

Difference In Assessment: Towards A Dynamic Assessment Approach In Education

Minoo Alemi

Languages and Linguistics Department, Sharif University of Technology, Tehran, Iran
alemi@sharif.ir

Abstract

Discouraging the divorce between Language teaching and language testing, dynamic assessment has changed into a strong tool not only for language assessment but also for language development and learner involvement in the learning process. This study aimed to investigate (a) Iranian EFL (English as a Foreign Language) learners' off-task self-assessment and on-task self-rating of their writing ability and the impact of a dynamic-assessment-based course on the accuracy of their self-assessment and self-rating, and (b) the interrelationships among teacher rating, self-assessment, and self-rating in the writing performance of 22 engineering students taking a writing course as part of an EAP (English for Academic Purposes) program. Application of descriptive statistics accompanied by correlational analyses and *t*-tests showed that the participants overrated their on-task and off-task writing ability as measured against teacher rating before their exposure to the dynamic-assessment-based course. However, the difference between teacher rating on the one hand, student self-assessment and self-rating on the other decreased as a result of dynamic assessment, and strong, significant correlations were found among participants' self-assessment, their self-rating, and teacher rating. The results suggest that dynamic assessment helps Iranian EFL learners gain a better awareness of the criteria for writing evaluation and hence become more accurate in assessing their own writing ability.

Key words: Dynamic Assessment, Self-Assessment, Self-rating, Teacher rating, Education, EFL Writing

Introduction

The genesis of Dynamic Assessment is under the influence of the Sociocultural Theory of Mind (SCT) proposed by a prominent Russian psychologist L. S. Vygotsky, which includes a chain of interrelated consequences of development shedding light on the new ontological trends of human abilities. Contrary to the previously held practices that are typically dubbed Non-Dynamic Assessment, the current driving force of development within educational settings is heavily dependent upon a dialogically integrated activity of assessment and instruction (Poehner, 2005 [8], 2008 [9]; Lantolf and Poehner, 2004 [4]; Lantolf and Thorne, 2006 [5]).

One of the watershed observations that deserves close attention to unravel the effect of Dynamic Assessment on EFL writing was carried out by Xiaoxiao and Yan (2010) [10]. The inseparability of instruction and assessment as a distinguishing feature of Dynamic Assessment has been in vogue recently. In spite of the widespread contributions of Dynamic Assessment to L₂ pedagogy, only meager resources give some insights into EFL writing (e.g. Donato, 2000 [3]; Lantolf and Thorne, 2006 [5]; Murphy and Maree, 2006 [7]). In sum, the ultimate objective of Dynamic Assessment in its all multitudinous forms can be finally achieved through interaction and mediation to facilitate the development of writers. What has been overlooked during the traditional assessment, that is, individuals' emotions and desires are being partially retained with the affective embodiment of Vygotsky's Zone of Proximal Development (Levykh, 2008) [6]. Lastly, there is no point in muddling on in the same old tradition of assessment for ever. On the other hand, although there is a robust research literature on DA in psychology and

general education, the approach is relatively unknown in second language (L₂) writing.

Purpose of the Study

To date, few studies are in progress in examining L₂ writing from a DA perspective. Therefore, studies of DA's implications for problems particular to the development of L₂ writing are only beginning, and there is a lot of room to work on it. Thus, this study was designed to investigate the following research questions:

1. Is there any relationship among students' self scores, teacher scores, and self-assessment in pre-DA phase of this study?
2. Is there any relationship among students' self scores, teacher scores, and self-assessment in post-DA phase of this study?
3. Is there any relationship between students' pre-DA self scores, teacher's scores, and self-assessment and those of their post-DA?

Methodology

Participants

This study was designed to gather evidence on the nature and effects of dynamic assessment on self-assessment and L₂ writing. It was conducted on 22 freshman-engineering students at Sharif University of Technology with the researcher. The participants consisted of an almost equal number of males and females. The course was General English, a three- credit course that met for three classroom hours per week in a fourteen-week semester.

Instruments

Two instruments have been used in this study, namely, seven paragraphs in different topics (as pre-test, treatment, and post-test) and a self-assessment questionnaire which was adapted from Bandura's (1995) [2] Self-Efficacy Scales (with reliability of .86) and then modified. The questionnaire consisted of 11 items tapping students' judgments of their confidence that they possessed regarding composition, grammar, usage, and mechanical skills appropriate to their academic writing skill. The students were asked to respond on a likert format consisting of 5 points ranging from 1 (no confidence at all) to 5 (completely confident).

Data Collection Procedure

Pre-DA/Pre-test Phase: At the very first session, students were asked to write a paragraph in half an hour in English (as the pre-test) about why they chose this university and this major. Dictionaries were not allowed, so that students would need to rely for word meaning on strategies such as prediction and hypothesis. Then they scored (on the 5-point scale ranging from 1 ("poor") to 5 ("excellent") their own papers for self-rating and filled out a self-assessment questionnaire. Students completed each scale as items were read aloud by the researcher. The instrument took approximately 5 minutes to complete. It was used to collect information about the starting points of each learner. Indeed, it was a record of their skills at the commencement of the course.

Treatment/Mediation Phase: In the session that followed, their papers were corrected along with feedback and the following errors in the class on the board: (1) language errors, such as word choice, verb tense, verb form, word form,

articles, singular-plural, pronouns, fragments, punctuation, spelling, sentence structure, idioms, and subject-verb agreement, and (2) writing organization. The researcher scored the papers and compared the students' scores with the researcher's scores in order to raise consciousness among the students. The chart of error categories was selected by the researcher in consultation with error analysis specialists. Then the researcher assigned another topic based on assigned readings of their textbook. Their textbook was *General English for Science and Engineering Students* (Alemi et al., 2009) [1]. For seven sessions the students wrote seven draft essays on varied topics and submitted their assignments in courseware (an academic forum at the university site). Each time the researcher and the students scored the papers.

Post-DA/post test Phase: At the last session, students were asked to write the last topic (what they had learned from their English class during this semester) as the post-test in the class, scored them for self-rating, and filled out a self-assessment questionnaire again. The researcher also scored the last papers.

Data Analysis

To investigate the reliability of the self-assessment questionnaire, the responses were analyzed through Cronbach's Alpha. The reliability index was .81 in pretest and .83 in posttest, respectively. The researcher also analyzed descriptive statistics of the self-assessment questionnaire, students' self-rating, and teacher rating. Then, she examined the relationship between students' self-rating, teacher scores, and students' self-assessment in the pretest and posttest. Next, she examined the relationship between students' self-rating in the pretest and posttest, students' self-assessment in the pretest and post test, all through correlation. Finally, the researcher used t-test between teacher rating in the pretest phase and that in the posttest phase, students' pre self-rating scores and post self scores to see if there are any significant differences.

Results and Discussions

The aims of this paper were to investigate the following relationships before and after a teacher-scaffolded self-assessment course in EFL writing: (1) the relationship between EFL learners' general self-assessment of their writing ability and their task-specific assessment of their writing (self-rating); (2) the relationship between EFL learners' general self-assessment of their writing ability and the teacher's task-specific assessment of their writing (teacher rating); and (3) the relationships between EFL learners' task-specific assessment of their writing (self-rating) and the teacher's task-specific assessment of their writings (teacher rating). First, the results of the pre-course phase are reported. Next come the findings related to the relationships among the three variables subsequent to the writing course in which students were asked to self-rate their writing papers, a sample of which was discussed every session to measure the soundness of the student rating against the teacher rating in order to raise students' consciousness about the accuracy of their self-rating through dialogic interaction and the scaffolding provided by the teacher in the process of dynamic assessment.

Pre-course Phase: The Relationship among Self-assessment, Self-rating, and Teacher Rating

To investigate the relationship among the variables, at first, descriptive statistics on self-assessment, students' self-rating, and teacher rating are given. As Table 1 shows, on a 5-point Likert scale, students' general self-assessment of their writing ability was an average of 3.27.

Table 1: Descriptive statistics for pre-course self-assessment

	Mean	N	Std. Deviation	Std. Error Mean
Pre-course Self-assessment	3.276860	22	.3972329	.0846903

An examination of students' self-assessment of the individual components making up their writing ability (Table 2) shows that they evaluated themselves as having the highest ability in Item 4 (M=3.68), as described below:

Item 4: I can write simple sentences with good grammar

By contrast, they ranked themselves as lowest in Item 7 (M=2.18), which was concerned with the use of idioms and expressions:

Item 7: I can correctly use idioms and expressions in my composition.

Table 2: Descriptive Statistics for pre-course self-assessment items

	N	Mean	Std.
Pre-course Self-assessment Items	Q1	22	3.59
	Q2	22	2.86
	Q3	22	3.59
	Q4	22	3.68
	Q5	22	3.64
	Q6	22	3.23
	Q7	22	2.18
	Q8	22	3.55
	Q9	22	3.55
	Q10	22	2.95
	Q11	22	3.23

Among the high-ability items, Item 4 falls within the Language Section of the self-assessment questionnaire. The selection of this item indicates that EAP students exhibit a high confidence in writing simple sentences in English, particularly due to the fact that writing individual sentences with correct grammar is one of the focal points in high school English education and in many of the EAP English courses for university students. By contrast, the lowest rank was allocated to correctly use idioms and expressions in compositions. This self-assessment is compatible with the low ability of EAP students in particular and EGP students in general in using idioms and expressions due to the meager attention allocated to them in their previous English education in high school and EAP courses in universities which are primarily focused on grammatical knowledge and non-idiomatic lexical knowledge.

The second variable in the pre-course phase was students' task-specific rating of their writing (self-rating). While, they evaluated their general writing ability without doing any writing task in self-assessment part of this phase, they were asked to write a composition and then rate it on a 5-point scale of 1 (very poor) to 5 (excellent) in the self-rating part. As shown in Table 3, they assigned the average of 3.36 to their writing.

Table 3: Descriptive statistics for pre-course self-rating

	Mean	N	Std. Deviation	Std. Error Mean
Pre-course Self-rating	3.36	22	1.177	.251

This rating is high by nature, showing that students considered their actual writing performance to be very good. With the average self-rating of 3.36, they showed consistency in their evaluation of their writing ability in that their evaluation of their general writing ability prior to writing task performance amounted to nearly the same average, i.e. 3.27. The similarity of the averages may lend support to the fact that task-specific writing with the demands it imposes on the students cannot change their general perception of their writing ability expressed through self-assessment. However, the mere similarity of the averages may be misleading because the similarity in the whole group may not prove to hold true as far as the correlation between the two sets of scores is concerned. In other words, despite differences between individuals' self-assessment and their task-specific self-rating, the average of the whole group in the two variables may be similar. As revealed later, the correlational study has the potential to show whether the two sets of scores are interrelated.

The comparison of students' self-assessment and self-rating with teacher rating will function to evaluate the accuracy of the first two against teacher rating. Table 4 depicts the finding related to teacher rating. As seen in the table, the average score the teacher assigned to students' writing task in the pre-course phase was 2.36. This comparison shows that the teacher's evaluation of students' writing performance is much lower than students' self-assessment and self-rating. It follows that EFL learners over evaluate their language skills, which is particularly a striking feature of students at the low-proficiency level.

Table 4: Descriptive statistics of pre-course teacher rating

	Mean	N	Std. Deviation	Std. Error Mean
Pre-course Teacher Rating	2.36	22	.790	.168

Despite these descriptive findings, the answer to the first research question requires the calculation of the correlations among the three variables of self-assessment, self-rating, and teacher rating. Table 5 shows the correlations among the three variables. The correlation between students' self-assessment and their task-specific self-rating was .23, which is not significant at the $p < .05$. The correlation between students' self-assessment and teacher rating was insignificant too, falling at .31. On the contrary, a significant correlation was found between students' self-rating and teacher rating (.62**).

Table 5: Correlations among pre-course self-assessment, self-rating, and teacher rating

		Pre-course Teacher Rating	Pre-course Self-assessment
Pre-course Self-rating	Pearson Correlation	.620**	.237
	Sig. (2-tailed)	.002	.287
	N	22	22
Pre-course Self-assessment	Pearson Correlation	.312	.276
	Sig. (2-tailed)	.157	.214
	N	22	22

There are reasons for the lack of significant correlation between students' self-assessment and their task-specific self-rating on the one hand and self-assessment and teacher rating on the other. First, students in the Iranian ELT context have a vague understanding of their writing ability as writing does not feature high in their language curriculum. It follows that when it comes to rating their actual writing papers, their rating behavior may change because they encounter the challenges of writing a well-organized paragraph and gain the opportunity to read and rate their actual writing product. Second, as the ELT program in Iran does not give any weight to self-assessment, ELT students neither have a proper awareness of the nature of self-rating and self-assessment scales nor are able to close up their general self-assessment and their task-based rating.

By contrast, the correlation between students' self-rating and teacher rating was found to be statistically significant ($r = .62$, $p < .05$). This finding may be explained on many grounds. First of all, when it comes to rating a writing task they have performed, EFL learners are on a better footing to evaluate their writing ability as evidently manifested in their piece of writing. As a result, their rating gets closer to teacher rating. Another reason may be that while self-assessment is an off-task assessment measure, both students' self-rating and teacher rating are on-task or task-based assessment of writing ability as realized in compositions. Consequently task-based ratings are more likely to correlate with each other.

Post-course Phase: The Relationship among Self-assessment, Self-rating, and Teacher Rating

The second purpose of this study was to explore the relationship among students' self-assessment, students' self-rating, and teacher rating. To this end, first there will be a report on the averages found as to these three ratings. Next will come the correlational indexes showing the relationships among these three variables in the post-course phase of the study, i.e. subsequent to the participants' writing course in which they were supposed to self-rating, there was a dynamic assessment of their writing, and there was a dialogic interaction between the teacher and the students on the accuracy of their self-ratings against teacher ratings.

Table 6: Descriptive statistics of post-course self-assessment

	Mean	N	Std. Deviation	Std. Error Mean
Post-course Self-assessment	3.314050	22	.3467137	.0846903

As Table 6 shows, the self-assessment average at the end of DA-based writing course reached the amount of 3.31. Compared with the average of 3.27 for pre-course self-assessment, there was not a big change in students' self-assessment of their writing ability despite the writing course they had received.

An examination of students' self-assessment of the individual components of their writing ability, as revealed through the self-assessment questionnaire, shows that the students highest evaluation of their writing ability fall on Item 4 (M=3.64), evaluating their ability to write simple sentences with good grammar. In line with the discussion of self-assessment results in the pre-self-assessment phase, the participants considered themselves to have the highest ability in writing simple sentences with good grammar due to the grammar- and sentence-oriented English language teaching and testing in Iran.

Table 7: Descriptive Statistics for post-course self-assessment items

	A1	22	3.59	.666
Post-course Self-assessment Items	A2	22	3.14	.774
	A3	22	3.55	.739
	A4	22	3.64	.790
	A5	22	3.59	.666
	A6	22	3.36	.727
	A7	22	2.36	.790
	A8	22	3.23	.813
	A9	22	3.59	.666
	A10	22	3.23	.813
	A11	22	3.18	.501

Whereas Item 4 received the highest average, Item 7 was ranked as the lowest. With the average of 2.36, it relates to the ability to correctly use idioms and expressions in compositions. Ranking as the lowest in the pre-course phase of self-assessment as well, the ability to draw on idioms to enrich writing tends to be the highest demand on the part of EFL students. The reasons are multiple, including the very low weight given to idiomatic knowledge in high school English education and English instruction and testing in University general English courses as well as linguistic variation in the idiomatic expression of concepts in Persian and English.

Students' self-rating of their writing ability reached an average of 3.45 (Table 8). This rate is higher than that of their self-assessment in the post-course phase. Compared with the average of 2.77 for the teacher score of their writing (Table 9), the self-rating average shows a much higher rate. The lower average score given to the writing papers by the

teacher indicates that EFL students are less aware of complicated nature of writing ability and hence tend to overrate their writing performance.

Table 8: Descriptive statistics of post-course self-rating

	Mean	N	Std. Deviation	Std. Error Mean
Post-course Self-rating	3.45	22	.800	.171

Table 9: Descriptive statistics of pre-course teacher rating

	Mean	N	Std. Deviation	Std. Error Mean
Post-course Teacher Rating	2.77	22	.973	.207

To obtain a deeper understanding of the relationships among the three variables in the post-course phase, a correlational analysis was employed. The results of correlational analysis (Table 10) show the correlational index of .540 between students' post-test self-assessment and their post-test self-rating of their writing ability. The index is statistically significant at $p < .05$. This shows that, after taking the DA-based writing course, the participants' general assessment of their writing ability went hand in hand with their task-specific rating of their writing ability. This is significant because it indicates that the same criteria were involved in general assessment and the task-based assessment of their writing ability. Irrespective of the match or mismatch between these two ratings and teacher rating, this finding lends support to the effect of dynamic assessment and continuous self-assessment on the greater going-togetherness of students' assessment ability in both off-task and on-task situations as a result of the insights they gained into their writing ability throughout the DA-based course in order to apply a more informed, systematic measure to evaluate their writing ability.

Table 10: Correlations among post-course self-assessment, self-rating, and teacher rating

		Pre-course Teacher Rating	Pre-course Self-assessment
Post-course Self-rating	Pearson	.384	.540**
	Sig. (2-tailed)	.078	.010
	N	22	22
Post-course Self-assessment	Pearson		.472*
	Sig. (2-tailed)		.027
	N		22

In addition, through correlational analysis, the significant correlational index of .472 was obtained between students' self-assessment and the teacher rating of their writing ability. The match between self-assessment and teacher rating in the post-course phase of the study is evidence to the effect of the DA-based writing course on students' awareness of their

writing ability and hence the closer approximation of their assessment to teacher rating.

Contrary to the significant correlations found between self-assessment on the one hand and the two variables of students' self-rating and teacher rating, the correlation between students' self-rating and teacher rating was not significant (.384). A number of reasons may account for this unexpectedly low correlation despite the insights the students were expected to gain through the course. The first reason may be that, unlike self-assessment as a questionnaire consisting of different clear items about the evaluation criteria, students' self-rating was only based on a Likert scale about how good or poor the composition was. As self-rating, unlike self-assessment, was based on Un-informed assessment, it had a lower correlation with teacher rating. The second reason may be the conscious application of the criteria for rating. Whereas the students might have failed to draw on rating criteria instructed to them through the DA-based course, the teacher's awareness of writing skills of and assessment put her in a far more advantageous position to activate the criteria while assigning different rates to compositions.

Pre-self-assessment and Post-self-assessment Phases: The Impact of Dynamic Assessment on Self-assessment and Self-rating

The main purpose of the study was to measure the effects of a DA-based writing course on the accuracy of EFL students' self-assessment and self-rating and the match of these two with teacher rating. To this end, a number of t-tests were applied to the scores for these three variables in the pre-course and pos-course as shown in Table 11. The application of the matched t-test ($t=-.44$, $df=21$) shows that the obtained value of t did not exceed the critical value and hence there was not significant difference between the means score of pre-course ($M=3.27$) and that of the post-course ($M=3.31$) for students' self-assessment of their writing ability.

The very small differences between the two means, resulting in the t value not exceeding the F critical value, can be used to argue for and against the beneficial effect of dynamic assessment. The argument for dynamic assessment is that the participants came to exhibit a better evaluation of their writing ability. While they over assessed their ability in the pre-course phase ($M=3.36$), which was a far cry from teachers rating ($M=2.36$), they became more accurate in their self-assessment at the end of the course in two ways: by assessing their post-course ability as 3.31 which was close to the pre-course average of 3.27 despite their improvement in writing throughout the course, and by decreasing the pre-test mismatch between self-assessment (3.27) and teacher rating (2.36) to a smaller difference in rating from 3.31 for self-assessment to 2.77 for teacher rating. The second argument for efficacy of the DA-based course relates to the decrease in the difference between self-assessment and self-rating as a result of the course. Before the start of the course, the participants did not evaluate their ability for writing in an informed way because of their insufficient consciousness about the evaluation criteria for good writing. This caused not only overrating, as discussed above, but also haphazard variation in assessment depending on variation in the context for assessment. The consequence of this contextual variation was the mismatch between off-task general self-assessment and on-task self-rating and hence non-significant correlation between the two variables in the pre-test, i.e. before the DA-based instructional course ($r=.23$). The situation, however, changed at the end of the course in that the participants gained insight into assessment criteria through regular self-assessment and the dialogic discussion between the teacher and the students on the accuracy of their self-assessment. This leads to an overwhelming increase in the correspondence between self-assessment and self-rating at the

end of the course, with the significant correlation between the two variables ($r=.54$, $p<.01$) as previously showed in table 2.

Table 11: t-tests for self-assessment, self-rating and teacher rating

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of				
				Lower	Upper			
Pre-course and Post-course Self-assessment	-.0371901	.3894103	.0830226	-.2098449	.1354648	-.448	21	.659

The argument against for the sufficient impact of the DA-based writing course on the participants' writing awareness emerges from the data on the low correlation between self-rating and teacher rating. This indicates that the DA-based course fell short of providing the participants with enough insight into writing assessment criteria to bring their on-task rating of their rating significantly closer to teacher rating.

Table 12: t-test for pre-teacher score and post-teacher score

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence				
				Lower	Upper			
PRETEACHER SCORE - POSTTEACHER SCORE	-.409	.796	.170	-.762	-.056	2.409	21	.025

Table 13: *t*- test for pre-self score and post- self score

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
PRESELF SCORE - POSTSELFSCORE	-.174	.650	.136	-.455	.107	1.283	22	.213

Conclusion

One purpose of this study was to explore EAP students' evaluation of their writing ability in terms of their general perception of their writing ability in general and based on a specific writing task before and after taking a writing course characterized by dynamic assessment. The findings revealed that the consistent self-rating and teacher-scaffolded dynamic assessment provided EAP students with insights into their writing ability, resulting in their more accurate assessment of their writing ability. This has many implications. First, the development of L₂ writing skills can benefit from L₂ students awareness of the criteria involved in the evaluation of writing. The second implication is that L₂ students, particularly low-level students, need to be involved in self-rating in order to evaluate their own writing ability as a metacognitive strategy and to eventually move toward autonomous learning. Third, as low-level L₂ students tend to overrate their writing ability, a dynamic-based assessment course can help them become more accurate in self-assessment.

The second purpose of the study was to find the degree of correspondence between teacher rating and student self-rating of writing ability. The results provided evidence as to the effect of dynamic assessment on the lowering of mismatch between the two variables. As students became more familiar with the criteria the teacher applied to the evaluation of their writing and received teacher scaffolding, their self-rating turned out to be more informed practice. This suggests that dynamic assessment and teacher scaffolding of the self-rating process can improve the match between teacher rating and students' self-rating.

References

[1] Alemi, M., Salehi, M., Khosravizadeh, P., Hatf, A. (2009). *General english for science & engineering students*. Tehran: Farhange Alem publication.

[2] Bandura, A. (1995). *Self-efficacy in changing societies*. New York: Cambridge University Press.

[3] Donato, R. (2000). Sociocultural contributions to understanding the foreign and second language classroom. In J. P. Lantolf (ed.), *Sociocultural Theory and Second Language Learning*. Oxford: Oxford University Press.

[4] Lantolf, J.P. and Poehner, M.E. (2004). *Dynamic assessment in the language classroom*. CALPER, PA: The Pennsylvania State University.

[5] Lantolf, J.P., and Thorne, S.L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.

[6] Levykh, M.G. (2008). The affective establishment and maintenance of Vygotsky's Zone of proximal development. *Educational Theory*, 58 (1), 83-101.

[7] Murphy, R. and Maree, D.J.F. (2006). A review of South African research in the field of dynamic Assessment. *South African Journal of Psychology*, 36 (1), 168-191.

[8] Poehner, M.E. & J.P. Lantolf. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, Vol. 9 (3): 1-33.

[9] Poehner, M. E. (2008). *Dynamic Assessment: A Vygotskian Approach to Understanding and Promoting L₂ Development*. Chicago: Springer.

[10] Xiaoxiao, L. and Yan, L. (2010). A case study of dynamic assessment in EFL process writing. *Chinese Journal of Applied Linguistics*, 33 (1), 24-40.