cited in Vanderveen, 2006; Shapiro & Ackerman, 1983), but Cresswell (2000) and Charles (1990) had emphasized the importance of self-assessment by evaluating the written notes or annotations (as cited in Vanderveen, 2006). Extensively argued that self-assessment is good to enhance learning (as cited in Vanderveen, 2006; Wenden, 1991; O'Malley & Chamot, 1990; Blanche & Merino, 1989) by providing an estimation of time needed for self-assessment and its relationship with the intervention size (Vanderveen 2006).

Traditional assessment is often considered as the monarchy of a teacher which captured the attention of scholars, but it can be triggered a shift to an alternative assessment i.e., self-assessment, portfolio assessment, peer-assessment, performance assessment, and so forth. Self-assessment is a kind of assessment tool for evaluating learners' language learning competencies (as cited in Baleghizadeh & Masoun, 2013; Huerta-Macias, 1995). Oscarson (1997) advocated self-assessment based on effective learning and for achieving better results. According to this condition if learners are engaged in a continuous process of learning then all other types of assessments are considered lesser. The most important advantage of self-assessment is the achievement of a confident performance. Moreover, the perceived self-mastery and confidence are the outcomes of self-assessment which would ultimately contribute self-efficacy of learners (Baleghizadeh and Masoun 2013).

Learning

Assessment is a matter of supreme importance because it affects the process of instructions. It is needed for the process and product of learning to know what is learned. Learning and assessment are intertwined therefore there is a growing demand for knowing lifelong learning. It is achieved when reevaluating the relationship between learning and assessment (as cited in Baleghizadeh & Masoun, 2013; Dochy et al., 1999).

Language Learning

Mostly, reading leads towards rote memorization and retaining materials for meaningful learning. Different research have been done for the improvement of learning using innovative strategies. Mind mapping is an important strategy to connect different ideas. Two main theories support this concept in language learning i.e., constructive and assimilation theory. The "Constructivist theory" implies that learners take their prior knowledge in the class, that is considered highly influenced with cultural and

ethnographic factors, but they believe in individual assessing ways. In other words, a learner's knowledge construct with their personal experiences. So, the connections would be made between the previous and novel information when learners wanted to learn in a meaningful manner. Therefore, the "assimilation theory" introduced by Ausubel classifies learning in two ways, i.e., (i) rote learning, and (ii) meaningful learning. So, the "meaningful learning" occurs when the learner intentionally relates the new knowledge with the prior knowledge and "rote learning" occurs in response to senseless cramming. It has been considered that concept mapping may contribute by using these theories (Khajavi and Ketabi 2012).

Language learning achieved its goals through communication. It is a key symbol for learners' engagement in learning another language. Group activities have been considered a basic tool in language learning because it provides various chances to the learners for better communication (as cited in Baleghizadeh & Masoun, 2013; Harmer 2001; Jacobs 1997; Jacobs, Crookall & Thiyaragarajali 1997). A cooperative group is committed to a common purpose for maximizing learning. Therefore, measures have been taken for making a cooperative group which ultimately proves the key to successful learning (Dabaghmanes et al., 2013).

Backwash Effect

"The effect of testing and learning is known as backwash". It may prove beneficial as well as harmful for the learners therefore its preparation must be considered more important. If testing techniques are wrongly adopted, it will harmfully affect the students' learning. So, there is great pressure for practicing the desired language skills. The backwash effect also has a positive effect therefore, a test is intensively designed which must be based on direct assessment of the required skills. Language testing has been conducted differently for different syllabus, chosen books, selected classes, level of students, types of assessment, time of assessment, and their proper use would cause a beneficial backwash effect. There exist a strong relationship between assessment and teaching. Sometimes teaching is good but testing may be bad because "learners may come out of the bad effects of teaching but never come out of the bad effects of assessment". Therefore it would be more supportive because it has been considered that even in the case of bad teaching, good testing leaves a positive backwash effect on learning (Hughes 1992). For better assessment and learning, the present research is also dealt with the implementation of self-assessment in the traditional evaluation system of L2 learners in Pakistan.

Methodology

Population

L2 learners of the public sector university of Pakistan are selected as the population of this research.

Sample

One teacher and fifty undergraduates of the second language are selected as a sample for the assessment of written essays. The participants are both male and female students of BS 3rd semester and their age ranges between 18-22 years.

Statistical Tool

SPSS software (version 20) is used for data analysis and interpretation.

Instrumentation

Instrument 1

The essay writing rubric of Punjab University is selected for the assessment of students. The same rubric is shared with learners as well as with the teacher. This assessment rubric has three main scales. Though, the researchers thought that students would need to know how these three components (organization, language, and vocabulary) would break down into smaller sub-scales for assessing an essay. Therefore, it becomes clear that the organization refers to sub-scales, such as the introduction,

body paragraphs, and conclusion. Similarly, language refers to the selection of words which refers to the choice and variety of appropriate vocabulary. For the participants, it is quite easy to use this checklist to assess the sub-categories of content.

Table 1.

Content	Marks Allotted
Organization	7
Language	3
Vocabulary	3
Writing Format	2

Instrument 2

A questionnaire is the second instrument of this study (see Appendix B). The researcher has been developed a structured and quantitative questionnaire. It is a controlled questionnaire in simple language to ensure easy comprehension. The main purpose behind the formation and conduction of this questionnaire is to elicit the learners' attitude for self-assessment, their opinions about its worth informal evaluation, how they had self-assessed their performance, and how self-assessment improves their learning.

The questionnaire is primarily useful for saving time as compared to collecting information by interviewing all the participants individually, but it works as a semi-structured interview that the researcher has conducted with the participants. It is also considered that the use of a questionnaire may enlighten some dark points, i.e., students would either overestimate or underestimate their performances.

Procedure

The study is conducted with the second language (L2) learners in a public sector university of Pakistan and has been completed in three weeks. Therefore, the methodology is divided into six steps (i.e. 1st Essay writing, self-assessment1, teacher-assessment1, 2nd Essay writing, self-assessment2, teacher has beeassessment2, along with their comparisons). Firstly, students have written another essay using the shared rubric. Then, written essays have been photocopied, 1 copy for the learner, and 1 for the teacher. Moreover, students have assessed their essay by using the already-given rubric, but without giving any instruction to use the rubric correctly. Furthermore, copies of the (firstly written) essay have also been given to the teacher for assessment 1. Later, both results are compared with each other to cross-check the difference between self-assessment and teacher-assessment. The results of self-assessments 1 are not favorable because students are not acknowledged with the self-assessment and the usage of the assessment rubric. Therefore, the teacher (or researcher) has trained learners by practicing it for fifteen days.

Though the students are undergraduates of BS 3rd semester some of them have problems with writing an essay. Therefore, the teacher (researcher) has devoted a little time to provide some instructions on an appropriate format, length, content, and organization of the essay. Afterward, the students are introduced to the rubric (Appendix A) and its usage. The researcher has explained the category and its sub-categories. Then after initial training of 15 days, students have been asked to write another essay again with the same topic keeping in mind their practice. The reason for selecting a limited period is to control new 'learning' or 'de-learning' which would appear after a long time. Afterwards, written essays have been photocopied again, one for the learner's self-assessment2 using a rubric after training. Again, the second copy of each essay has been given to the language teacher for a second assessment to cross-check the students' evaluation criteria. Later, they were asked to evaluate their writings after two days for making a possible objective assessment. This would help the students to detach themselves from their writings and mark them critically. The essay topic has been kept the same for the acknowledgement of learning improvements. Furthermore, the researcher himself evaluated the essays by using an assessment rubric for grading the learning outcomes of students. It is also worth mentioning that assessors are asked to give a reason for justifying the numbers they have given to their essay. Consequently, the researcher and students have been commented on

in the margin of the paper. Later on, the results of both tests are compared to finding the backwash effect of self-assessment on the learning outcomes of learners.

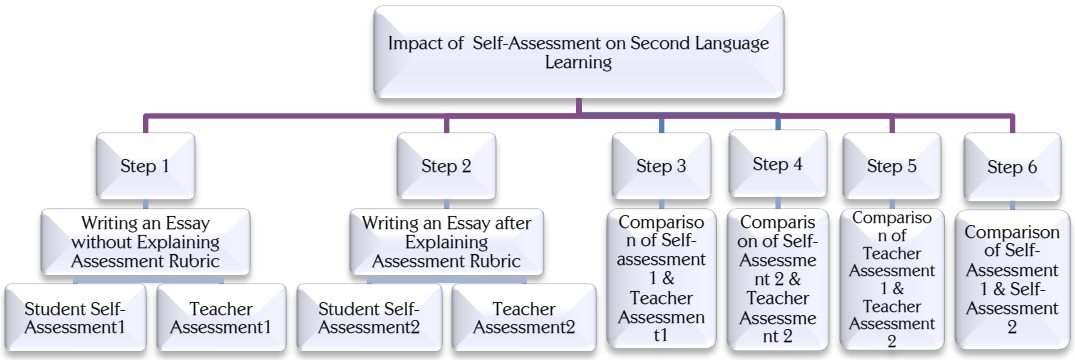

Figure 3: Backwash Effect of Self-assessment on Students' Learning

Then, the participants have given back their writings which have been evaluated two times i.e. firstly by the student and secondly by the assessor for knowing the possible similarity and difference. Moreover, the researcher has asked the learners to read their writings more than one time. The questionnaire has also been given to students to reflect on the assessment marks awarded by the teacher. Both assessments have been compared for finding out whether students have overestimated or underestimated their performances. The second main reason is the knowledge about the effect of self-assessment on students' learning by finding out the backwash effect. The whole process may be concluded as a cyclical process. The cycle of essay writing, self-assessment, and teacher assessment has been completed in three weeks.

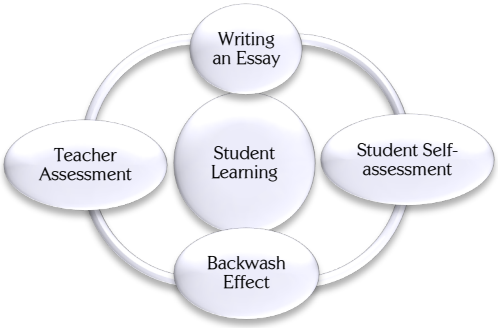

Figure 4: Learning Cycle

The students have not been informed about the research criteria. The majority of students have no earlier experience of self-assessment therefore, they have been trained. Along with the learners, a native Pakistani teacher (with five years' experience of teaching English as a Second Language) has a charge of teaching essay writing to the students and evaluating their written scripts. Therefore, an analytical scoring offers.

1. low inter-rater reliability between students and teacher,
2. low intra-rater reliability in both evaluations by the learners themselves
3. high intra-rater reliability in both evaluations by the teacher but only one rater (teacher) has assessed an essay.

After self-assessment 2, students are asked to fill a structured questionnaire to take a clearer idea about the self-assessment. Results and data analysis of the questionnaire claim that all students have been learning English since nursery class. 5 students said their assessment is close to the teacher. 25

students said that it's their first chance to assess their written essay formally therefor; their assessment is not close to the teacher. According to 10 students, previously 2-3 times they have assessed themselves but not equal to teachers' assessment criteria. Only 7 students have claimed to assess themselves 2-3 times and achieved equal marks to their estimation. Only 3 students denied any experience of self-assessment in their academic lives and it's their first chance. But remaining 47 students have confirmed their prior experience with self-assessment but also accept; that assessment was not part of their official evaluation. 28 students considered it as a good experience and tool for learning, but 20 students considered it good but effective for learning and only 2 students refused its importance by claiming no learning at all.

13 students admitted that they have overrated themselves, 35 students have told that they don't know about their assessment criteria. 2 students claimed that they don't know about their assessment. Thirty-five, second language learners have admitted that the teacher has done fair assessments while the remaining 15 claimed good assessments. A comparison of the self-assessment and teacher-assessment seems good for the improvement in learning. 35 students have claimed that self-assessment proves effective for the improvement in learning while 15 claimed it as an ineffective way. Students have accepted that being part of this experiment makes them aware of their mistakes and the demands of the teacher for writing an essay. 33 students have suggested that they must be trained for learning better while 17 have suggested that students must be trained for becoming active participants in classroom activities and discussion. In other words, all learners want to become part of this innovative technique of assessment. This would ultimately motivate them to participate in learning a second language by keeping them updated with the demand for evaluation.

Data Analysis and Results

Afterward, contrastive analysis has been done for cross-checking the results of both rounds of the self-assessment and the teacher-assessment methods. Therefore, statistical analysis of data has been done by using a t-test.

The Comparison of Self-assessment 1 and Self-assessment2

SPSS software is used for statistical analysis and dependent t-test pairs which is equal to Self-Assessment 1 with Self-Assessment 2 (paired). This command is used to test the null hypothesis of whether the self-assessment of both groups is equal. This t-test is done to know the intra-rater reliability of two assessments by the same group of students. Difference scores have been computed by subtracting the self-assessment1 from self-assessment 2. Firstly, SPSS has produced the mean, the number of pairs, the standard deviations, and the standard error to the self-assessment1 and self-assessment2.

Table 2.

A Paired Samples Statistics

		Mean	N	Std. D	Std. Error Mean
Pair 1	With Rubrics Self-Assessment 1	12.62	50	1.398	.198
	With Rubrics Self-Assessment 2	10.52	50	1.432	.203

Table 3.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Differrence				
					Lower	Upper			
Pair 1	With Rubrics Self-Assessment 1 With Rubrics Self-Assessment 2	2.100	1.607	.227	1.643	2.557	9.242	49	.000

A paired t-test is done by computing the set differences score where self-assessment 2 has been subtracted from self-assessment1 and mean value is shown under the "Paired Difference" and its value is equal to the difference between the mean of self-assessment 1 and self-assessment 2 (2.100), their standard deviation (1.607) and standard error (0.277), confidence interval (95 per cent) for a population mean of difference (self-assessment 1 – self-assessment 2).

The observed t-value has been calculated (9.242) wherein a mean difference (2.100) is divided by its standard error (0.227). 49 is the degree of freedom which means several pairs of observations minus 1) and the two-tailed *p*-value is "0.000". It does not mean the actual zero but if *the p*-value is less than .005 then SPSS rounds it off as zero (.000). The Critical t-value with 49 degrees of freedom and an alpha level of .001 (the smallest level of significance listed in most textbooks) for a two-tailed text test. Moreover, the t-value is greater than the critical t-value so the research null hypothesis is rejected. Consequently, the results of self-assessment 2 are better than self-assessment 1, and results also denied the intra-rater reliability of the same group. The reason is the objective and critical self-assessment in the second round. The other reason may be the practice with the assessment rubric which also proves helpful.

A Comparison between Self-Assessment 1 and Teacher-Assessment 1

SPSS software is used for statistical analysis and dependent t-test pairs which is equal to Self-Assessment 1 with Teacher-Assessment 1 (paired). This command is used for testing a null hypothesis whether the assessments of both groups are equal or not. This t-test has been done to know the inter-rater reliability of the teacher and students. "Difference scores" have been computed by subtracting self-assessment 1 minus teacher-assessment 1. Firstly, SPSS has produced a mean, standard deviation, number of pairs, and standard error of self-assessment 1 and teacher-assessment1.

Table 4.

A Paired Samples Statistics

		Mean	N	Std. D	Std. Error Mean
Pair 1	With Rubrics Self-Assessment 1	12.62	50	1.398	.198
	With Rubrics Self-Assessment 2	10.52	50	1.432	.203

Table 5.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	With Rubrics Self-Assessment 1 With Rubrics Self-Assessment 2	2.100	1.607	.227	1.643	2.557	9.242	49	.000

A paired t-test has been calculated after calculating a set difference where teacher-assessment 1 has been subtracted from self-assessment 1 and the mean value is shown under the "Paired Difference" and its value is equal to the differences of the mean of self-assessment 1 and the mean of teacher-assessment 1 (3.420), their standard deviation (1.751) and standard error (0.248), 95 percent confidence interval is for the population means of difference (self-assessment 1 – teacher-assessment 1).

The t-value (13.813) has calculated with a mean difference (3.420) is divided by its standard error (0.248) value. 49 is its degree of freedom and 0.000 is the two-tailed *p*-value. It does not mean that the *p*-value is zero but SPSS rounds of the figure if it's less than 0.005 (sig. value). Therefore, the Critical t-value with 49 degrees of freedom and an alpha level of .001 for a two-tailed text test. The t-value

(13.813) is measured greater than the critical t-value therefore the null hypothesis has been accepted and concluded that there is a huge difference in the self-assessment 1 and teacher-assessment 1 marking criteria.

A Comparison between Self-Assessment 2 and Teacher-Assessment 2

SPSS software has used for statistical analysis and dependent t-test pairs which is equal to Self-Assessment 2 with Teacher-Assessment 2 (paired). This command has used for testing the null hypothesis, "the assessments of both groups are equal or not?" This t-test has been done to know the inter-rater reliability of the teacher and students. "Difference scores" have been computed by subtracting self-assessment 2 minus teacher-assessment 2. Firstly, SPSS has produced mean, standard error, number of pairs, and standard deviation of self-assessment 1 and teacher-assessment1.

Table 6.

A Paired Samples Statistics

		Mean	N	Std. D	Std. Error Mean
Pair 1	With Rubrics Self-Assessment 2	10.52	50	1.432	.203
	With Rubrics Teacher-Assessment 2	9.44	50	1.459	.206

Table 7.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	With Rubrics Self-Assessment 2 With Rubrics Teacher Assessment 2	1.080	1.275	.180	.718	1.442	5.989	49	.000

A paired t-test has been measured after calculating a set of differences scores wherein teacher-assessment 1 has been subtracted from self-assessment 2 and the mean value is shown under the "Paired Difference" and its value is equal to the difference of the mean of self-assessment 2 and the mean of teacher-assessment 2 (1.080), their standard deviation (1.275) and standard error (0.180), 95% is the confidence interval for the population mean of difference (self-assessment2 – teacher-assessment2).

The t-value has calculated 5.989 after dividing the mean difference (1.080) with the value of its standard error (1.275). The value of 49 is the degree of freedom and 0.000 is the two-tailed *p*-value. It does not mean that the *p*-value is zero but SPSS rounds off the figure if it is less than 0.005 which means the difference is highly significant. The Critical t-value has 49 degrees of freedom and an alpha level of .001 for a two-tailed test. The t-value (5.989) is larger than the critical t-value therefore the null hypothesis has been accepted and concluded; the difference between the self-assessment 2 and teacher-assessment 2 marking criteria is huge and larger. But the results of self-assessment 2 are better than self-assessment 1 due to the familiarity and practice with the assessment rubric.

Comparison between Teacher-Assessment 1 and Teacher-Assessment 2

SPSS software has used for statistical analysis and dependent t-test pairs which is equal to teacher-assessment 1 with teacher-assessment 2 (paired). This command is used to test the null hypothesis of whether both assessments by the same group are equal or not. If the assessment scores are different

by having more marks, then it means that learning has taken place among students. Difference scores have been computed by subtracting the teacher-assessment 1 minus teacher-assessment 2. Firstly, SPSS has produced a mean, standard deviation, standard error, and several pairs for teacher-assessment 1 and teacher-assessment2.

Table 8.

A Paired Samples Statistics

		Mean	N	Std. D	Std. Error Mean
Pair 1	With Rubrics Assessment 1	9.20	50	1.539	.218
	With Rubrics Teacher-Assessment 2	9.44	50	1.459	.206

Table 9.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	With Rubrics Teacher Assessment 1 With Rubrics Teacher Assessment 2	-.240	1.533	.217	-.676	.196	-1.107	49	.274

A paired t-test is extracted after calculating a set of differences' scores by teacher-assessment 2 has been subtracted from teacher-assessment 1 and the mean value is shown under the "Paired Difference" and its value is equal to the difference for the mean of teacher-assessment 1 and mean value of teacher-assessment 2 (-0.240), their standard deviation (1.533) and standard error (0.217), the 95 percent is a confidence interval for the population mean of differences (teacher-assessment 1 – teacher-assessment 2).

The t-value (-1.107) has been calculated after dividing the mean difference (-0.240) with the standard error (0.217). The value 49 is a degree of freedom and the two-tailed *p*-value is 0.274. It means that the *p*-value is larger than .005 which is representing insignificant results. Critical t-value with 49 degrees of freedom and an alpha level of .001 for a two-tailed text test. The t-value (-1.107) is larger than its critical t-value therefore, the fourth null hypothesis of this study has been rejected. It concludes no difference in the marking criteria of teacher-assessment 2 and teacher-assessment 1. It might be the teacher's earlier familiarity with the assessment rubric.

Conclusion

Self-assessment of second language learners has a positive backwash effect on their performance in essay writing assessment. After analyzing the data, it is concluded that students can evaluate their written performance with some instructions and guidance. They learn how to use an assessment rubric with little instructions which also proves its effectiveness. It is also concluded a significant difference between self-assessment and teacher assessment. In the second round of assessment, this difference reduces after understanding the rubric and its usage. There is a huge difference in the self-assessment 1 and teacher-assessment 1 marking criteria. It is also concluded; there is no inter-rater reliability between teacher and students, therefore a student cannot be a part of the formal evaluation (4.3.). The results of self-assessment 2 are better than self-assessment 1 and suggest the good results would be changed with more practice and instructions (4.4.).

Comparative analysis of self-assessment 1 and self-assessment 2 shows improvement in their evaluation criteria. For example, L2 learners have over-marked their written essays in the first session while in the second round they have evaluated themselves more objectively. The results of self-assessment 2 have improved than the results of self-assessment 1 because students have evaluated themselves more objectively and critically in the second round. Although, they have given themselves

fewer marks than earlier it shows their level of better understanding the self-assessment (4.1.). 70% of L2 learners assess (35 out of 50) different parts of written essays very similar to the teacher. 20% (10 out of 50 students) show improvement in self-assessment which could not be generalized but it can improve with practice and concentration. These results have also denied the opinions of L2 learners shared in the questionnaire that they have evaluated themselves more objectively, but in actual they have overrated themselves in both assessments. It is also true that learners become more realistic in evaluating themselves during the second round. It is also concluded no difference in the marking criteria of teacher-assessment 2 and teacher-assessment 1 so the teacher's intra-rater reliability is more than the students' intra-rater reliability.

The purpose of research is to improve the quality of learning and formal assessment. Self-assessment is also needed to reduce the cost of teachers' summative assessment. Therefore, this research gives enough information for self-assessment by answering the questions which have been outset earlier. The results of the research show the reliability of the self-assessment assessment technique, but it is not enough to be a part of a formal summative assessment. In the future, it may achieve consistency in results across items by doing practice. Self-assessment reliability can be achieved by engaging L2 learners in the construction and use of the assessment rubric. Then the assessment discrepancy can ultimately be reduced with training and practice.

Besides this, awareness about the assessment differences would lead to productive and healthy conversations between teachers and students. Then, students will ultimately know their individual needs and lacking. The main finding is; the teacher would know the required skills for language teaching with the discussion but accurate self-assessment is the prior condition for it which would be achieved with practice. It has also been concluded that very limited work has been done in this area especially in Pakistan. The Self-assessment method is an effective way of improving the written performance of students. It also helps the learners for giving exact responses and if they will realize it then they would understand that they have not been judged only with marks. Moreover, it helps students to become more involved and motivated in learning by assessing them in the classroom. These are practical ideas that have been supported concerning literature review. Self-assessment can be established by linking theory with practice. It's a unique feature to include in the formal assessment typology of language learning by following the recommendations to implement. But in actual it can be useful for anyone concerned with learning.

Recommendations and Future Work

The research proves suggestive for teachers by giving evidence; self-assessment produces more valid and reliable information about the language learning achievement of L2 learners.

- There is a big need to provide proper training to students but before them, teachers' training is most essential.
- Teachers should make a thoughtful and grave commitment to using the self-assessment for improving students' learning.
- The promotion of self-assessment will enhance self-confidence, motivation, and a sense of achievement.
- Students would be instructed by the teachers to do practice in class and at home on their own.
- Last but not least, self-assessment should make a part of the formal evaluation policy for achieving better results and learning.

References

Alibakhshi, G., & Sharakipour, H. (2014, 08 06). The Effect of Self-Assessment on EFLLearners' Receptive Skills. *Jurnal Pendidikan Malaysia, 1*(39), 9-17.

Andrade, H. G. (2000). Using Rubrics to Promote Thinking and Learning. *Educational Leadership, 57(5)*, 13-18.

Baleghizadeh, S., & Masoun, A. (2013). The Effect of Self-Assessment on EFL Learners' Self-Efficacy. *TESL Canada Journal/Revue TESL DU CA, 31*(1), 42-58.

Baniabdelrahman, A. A. (2010, September). The Effect of the Use of Self-Assessment on EFL Students' Performance in Reading Comprehension in English. *The Electronic Journal for English as a Second or Foreign Language (TESL_EJ), 14*(2).

Becker, A. (2011). Examining Rubrics used to Measure Writing Performance in US Intensive English Programs. *CATESOL Journal, 22(1)*, 113-130.

Blanche, P., & Merino, B. J. (1989, September). Self-Assessment of Foreign-Language Skills: Implications for Teachers and Researchers. *A Journal of Research in Language Studies, 39*(3), 313-338.

Comer, K. (2009). *Developing Valid and Reliable Rubrics for Writing Assessment.* October 28, 2016, http://www.akoatearoa.ac.nz/ako-hub/good-practice-publication-grants-ebooks/resources/pages/developing-valid-and-reliable-rubrics-writing-assessment-researc

Cumming, A. (1997). The Listing of Writing in a Second Language. *Encyclopedia of Language and Education: Language Testing and Assessment, 7*, 131-139.

Dabaghmanesh, T., Zamanian, M., & Bagheri, M. S. (2013, December 8). The Effect of Cooperative Learning Approach on Iranian EFL Students Achievement Among Different Majors in General English Course. *International Journal of Linguistics, 5*(6), 1-11.

Dieten, A.-M. J.-v. (1989, June). The Development of a Test of Dutch as a Second Language: the Validity of Self-assessment by Inexperienced Subjects. *SAGE Journals: Language Testing, 6*, 30-46.

Dixon, T., & Hara, M. O. (2016). *Making Practice-Based Learning Work: Communication Skills.* Making Practice-Based Learning Work, University of Ulster, an educational development project funded through FDTL. www.practicebasedlearning.org

East, M., & Young, D. (2007). Scoring L2 Writing Samples: a scoring the Relative Effectiveness of Two Different Diagnostic Methods. *New Zealand Studies in Applied Linguistics, 13*, 1-12.

Fahimi, Z., & Rahimi, A. (2015, December 11 –13). On the Impact of Self-assessment Practice on Writing Skill. (A. W. Center, Ed.) *ELSEVIER, Procedia - Social and Behavioral Sciences, 192*, 730 – 736.

Gardner, D., & Miller, L. (1999, 3 11). *Establishing Self-Access from Theory to Practice*. January 19, 2017,https://books.google.com.pk/books?hl=en&lr=&id=qFlcoCN1JUoC&oi=fnd&pg=PR13&dq=A+project+in+self-access+learning+for+English+language+and+study+skills&ots=RFu48D-I0o&sig=b Gt9WujFzrmniJW _NGwZkAk8GTw&redir _esc=y#v=onepage&q&f =false

How to Improve College Reading Skills in 10 Steps. (2017, September). http://www. https://www.theclassroom.com/how-to-improve-college-reading-skills-in-10-steps-12138710.html

Hughes, A. (1992). Teaching and Testing. *Testing for Language Teachers, 4*. Cambridge University Press.

Hyland, T. A. (2009). Drawing a Line in the Sand: Identifying the Boarderzone between Self and Others in EL1 and EL2 Citation Practices. *Assessing Writing, 14*, 62-74.

James, C. L. (2009). Electronic Scoring of Essays: Does topic Matter? *Assessing Writing, 13*, 80-93.

Jeary, T. (1996). Inspire any Audience: Proven Secrets of the Pros for Powerful Presentation.

Jonsson, A., & Svingby, G. (2007). The Use of Rubrics: Reliability, Validity and Educational Consequences. *Educational Research Review, 2*, 130-144.

Khajavi, Y., & Ketabi, S. (2012). Influencing EFL Learners' Reading Comprehension and Self-efficacy Beliefs: The Effect of Concept Mapping Strategy. *Porta Linguarium, 17*, 9-27.

LeBlanc, R., & Painchaud, G. (1985, December). Self-Assessment as a Second Language Placement Instrument. *TESOL Quarterly, 19*(4), 673-786.

MacIntyre, P. D., Noels, K. A., & Clement, R. (1997, June). Biases in Self-Ratings of Second Language Proficiency: The Role of Language Anxiety. *Language Learning: A Journal of Research in Language Studies, 47*(2), 265-287.

Paolo, F. (1994). *How to Make a Great Presentation in 2 Hours*. Hollywood Fla: Lifetime Books.

Peirce, B. N., Swain, M., & Hart, D. (1993). Self-Assessment, French Immersion, and Locus of Control. *Oxford Journal of Applied Linguistics, 14*(1), 25-42.

Perfetti, C. A. (2001). Reading Skills. (N. J. Smelser, & P. B. Baltes, Eds.) *International Encyclopedia of the Social and Behavioral Sciences*, 12800-12805. Retrieved from Open Learn: http://www.openlearn.open.ac.uk/mod/resource/view.phd

Petruzzi, A. (2008). Articulatinf=g a Hermeneutic Theory of Writing Assessment. *Assessing Writing, 13*, 219-242.

Razi, S. (2015, April-June). Development of a Rubric to Assess Academic Writing Incorporating Plagiarism Detectors. *SAGE Open, 10*, 1-13.

Ross, J. A. (2006). The Reliability, Validity, and Utility of Self-assessment. *Practical Assessment, Research, and Evaluation, 11*(10), 1-13.

Routeledge. (2002). *Business Communication.* (P. Hartley, & B. G. C., Eds.) London.

Shahrakipour, H. (2014, Jan). On the Impact of Self-Assessment on EFL Learners' receptive Skills Performance. *AJTLHE, 6*, 1-13.

Silva, R. D. (2014). Rubrics for Assessment: Their Effects on ESL Students' Authentic Task Performance. *CELC Symposium*, 136-141.

Spitzberg, B. H. (1994). The Dark Side of Incompetence in Adler. (J. M, & E. B. R, Eds.) *Communicating at Work: Principles and Practices for Business and the Professions*.

Topping, K. (2003). Self and Peer Assessment in School and University: Reliability, Validity, and Utility. In M. Segers, & et al. (Eds.), *Optimizing New Modes of Assessment: In Search of Qualities and Standards* (pp. 55-87). Kluwer Academic Publishers, Printed in the Netherlands.

Vanderveen, T. (2006, 11 01). The Effect of EFL Students' Self-Monitoring on Class Achievement Test Scores. *Japan Association for Language Teaching*(28.2), 197.

Willkommen, K., & Team. (2010). *MTD Training Advanced communication Skills.* MTD Taining. Ventus Publishing ApS.

Worth, R. (2004). *Communication Skills* (2nd ed.). New York, United States of America: Career Skill Library, Ferguson.

Appendix

Rubrics of Essay Writing for L2 Learners

CRITERIA		ATTRIBUTES	4	3	2	1	0
ORGANIZATION	A.1. INTRODUCTION						
	A.1.1. Introductory Sentences	Effective introductory sentences					
	A.1.2. Thesis Statement	Appropriate thesis statement (thesis and central idea)					
	A.2. BODY PARAGRAPHS						
	A.2.1. Topic Sentence	Appropriate topic sentence (possibly implied) supporting the thesis and the central idea					
	A.2.2. Supporting Sentences	Appropriate sentences supporting the topic (possibly major and minor)					
	A.3. CONCLUSION	Appropriate conclusion related to thesis					
LANGUAGE USE	B.1. Word Order	Correct word order					
	B.2. Pattern Variety	Using different patterns					
	B.3. Verb Form	Using verb forms correctly					
	B.4. Tenses	Using tenses appropriately					
	B.5. Articles	Using articles correctly					
	B.6. Pronouns	Using pronouns correctly					
	B.7. Prepositions	Using prepositions correctly (verb + preposition, adjective + preposition)					
VOCABULARY	C.1. Word Choice	Selecting the appropriate words					
	C.2. Word Variety	Having a rich vocabulary					

Appendix B

Questionnaire

1. You have been learning English from which class?

a. Nursery b. primary c. Middle d. Matric

2. Have you ever evaluated your performance in an exam? If yes, has your estimation closed to your received marks?

a. Never assess me before
b. My first chance and my estimation are close to the teacher
c. My first chance and my estimation are not close to the teacher
d. Yes 2-3 times and never close to received marks
e. Yes 2-3 times and close to received marks

3. Was this experiment new to you? If not, then how did your self-assessment?

a. First experience
b. self-assessed before but as part of the official evaluation
c. self-assessed before but not as part of the official evaluation

4. What did you feel when you have assessed your essay?

a. Feeling Good, learning improved
b. Feeling good, but no learning
c. Feeling bad, no learning
d. Feeling bad, learning improved

5. Have you tried to be honest and objective for self-assessment? If yes, how?

a. Objective
b. biased about checking
c. don't know about my assessment

6. What are you thinking about your scores assigned by the teacher to you?

a. bad b. fair c. Good d. excellent

7. What do you think about comparing your self-assessment with the teacher's assessment will help you to evaluate your writings better?

a. Un-effective b. effective c. Good d. excellent

8. Will self-assessment help you to learn better?

a. Un-effective b. effective c. Good d. excellent

9. After doing the practice of self-assessment, will you like to have a chance to assess yourself by evaluating your written scripts? If yes, Why?

a. Liked without reason
b. don't like
c. liked as it makes me aware of my mistakes and the demands of a teacher

10. What's your opinion about the need for students' training in self-assessment and why?

a. Yes, students must be trained
b. no need to train
c. students must train for learning better
d. students must be trained for making them an active participant