Holistic vs. Category-Based Self-Assessment of Expository L2 Writing: Validity & Reliability Considerations

Massoud Yaghoubi-Notash

Department of English, Faculty of Persian Literature & Foreign Languages, University of Tabriz

P .O. BOX 51665331, Tabriz, Iran

Tel: +98 411 3392192   E-mail: masoud.yaghoubi@gmail.com

Abstract

A considerable body of research in EFL assessment seems to be motivated by the notion of self-assessment (see Sung et al., 2010, for example). In this research, essay writings of sixty-four major English learners were subjected to self- and teacher-assessments employing holistic vs. category-based scoring. The average of teacher scorings was used as the criterion for validity. Statistical analysis indicated that self-assessments were fairly valid, but not reliable. Also, holistic and category-based self-assessments correlated but not very highly. Findings imply that while self-assessment may provide a valid method for measuring learner performance in EFL, an unthinking application of self-assessment as a primary means of measuring learners' performance would be questionable. Another implication might be that in the cautious application of self-assessment as a partial representation of learners' performance, teachers and testers may instruct the learners to use both types of scoring as they empirically evoke similar self-judgments on the test-takers' part.

Key words: Self-assessment, writing, EFL

1. Introduction

Modern trends in education seem to emphasize the learners' role in the learning process on the one hand, and focus on the real performance of the learners on the other. These two dimensions account for a radical influence in many aspects of education including assessment. According to Sung et al. (2010)

> Recent learning perspectives have emphasized active learning, participation, social interaction, self-monitoring and self-regulation by students. . . . The new learning culture has not only influenced curriculum design and teaching methods, but has also encouraged innovation of classroom assessments. (p. 135).

The impact of alternative assessment a dimension of which is self assessment can also be felt in language testing literature (Davies, 2003a). Two influential meta-analysis studies, namely Blanche & Merino (1989) and Ross (1998), bear witness to the significance of the notion of self-assessment in ESL/EFL literature. Also, Ross (1998) points to the "cyclical regularity" (p. 1) with which the theme recurs in the literature.

Various definitions have been offered to characterize self assessment. Lindblom-Ylän, Pihlajamäki, & Kotkas (2006) define self-assessment as "the process in which students assess their own learning particularly their achievements and learning outcomes (p.52). Self assessment according to Blanche & Merino (1989) "…. is a condition of learner autonomy" (p. 313) with the latter notion, i.e. autonomy, fitting hand-in-glove with the modern language teaching paradigms (see also Harris, 1997). Baily (1998) argues that it is the set of "procedures by which the learners themselves evaluate their language skills and knowledge" (p. 227).

Despite the fact that self-assessment is now becoming a norm in the context of education, there are certain reservations about its use. Ross (2006) argues that

> A large proportion of teachers . . . . reports using self-assessment at least part of the time, even though teachers express doubt about the value and accuracy of student self-appraisals. The doubts center on the concern that students may have inflated perceptions of their accomplishments and that they may be motivated by self-interest.

Concerns about the value of self assessment can be divided into two types, that is reliability and validity.

Reliability is viewed as the internal consistency. Findings on reliability self-assessments within the broad context of education are unified, but not categorical. A number of studies report a high degree of internal consistency for self-assessment., for example, grade 5-6 students self-rated performance on five dimensions of mathematical problem solving and came up with an internal consistency of 0.91. In an earlier study, they obtained the same results with grades 4-6 doing English as their subject (Ross, Rolheiser, and Hogoboam-Gray, 1999). Fitzgerald, Gruppen, and White (2000) studied the self-assessment of medical students on performance tasks vs. cognitive tasks reporting consistency over a range of skills or tasks. Ross believes that "the evidence in support of the reliability of self-assessment is positive in terms of consistency across tasks, across items, and over short time periods" (Ross, 2006, p. 3). In their study, Butler and Lee (2006) showed that "the correlations between the on-task SA and the criterion measures appeared to be higher in general than those between the off-task SA and the criterion measures (p. 511). Miller and Ng (1994) focused on the reliability of pronunciation self-assessment which they found to be high. Dlaska and Krekeler (2008) examining the self assessment of advanced German learners' pronunciation came up with high correlation; however, they went on to argue that "If the reliability of the self-assessments is regarded as the most important factor, this study confirms that self-assessments of L2 pronunciation ought to be used cautiously" (p. 515). Brantmeier and Vanderplank (2008) concluded that "the nature of self-assessment has useful but ultimately limited reliability for reading placement" (p. 473).

On the other hand, validity is generally defined as the degree to which self-assessment scores agree with teacher judgment (Magin and Helmore, 2001; Topping, 2003; Lindblom-Ylänne et al., 2006 among others). Evidence surrounding the agreement between self-assessment and teacher's assessment (as reference point) is contradictory. Quite a lot of researchers have found that correlation between self-assessed scores and expert scoring are so low that it sounds unadvisable to rely on them in summative assessment (e.g. Lejk and Wyvill, 2001). A review of research on self-assessment dealing with qualitative analysis of learning products, like essays, indicated high accuracy on the students' part in grading their own essays (Dochy, Segers, & Sluijsmans, 1999). Some other studies indicate that students tend to overestimate their performance as compared with teacher-assessments (Zoller and Ben-Chaim, 1997). Bond and Falchikov's (1989) review of 48 studies supported the agreement between self- and the teacher; however, the reviewers were not confident about the quality of the studies. In the same study, Bond and Falchikov's (1989) suggest that there is a higher probability of overestimation on the students' part if their self-rating would mean a higher grade in the course. Topping (2003) states that self-assessed grades tend to be higher than staff grades. Sullivan and Hull (1997) reported that 39% of the students overestimated their performance and Oldfield and Macalpine (1995) found a low correlation between self- and teacher-assessments. Self-raters according to (Matsuno, 2009) "tended to assess their own writing more strictly than expected" (p. 91). Agreement of teacher and student self-assessments are reportedly higher when a) students are provided with instructions on how to assess their work (Ross et al., 1999; Sung et al, 2005), b) when students have information on the content and domain of the task (Longhurst and Norton, 1997; Ross, 1998), c) when there is an anticipation of comparison between assessment by the self with those of the peer or supervisor (Fox and Dinur, 1988) and d) when the application of the assessment criteria involves low level inferences (Pakaslati and Keltikangas-Järvinen, 2000).

Despite a wealth of literature on self assessment, not enough studies have ever investigated consistency of self assessment with different scoring procedures/approaches. The only study which roughly touches the issue is Lejk and Wyvill (2001) which is in favor of a holistic approach rather than a category-based, but of course for peer-assessment rather than self-assessment. The present study seeks to determine the reliability and validity of holistic and analytic scores obtained from the Iranian students' self-assessment of essay writing tasks, and the degree to which they are congruent. Therefore the research questions posed are:

R.Q. 1. Are the scores from holistic and category-based L2 writing self-assessments reliable?

R.Q.2. Are the scores from holistic and category-based L2 writing assessments valid?

R.Q. 3. Are the scores obtained from holistic and category-based L2 writing self-assessments congruent?

2. Method

2.1. Participants

Participants in the study were 64 (23 male and 41 female) Iranian students majoring in English language and literature at the University of Tabriz, Tabriz, Iran. They were doing their second and third year of their studies. Since they had already been admitted to university based on a standard entrance exam, no other test was administered to establish their level of proficiency. First language backgrounds constituted Azerbaijani (mother tongue of the natives to Tabriz and the surrounding cities and towns across Northwest of Iran), Persian, and very

few Kurdish (n = 3). The males' average age was 21.3 and that of the females 20.5. (The information contained here is for demographic uses to clarify the context of the study, and none of the variables such as age, L1 background or gender served as variables of any kind in the study).

3. Procedures

The writing task required of the learners was a 5-paragraph standard expository paragraph. To decide on the topic, several steps were followed. First, a pool of 50 topics from different textbooks and IELTS writing modules were chosen. Then, three English teachers were requested to choose 15 out of the 50 topics. Finally, the 15 topics were presented to the students themselves and they were asked to number them from (1) least interesting to the most interesting (15). The topic with the highest average rating '*What is the best way to choose a marriage partner?*' ($\bar{x}$= 13.57) was chosen as the topic to be developed by the participants in the essay writing task. The students were instructed to complete the task in 60 minutes, 20 minutes of which could be spent on drafting. They were not allowed to use any dictionary, or consult any reference grammar. When the allotted 60 minute was up, they were asked to stop writing, and hand in their essays (fair copies) and drafts.

Self-assessments occurred based on holistic, and then category-based scoring approaches with a two week interval between the two. This is because, methodologically, a shorter time in-between the ratings would run the risk of memory effect and a longer time would be impractical. For holistic scoring, they were asked to give their essays a score from 0-100. With the category-based self-assessment taking place two weeks later, each student was provided with the analytic essay evaluation sheet adapted from Jacobs, Zinkgraf, Wormuth, Hartfiel, & Hughey (1981) (see Appendices). To set a criterion as a basis for validity, the author (as the teacher) also assessed each of the essays both holistically and in a category-based fashion. For the latter the same adapted essay evaluation sheet by Jacobs et a. (1981) was employed. Then the mean of the teacher's two scorings were adopted as the criterion score. A corresponding time gap of almost two weeks was set for analytic and category-based scoring on the teacher's part. Also, the students' writing scores from the earlier semester were obtained which was based on the students' continuous assessment of the students writing during the whole semester. The mean of all three scores (i.e., teacher's holistic, teacher's category-based, and previous semester score on writing) was used as the criterion. Finally, the three sets of scores were fed into SPSS 17 for statistical analysis.

4. Data Analysis and Results

To determine reliability, internal consistency of the HSA and CSA scores were calculated using split half method. As for the internal consistency of HSA, the correlation between even- and odd-numbered HSA scores, the r value turned out to be − 0.10. With CSA, the correlation between two halves of the test is 0.09. Neither of the Pearson Correlation values was significant at $p < 0.05$.

Descriptive statistics for holistic self-assessment (HSA), category-based self-assessment (CSA), and criterion (C) show a trend of highest to lowest of means, that is 70.67, 69.42, and 63.05, respectively. However, with the standard deviations, the pattern is different. The highest SD belongs to the criterion (16.01), the lowest is represented by CSA (14.82), and slightly higher than that is HSA, i.e. 15.02 (see Table 1).

Table 1. Descriptive statistics for the four sets of assessment scores.

	Mean	Median	Mode	SD
Holistic self-assessment (HSA)	70.67	73.00	60	15.02
Category-based self-assessment (CSA)	69.42	70	82	14.82
Criterion*	63.05	63.41	58.17	16.01

* Criterion is the average of teacher's holistic, category-based scores, and the writing scores from the previous semester

A further look at the Table 1, shows that the closest distribution of set of scores to normal curve is that of the criterion, while the HSA is positively, and CSA is negatively skewed (see also Figures, 1, 2, and 3).
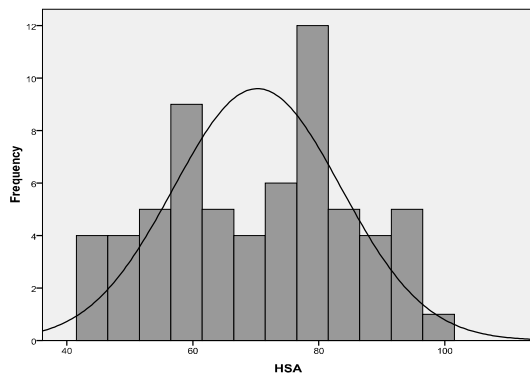
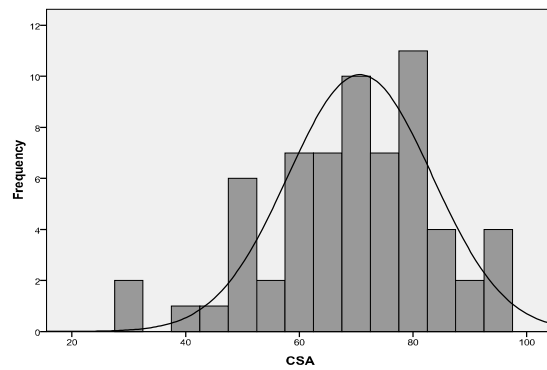Figure 1. Distribution of holistic self-assessment scores

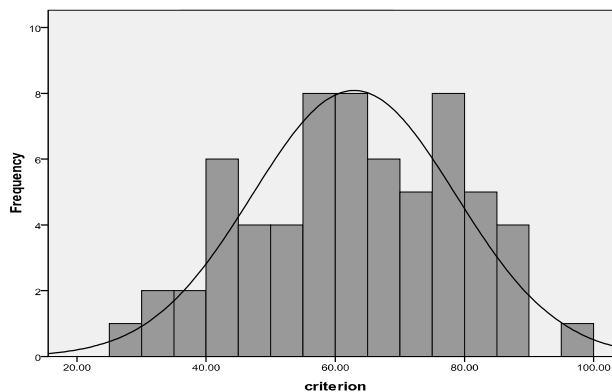Figure 2. Distribution of category-based self-assessment scores

Figure 3. Distribution of criterion scores.

Inferentials: The correlations (r value) among the three sets of score were analyzed using Pearson Product Moment. See Table 2 below:

Table 2. Correlations among holistic self-assessment, category-based self-assessment, and criterion.

	HSA	CSA	C
HSA			
CSA	0.52**		
C	0.29*	0.41**	

* significant at $p < 0.05$, ** significant at $p < 0.01$

The Pearson values (r) presented in the table above indicate that there is a correlation among all three sets of scores. The highest r value corresponds to the correlation between HSA and CSA (0.52) and the lowest pertains to that between HSA and C (0.29).

It is already an established fact that correlations do not provide a very strong basis for conclusions about differences or causality. Therefore, a further step is taken here to compare the significant mean differences between a) HSA, and CSA, b) HSA, and C, c) CSA and C using paired sample t-test. Results appear in Table 3.

Table 3. Paired-sample t-test results comparing the means between airs of assessments

Pair	t-value	df
HAS and CSA	0.68 (ns)	63
HSA and C	3.28**	63
CSA and C	3.04**	63

*ns = not significant; ** significant at $p < 0.01$

It can be understood from Table 3 that the mean difference between HSA and CSA is not statistically significant whereas HSA and C, and CSA and C pairs turn out to be significantly different regarding the mean of scores.

5. Discussion and Conclusion

Research question 1: Are the scores from holistic and category-based L2 writing self-assessments reliable?

The r value (Spearman rho) of the two halves of holistic self-assessments was − 0.10., which means the two halves of the test did not highly correlate. The same held true for category-based self-assessment in which r equaled 0.09. Thus, it can be concluded both holistic and category-based self-assessments lacked internal consistency or reliability. This finding goes against the findings by Ross, Rolheiser, and Hogoboam-Gray, (1999), Fitzgerald, Gruppen, and White (2000), Miller and Ng (1994), Dlaska and Krekeler (2008). However, it can be claimed coming up with no reliability in this study is perhaps one proof that the cautions for the limited reliability of self-assessments (Dlaska and Krekeler, 2008; Brantmeier and Vanderplank, 2008) must be overlooked.

R.Q.2. Are the scores from holistic and category-based L2 writing assessments valid?

Validity is defined as the degree to which the students' self-assessments go hand in hand with the teacher judgments (Magin and Helmore, 2001; Topping, 2003; Lindblom-Ylänne et al., 2006). That teacher rating is a reference point is further supported here since the criterion score distribution looks more inclined to normality due to the closer mean, mode and median values (see also, Figures 1, 2, and 3).

Results indicated that there is a correlation between HSA and criterion (r = 0.29), as well as between CSA and criterion (r = 0.41), with the former being significant at p < 0.05, and the latter at p < 0.01. However, these quantities are not high enough because normally, a value of 0.80 or above is considered an acceptable degree of correlation. For more empirical support, paired-sample t-test results were employed which clearly point out that mean differences between both HSA and C and CSA and C is high enough for them to be related. Such a finding is completely in line with Lejk and Wyvill, (2001) and Macalpine (1995) that report a low correlation between self-assessment scores and expert scores. Oldfield and Conversely, it contradicts (Dochy et al., 1999) who claimed a high agreement between self and teacher assessments. Again, low agreement supports the doubts expressed by Bond and Falchikov's (1989) about the qualities of 48 studies they had reviewed.

The means of HSA and CSA amounted to 70.67, and 69.42, respectively which are larger than the criterion (63.05), which stands for the fact that in both holistic and category-based assessments, the participants have overestimated their performance in essay writing. This is supported by (Zoller and Ben-Chaim, 1997; Topping, 2003; Sullivan and Hull, 1997) but disconfirmed by (Matsuno, 2009) who claimed that students tend to be strict in assessing their own writing.

Research Question. 3. Are the scores obtained from holistic and category-based L2 writing self-assessments congruent?

The correlation between the holistic and category-based self assessments was 0.52 which is significant at p < 0.05, showing that there is a correlation between the two sets of scores. However, r below 0.70 is normally considered moderate and not very strong.

To conclude, the findings of this study indicated that the scores from both holistic and category-based self-assessed scorings lacked internal consistency, but were valid, and that increase in one moderately predicated a corresponding increase in the other. Aside from supports and refutations in the literature, these findings may suggest that

a) Scorings did not demonstrate any consistency perhaps due to the want of a clear set of instructions on how to score. b) The scores were shown to be valid because they were very much close to how the teachers treated their writings. It can be claimed that teachers and students might have a shared understanding of writing quality in L2 shaped during years of instruction. And c) There was a fair level of correlation between their holistic vs. category-based assessment which might have been due to the fact that they followed their own perception of the quality of their writing more than they did the scoring scale.

Generally, self-assessment as it was partially represented by the findings can be a valid instrument for assessment in its own right; however, there is a long way to go to achieve consistency of measurement which is yet another requirement made of any and all types of assessment. In other words, findings from this study suggest that while such an assessment might be quite well-directed in what it claims to assess (as the definition of validity goes), the accepted level of consistency of scoring cannot be insured. Therefore, a clear implication

for assessment in ELT would be to remain on the safe side of overestimating its utility in contexts particularly when eliciting learners' performance and the related assessment is a high-stakes one. Teachers and assessment experts must see this technique as just one (and indeed not the only) source of assessing alongside others which merely give them one part of the "more comprehensive picture of the test-taker…" (Brown, 2004, p.111). In treating self-assessment with caution, assessment experts and teachers judiciously refrain from falling into the trap of traditional, single-shot test events thus fulfilling their commitment to the theoretical underpinnings of more modern and forward-looking alternative assessment.

Concentrating on learner-specific variables such as proficiency levels, gender, achievement, motivation and a host of other factors might have contributed to a better understanding of variations in self-assessment. Nevertheless, such a concern was outside the scope of analysis since each of the variables can constitute the main focus of further research, and in the context of this study they were seen as delimitations.

References

Bailey, K. M. (1998). *Learning about language assessment.* Cambridge, MA: Heinle & Heinle.

Blanche, P. & Merino, M. (1989). Self-assessment of foreign-language skills: implications for teachers and researchers. *Language Learning, 39(3)*, 313-340.

Boud, D. & Falchikov, N. (1989) Quantitative studies of student self assessment in higher education: A critical analysis of findings.   *Higher Education 18,* 529-549

Brantmeier, C., & Vanderplank, R. (2008). Descriptive and criterion-referenced self-assessment with L2 reader. *System, 36,* 456-477.

Brown, D.H. (2004). *Language assessment: Principles and classroom practice*. New York: Longman Pearson.

Bulter, Y. G., & Lee, J. (2006). On-Task Versus Off-Task Self-Assessments Among Korean Elementary School Students Studying English. *The Modern Language Journal, 90*, 506-518.

Butler, Y. G., & Lee, J. (2006). On-task versus off-task self-assessments among Korean elementary school students studying English. *The Modern Language Journal, 90*, 506-18.

Davies, A. (2003). Three heresies of language testing. *Language Testing, 20*(4), 355-36.

Dlashka, A., & Krekeler, C. (2008). Self-assessment of pronunciation. *System, 36*, 506-516.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24, 331e350.

Fitzgerald, J. T., Gruppen, L. D., & White, C. B. (2000). The influence of task formats on the accuracy of medical students' self-assessments. *Academic Medicine, 75(7)*, 737-741.

Fox, S. & Dinur, Y. (1988). Validity of self-assessment: A field evaluation. *Personnel Psychology, 41*, 581-592.

Harris, M. (1997). Self-assessment of language learning in formal settings. *ELT Journal, 51(1)*, 12-20.

Jacobs, H., Zinkgraf, S., Wormuth, D., Hartfiel, V., & Hughey, J. (1981). *Testing ESL composition: a practical approach.* Rowley, MA: Newbury House.

Lee, Y. (2006). The process-oriented ESL writing assessment: Promises and challenges. *Journal of Second Language Writing, 15*, 307-330.

Lejk, M., & Wyvill, M. (2001). The effects of the inclusion of self-assessment with peer assessment of contributions to a group project: A quantitative study of secret and agreed assessments. *Assessment and Evaluation in Higher Education, 26*, 551–561.

Lindblom-Ylän, S., Pihlajamäki, H., & Kotkas, T. (2006). Self-, peer- and teacher-assessment of student essays. *Active Learning in Higher Education, 7, 51-62.*

Magin, D. & Helmore, P. (2001). Peer and teacher assessments of oral presentation skills: how reliable are they? *Studies in Higher Education, 26(3)*, 288-297.

Miller, L. & Ng, R. (1994). Peer assessment of oral language proficiency perspectives. *Working papers of the Department of English, City Polytechnic of Hong Kong, 6,* 41-56.

Matsuno, S. (2009). Self-, Peer-, and Teacher- assessments in Japanese University EFL Writing Classrooms. *Language Testing, 29(1),* 75-100.

Oldfield, K.A., & McAlpine, J.M.K. (1995). Peer and self-assessment at the tertiary level: An experiential report. *Assessment and Evaluation in Higher Education*, 20, 125-132.

Pakaslahti, L., & Keltikangas-Järvinen, L. (2000). Comparison of peer, teacher and self-assessments on adolescent direct and indirect aggression. *Educational Psychology, 20*, 177-190.

Ross, J. A., Rolheiser, C., & Hogaboam-Gary, A. (1999). Effects of self-evaluation training on narrative writing. *Assessing Writing, 6(1),* 107-132.

Ross, John A. (2006). The Reliability, Validity, and Utility of Self-Assessment. *Practical Assessment Research & Evaluation*, 11(10). Available online: http://pareonline.net/getvn.asp?v=11&n=10

Ross, S. (1998). Self-assessment in second language testing: a meta-analysis and analysis of experiential factors. *Language Testing, 15(1),* 1-20.

Sullivan, K., & Hall C. (1997). Introducing students to self-assessment. *Assessment and Evaluation in Higher Education, 22*, 289-303.

Sung, Y., Chang, K., Chang, T., & Yu, W. (2010). How many heads are better than one? The reliability and validity of teenagers' self- and peer assessments. *Journal f Adolescnce, 33*, 135-145.

Topping, K. (2003). Self- and peer-assessment in school and university: Reliability, validity and utility in M. Segers, F. Dochy and E. Cascallar (Eds). Optimizing new modes of assessment: In search of qualities and standards (pp. 55–87). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Zoller, U. & Ben-Chaim, D. (1997*). Student Self-assessment in HOCS Science Examinations: Is it Compatible with that of Teachers?* Paper presented at the 7th EARLI conference, Athens, Greece, 26–30 August.

Appendices

Appendix 1

Composition	HSA*	CSA*	Criterion
1.	80	67	55.00
2.	68	78	77.00
3.	78	75	45.17
4.	52	32	32.67
5.	80	78	63.33
6.	90	62	68.83
7.	75	82	64.00
8.	50	63	43.50
9.	68	64	62.00
10.	75	72	66.83
11.	72	68	67.67
12.	45	60	61.83
13.	82	68	72.00
14.	51	52	50.33
15.	45	61	54.50
16.	66	52	82.17
17.	76	71	76.17
18.	85	65	56.33
19.	74	85	76.67
20.	77	53	81.00
21.	77	68	74.83
22.	44	79	82.67
23.	54	67	84.50
24.	95	90	75.50

25.	82	75	59.67
26.	95	87	75.83
27.	80	82	84.00
28.	84	79	72.33
29.	80	86	77.33
30.	60	65	59.00
31.	70	78	40.33
32.	64	69	67.50
33.	60	74	37.50
34.	65	58	76.50
35.	60	75	58.17
36.	88	64	60.33
37.	75	69	63.50
38.	60	82	64.17
39.	80	94	44.33
40.	60	87	88.00
41.	50	76	86.17
42.	100	97	95.17
43.	55	62	44.83
44.	84	82	66.00
45.	50	47	44.67
46.	81	59	58.17
47.	90	93	89.67
48.	60	80	47.00
49.	80	74	65.67
50.	96	82	46.83
51.	52	39	87.67
52.	67	94	74.00
53.	64	68	61.00
54.	59	49	33.33
55.	77	52	59.33
56.	88	92	74.17
57.	94	74	79.83
58.	64	72	54.33
59.	80	71	39.67
60.	96	59	56.17
61.	53	48	25.50
62.	44	30	52.33
63.	57	55	48.00
64.	60	52	43.17

*HAS: Holistic Self-assessment, **CSA: Category-based Self-assessment

Appendix 2

Essay Evaluation Sheet

	Excellent	**Good**	**Average**	**Poor**
Content	30-27 clear thesis, thorough development of thesis, no irrelevant sentence, appropriate length	26-22 clear thesis, limited development of thesis, mostly relevant to the topic, lacks details	21-17 Not clear thesis, Inadequate development	16-13 No thesis, Not relevant, Not enough to evaluate
Organization	20-18 Introduction Development paragraphs with topic sentences Logical order, cohesive, conclusion	17-14 Somehow choppy, limited support, Loosely organized	13-10 Disconnected ideas. No logical order	9-7 Does not communicate, no organization, not enough to evaluate
Vocabulary	20-18 Good range, effective word/idiom use, word form mastery	17-14 Adequate range, occasional errors of word/idiom usage, meaning not obscured	13-10 Limited range, frequent errors of word/idiom usage, meaning confused	9-7 Little knowledge of vocabulary, not enough to evaluate
Language Use	25-22 Effective complex structures, few errors of agreement, tense, number, word order, article, pronouns, prepositions	21-18 Effective but simple structures, some errors of agreement, tense, number, word order, ….. meaning seldom obscured	17-11 Problems in simple/complex structures, frequent errors of agreement, tense, number, word order, …. Meaning confused	10-5 Dominated by errors, does not communicate
Mechanics	5 Few errors of spelling, punctuation, capitalization, paragraphing	4 Occasional errors of spelling, …. Meaning not obscured	3 Frequent errors of spelling, ….. meaning confused, poor handwriting	2 Dominated by errors …… does not communiate

Adapted from Jacobs *et al*. (1981)