



The Difference in Formative Writing Assessment Procedures among Novice and Experienced Teachers

Seyedeh Azam Hosseini

Department of English Language, Kharazmi University, Tehran, Iran

Received 02 February 2019 Accepted 10 June 2019

Abstract

The present research investigates the associations among teacher assessments for foreign language (L2) writing and relevant teaching experiences of L2 writing. This study also seeks to address teachers' employment of formative evaluations for lessons of General Studies amidst teaching practices, teacher formative evaluation perceptions upon carrying out formative evaluations for General Studies, the prevalent approaches of formative evaluations, the accordance of evaluation methods with learning tasks, the intervention of teachers upon receiving feedback from their students, issues in classes and student feedback. Thirty English teachers from a language institute in Iran assessed 10 essays by students and validated their marks using three qualitative statements. Findings of this study indicate the lower marks of the writing teachers with the most experience compared to the teachers of less experience for four of the ten total essays. Upon conducting an analysis for the qualitative statements pertaining to the four essays, it was shown that experienced teachers submitted more negative comments on issues such as general language, ideas, language fluency, and general organization. Essentially, despite various experiences in instructing composition, such groups of teachers may be similar or different in terms of other variables that may contribute to divergences and similarities for their ratings.

Keywords: *Formative assessment (FA), Writing Evaluation procedures, Novice, Experienced*

Introduction

Formative evaluations refer to continuous evaluations, summarizations, examinations, and analysis which provide teacher instruction and pupil feedback every day (Fisher & Frey 2007). Conventionally, evaluations are used to determine the extent of learning achieved by students based on a defined period (Stiggins, 2007). This is what he refers to as “assessment of learning” which is used in this study to determine if students are measuring up to the defined standards by

the state, region or the teacher. Formative evaluations are a learning constituent; thus they are not marked as summative evaluations (e.g. end of unit exams or quarterly exams). Hence, they function as student practice similar to purposeful homework (Chappuis & Chappuis, 2007,2008). These type of assignments are to aid in understanding and in guiding teachers in their decisions on future instructions. They also present pupils with feedback thus improving performances. Assessment standards are made concrete by providing teachers within a clear sense of the kinds of performances students need to demonstrate to show that they have met the standards.

The research questions provided below are used to guide the research:

- 1) What different FA writing procedures are used by novice and expert practitioners?
- 2) How do different FA writing procedures contribute to their professional practice development?

Despite the fact that instructor origination on evaluation is commonly characterized as the thoughts and mentalities that educators have toward what evaluation is and what it is really pursuing (Brown & Gao, 2015), unique purposes that appraisal serves may make confusion among instructors. At the end of the day, educators' apparent origination of appraisal in language teaching and its motivations might be characterized contrastingly by various analysts (Biggs, 1998). Torrance and Pryor (2001) depict the reasons for appraisal as "perception of procedure and items", "giving criticism and judgment" and "questioning" (p. 624). In another model, the points of evaluation are portrayed as "choice", "expanding the educators' adequacy" and "expanding the learners' potentialities" (Heaton, 1988, p.136). As indicated by Brown (2004), these reasons for the most part centre around the following criteria: "improvement of educating and picking up", "making students responsible for learning" and "accountability of schools and educators". Other basic meanings of these objects is recorded as (a) modification of the viability of the learning procedure and teacher training, (b) to guarantee responsibility of the schools' execution, (c) to give data about the learning procedure of the learners to the individuals who are concerned, for example, parents or the students themselves, and (d) to modify the principles of national instruction framework (Alhareth & Ibtisam, 2014).

There is broad evidence which suggests that there is a vital opportunity to get better in evaluation (George & Cowan, 1999). Most recent assessments declare that appraisal is regularly a frail region when contrasted with different parts of the educational modules and furthermore criticism is impressively moderate which neglects to give agreeable help to learners (Yorke, 2005). Preparing reasonable feedback can be complex and troublesome; numerous components are required to be considered, including the number of students, time and assets and course goals (Zou, 2008). While there is an expanding interest for the significance of having a differing evaluation framework, there is still a lot of accentuation on tests and standard tutor checked papers and reports (Hornby, 2005). There is a threat that "latent, exhausted students offer back to educators what they have just been given in a useless evaluation grubbing route unessential to their future lives" (Gibbs, 1995, p. 2).

In line with Zou (2008, p 83), evaluation practices are changing for six important motives, even though the ones are likely not to be particular to this area:

- A. The heightened reputation of its importance
- B. more need for accomplishing established skills which incorporates interaction and cooperation
- C. increase in capability for plagiarism
- D. need for time and fee, strong strategies of assessing huge classes
- E. up to date technology
- F. converting nature of students

A current-day look at in Scotland observed that changes to assessment are not well matched with the converting environment in higher education (Hornby, 2005). The boom in student numbers has stretched the unit of beneficial resource, resulting in unintended outcomes; there isn't enough formative assessment; feedback is truly too often slow-paced and not huge; studying effects are regularly assessed several times and now not using a purpose; there may be little correlation between credit score factors and student and personnel workload; mechanisms are horrible for coordination at some stage in modules; and bunching of assessments consequences in issues.

Experience or Years of Teaching Writing

In the experiential learning concept, an emphasis is placed on the significance of experience in learning. Kolb, Boyatzis, and Mainemelis (2001) stated if experienced teachers aim to be influential, then they require competency in four various areas: active experimentation (AE), abstract conceptualization (AC), reflective observations (RO), and concrete experience (CE). Tsui (2003) conducted a comparison between expert and novice teachers in terms of proactive and interactive instruction. According to this comparison, concerning the proactive phase, the experienced teachers exhibited different characteristics compared to novice teachers in four principle traits:

- 1- Concerning the preparation procedure, the experienced teachers exhibit higher independence levels whereas novice teachers plan according to models and rules.
- 2- Experienced teachers exhibit highly efficient planning skills compared to novice teachers despite using less time.
- 3- Experienced teachers exhibit higher flexibility when planning since they vary their plans based on context.
- 4- Experienced teachers employ an integrated and rich basis of knowledge. In regard to the interactive phase, experienced teachers exhibit differences to novice teachers in three facets. The most prominent facet is information processing efficiency during the classes. Experienced teachers are competent in transmitting information whilst they convey data. The other facet is experienced teachers are competent in selecting information for processing. The final facet is that experienced teachers take the students requirements into account and respond to events accordingly in class. Time and experience are other influential aspects in developing pedagogical knowledge. The domain of particular knowledge must be adhered to by experts via significant contextual experience. Prior to taking part in classes, experts should:

- 5- Have a thorough mastery over the content, and design a number of activities for teaching content (Berliner, 2004).

Amidst the development phase, pragmatic knowledge is built up and gradually achieved amidst the career of a teacher (Van Driel et al, 1998). Pragmatic knowledge is based on action and commonly achieved with no direct assistance. Moreover, pragmatic knowledge is context-bound and personal whilst presenting the necessary skills for teachers in adhering to relevant contexts of teaching.

Shi (2001) cited that there are no prior studies which investigate the experience of teachers in L2 writing in terms of a variable which may find variances in marking student compositions. He implied that the marks given by the common level of experience pertaining to English teachers in ESL may influence the assessment of writings. Research conducted for comparison purposes by Hughes and Lascaratou (1982) in terms of error gravity criteria stated that experienced teachers have the propensity to utilize intelligibility and grammar rules compared to novice teachers. Another research conducted by Cummings (1990) on comparing the decision making activities of experienced and novice teachers in marking ESL writing suggested that experienced teachers utilized efficient strategies as well as an extensive scope of knowledge resources for reading and marking student writings. As an example, teachers of greater experience frequently adhered to specific traits e.g. vital criterion, primary ideas, topic and English sentence structure command development. Contrarily, novice teachers were mainly focused on language characteristics analysis or perceiving the conveyed ideas from the textbooks. To summarize, the experienced teachers exhibited lower marks to different facets of sample composition compared to novice teachers.

In addition to the studies that investigated the influence of the level of teachers' experience levels, numerous studies indicated the effect of experience of raters in terms of refinement and EFL writers on assessing L2 texts. As an example, Hamp-Lyons (1989) suggested native English speakers might exhibit negative or positive bias in regard to ESL writing according to their experience levels pertaining to the language and culture of students. Identically, Land and Whitley (1989) implied that bilingual or multilingual readers may place value on various writing schemes. Scholars have also observed the fact that faculty members who have been exposed to ESL pupils might show more tolerance towards language mistakes (Vann, Lorenz & Meye, 1991) or exhibit higher leniency on ESL holistic assessment of essays (Song & Caruso, 1996). All in all, the conclusions of these studies imply the raters' judgments might be affected by the exposure level to ESL students' culture and language.

Teacher Effectiveness

Numerous surveys were conducted on self-efficacy and the effectiveness of teachers (Evers et al, 2002; Fives et al, 2005; Rotter, 1990). Based on a study by Rotter (1990), the efficacy of teachers is the level that teachers think they could preserve reinforcements of relevant natural processes within themselves i.e. reinforcement of internal control or the environment i.e. reinforcement of external control. Student performance and motivation are vital reinforcements

which pertain to the behaviors of teaching. Rotter states that the aspects that are controlled by teachers are more influential in teaching results compared to aspects of the students and the environment. Bandura (1977) defines teacher effectiveness as a character stemming from self-efficacy where individuals form a belief concerning their capability to conduct tasks with a gifted degree of accomplishment. Such self-efficacy future-aligned and impacts emotions and thought patterns.

Based on this concept, efficacy is affected rather easily when learning, thus the initial teaching years may be vital to the overall teacher efficacy development. The sense of efficacy of teachers is relevant to pupil performances e.g. accomplishments, inspirations, sense of effectiveness and the behavior of teachers during classes. A solid efficacy sense enables teachers to use novel theories whilst motivating them to test new approaches according to student requirements as well as causing higher provision levels and governance with less critique of student mistakes (Tschannen-Moran et al, 1998). Two parameters influence efficacy. The foremost ingredient is innate teaching effectiveness that is defined as an individual's capability notions as an instructor.

The other parameter is arguably referred to as general teaching efficacy. Various scholars such as Riggs and Enochs (1990) refer to this parameter as the outcome expectancy where individuals evaluate the outcomes of their conducted tasks which were meant to be accomplished. The efficacy perceived by novice teachers is linked to commitments to instruction and stress in addition to being satisfied with their received training and support. Beginner teachers with a significant teacher efficacy level are more satisfied when teaching, have a more positive outlook on teaching and exhibit fewer levels of stress. Novice teachers of higher efficacy give higher ratings to their preparations and consider teaching difficulty as more insignificant compared to teachers of less efficacy. Experienced teachers have more constant efficacy beliefs. Evidently, the input amidst primary training phases exhibits a differing influence compared to the inputs when the teachers are practicing in the field (Tschannen-Moran et al, 1998).

Methodology

A total of 30 teachers-raters in Tehran, Iran contributed to this research. Since the majority of English writing schemes in Iranian institutions are presented collectively via expatriate and local tutors, several volunteers were initiated from each group. They were mostly in-service teacher-trainees of a language institute. Every teacher who was participating in this study filled in a questionnaire which asked for demographic data e.g. native language, age and general experience of teaching English and experience in teaching English as a foreign language i.e. EFL. Among such variables, a revision was an autonomous variable. The following EFL teaching experiences were included.

Table 1

Teacher profile

Years of teaching experience		
5- 8 years	9- 12 years	13-16 years
9teachers	17teachers	4teachers

The years of teaching experience of EFL writing in terms of their L1 is summarized in Table 1. The EFL teaching experience in terms of years was gathered into three categories i.e. 5 to 8, 9 to 12 and 13 to 16 years on the basis of data distribution. This categorization can aid in determining the ways in which teachers are different in assessing student essays according to the various experience stages of their careers. According to Table 1, among the thirty participants in the 5 to 8 teaching experience years' group, nine teachers were between the age range of 27 to 30 and only one of them was 44, whereas seventeen of the teacher participants had taught English writing for 9 to 12 years and four teachers had 13 to 16 years of experience.

Procedures

A random selection of ten essays from 20 writing assignments in class was conducted. Each eighth essay was chosen among the total set in no defined order. The writing task was administered by three teachers in writing classes that lasted 50 minutes for advanced English students in a prominent language institute. Amidst the data collection time, pupils practiced argumentative essay writing as a section of their writing curriculum. The three teachers suggested the writing prompt to adhere to the project pertaining to relevant teaching procedures: "Nowadays, with the popularity of social media, particularly the internet, individuals have access to news on a daily basis in a more convenient manner. There are some people who dismiss the benefits of television, stating that the internet should take its place. How much of this statement do you agree or disagree too? Provide support for arguments"

All ten essays were given to the contributing teachers to evaluation via a 10-point scale whilst providing three reasons or comments to rationalize their marks. There were no assessment indices provided for the 10- point scale, thus the teachers could only utilize their personal experience to provide qualitative and quantitative assessments. The teachers participating in this survey had to abide by the instructions below when evaluating the essays:

The aim of this study is to determine the ways English teachers rate student essays. Please use a 10- point scale to read and mark the provided essays (10 is the highest mark) and comment in the order of significance, three traits or reasons for every essay which you regard as the most influential in marking that essay (priority of importance given to the first reason).

Coding of Qualitative Comments

Teachers' comments were subjected to comparison using a developed coding framework. According to primary observations, the comments of teachers mostly consisted of an adjective e.g. good, strong, poor, weak etc. which exhibited the positive or negative nature of the comment as well as a content expression showing a specific or general textual trait e.g. general quality, organization, content, length and language which teachers focused on. According to such keywords, the comments were subjected to coding as negative or positive words in five main

groups: comments pertaining to essay length, language, organization, content and overall quality. The language, organization and content groups were subjected to extended evaluations to determine the subgroups of various particular or general comments. Such subgroups pertaining to particular comments were arguments or ideas under the group of content; transition and paragraph under organization; and finally, fluency, precision, and intelligibility under language. Hence, twelve groups were produced.

To check inter-coder reliability, two of the experienced teachers autonomously accredited codes to their comments and obtained a 95% agreement. According to entire data set coding, an overall number of comments were identified which included both negative and positive comments on the basis of the twelve produced categories that showed how teachers rated the quality of essays. Less than three comments were presented by a few researchers for a number of essays.

Data Analysis

An analysis was conducted on the qualitative judgments and holistic marks given to the essays to discover the extent at which the number of years of teaching EFL writing skills influenced teacher assessments. In order to conduct a comparison, initially, tests of reliability were run to ascertain the degree of teachers' agreement on holistic marks given to the essays, defined by the number of years of teaching EFL writing skills. In the next stage, post hoc Tukey and ANOVA tests were conducted to evaluate the variances concerning the mean scores submitted for every essay via different categories of teachers. Post hoc Tukey and ANOVA tests were run for the essays that were different in terms of holistic scores for the purpose of comparing the mean frequencies of 24 comments (12 positive category comments and 12 negative category comments) to ascertain if variances in regard to the qualitative judgments of teachers may be the reason for the differences in terms of given scores.

Results and Discussion

This section includes the results and findings of the nature of teachers and the ways that teachers of different experience levels of teaching EFL writing provided various qualitative and quantitative judgments in their process of assessing the same student essays. The credibility of the teacher groups was determined based on differing teaching EFL/ESL writing years that was derived according to interclass correlation coefficients. Questionnaire findings in this survey were based on the Shi (2001) which is affirmed to be true.

Table 2

Teacher-raters reliability coefficient compared with differing teaching experience years (Alpha)

5-8 years (N = 9)	9-12 years (N = 17)	13-16 years (N=4)
.68	.84	.75

Table 2 reports the correlations and displays that reliability coefficients were within the 0.68 to 0.84 range implying various agreement levels for every group as a whole corresponding to ten

essays. The 5 to 8 years of experience group in terms of EFL writing attained the lowest reliability i.e. Alpha=0.68. Evidently, the teachers of the 9 to 12 years of experience group showed the most consistency i.e. Alpha=0.84. The group of highest experience (13 to 16 years) showed more consistency compared to the 5 to 8 years group but less consistency compared to the teachers of 9 to 12 years of experience. Apparently, experience in EFL teaching of writing skills (9 to 12 years) aided teachers in more consistency when assessing student essays. Unexpectedly, the teachers of most experience (13 to 16 years) did not gain the highest level of reliability.

To answer the first research question in this study, survey questionnaires were distributed among the novice and experienced teachers of an EFL institute. The purpose of the questionnaire is to determine the teachers' feedback toward FA (Formative Assessment) and likewise to test the first hypothesis that experienced teachers expose their students to write feedback or peer feedback and use formative assessment in order to cater to the dissimilar ways that they take in information.

Table 3

Attitudes toward writing

S	Y	N	
15	12	3	Are you keen on writing?
6	24	0	Generally, are you confident in your abilities in finding appropriate words, perceptions, and ideas?
18	12	0	Would you consider yourself being a writer?

Write Y, N, and S as denotations of Yes, No and Sometimes. In the case that the answer is unknown, which can happen often at the beginning of the course, apply a question mark.

Table 2 indicates that half of the teachers, enjoy writing, most of the teachers trust them as ways to come up with good words and ideas. More than fifteen respondents think of themselves as writers.

Table 3

Generating

15	15	0	On a topic that interests you, are you able to create many words promptly and freely-not hesitate?
18	2	0	When a topic interests you, are you able to think of insights and ideas which you could not come up with before?
18	12	0	When a topic does not interest you (maybe a topic that has been assigned), are you able to create numerous words in a prompt and free manner-without hesitation?
24	3	3	When a topic interests you, are you able to think of insights and ideas which you could not come up with before?
9	15	3	In the case that you do not know your opinion about a topic, are you able to think and write enough to reach a conclusion?
15	12	3	In the case that you know what you think about a topic, are you able to think adequately enough to change your mind?

Table 3 Reveals that half of the teachers agree to have interesting topics to write and half of them can sometimes generate lots of words easily when the topic is interesting to them. Less than 5 of the teachers are not much touched by the subject of their interest.

Table 4

Revising

18	6	6	Are you able to revise literally in “reset” sense- hence, change your mind and rethink the fundamental issues that you have stated?
6	18	6	Are you able to determine a prominent point in an unorganized mess of your writing?
6	15	9	Are you able to determine a new shape on a section of your writing that was organized previously?
7	15	8	Are you able to discover problems pertaining to your logic or rationale and appease such problems?
0	30	0	Are you able to clarify your sentences in order for readers to be able to read them on their first reading?
6	21	3	Are you able to liven up your sentences? Are you able to assign a human voice to them?
3	24	3	Are you able to correct the majority of grammatical, punctuation, spelling etc. errors? Are you able to edit your writing in a way as to not put off readers?
6	21	3	Are you able to virtually clear every error?
12	9	5	Are you able to predict the reaction of readers when they read your writing?
18	8	4	Are you able to make modifications to adhere to the needs of specific readers?

From Table 4, it was found that it is easy for more than fifteen teachers to revise and remember the mistakes and correct. Most of the instructors can write lively, get rid of most flaws and make sentences clearer to read. But, less than six respondents can revise and guess how readers react to what they have written.

Table 5

Questions on feedback

18	9	3	Are you able to enjoy sharing a draft of your writing with friends?
9	19	2	Are you able to read a draft of what you've written aloud to listeners thus making it clear and "given" instead of holding back and muttered?
15	15	0	Are you able to listen without judgment to reader reactions and aim to view it as they do, even if you consider their reactions as all wrong?
12	15	3	Are you able to provide non-critical feedback- informing the writer of your requests and reflecting or summarizing the things you hear from the words said?
18	9	3	Can you give noncritical feedback--telling the writer what you would like and summarizing or reflecting what you hear the words saying?
12	15	3	Are you able to provide “movies of your mind” in the position of a reader- a clarified story of occurrences within your mind whilst reading someone else’s writing?
12	15	3	Are you able to provide "criterion-based feedback"-informing readers about the ways in which the draft matches the most prevalent indices of high-quality writing?

Table 5 Indicates that fifteen and more prefer to read out loud to listeners, give noncritical feedback and criterion-based feedback. 18 respondents sometimes share their drafts and give a clear report of what was going on in their minds. Three respondents and less answered that they cannot give movies of their minds, criterion-based feedback and enjoy sharing with friends what they have written. Table 6 shows that in contrast to the population of opposing respondents (3), more than half of the respondents can work on a task collaboratively and nine of them can sometimes work in a group.

Table 6

Question on cooperation

9	18	3	Are you able to cooperate with a small group on conducting a task whilst pitching in, sharing words, helping in the group cooperation and maintaining the group in line with the task?
---	----	---	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Table 7

Consciousness and command of the writing procedure

9	20	1	Are you able to provide a comprehensive account of your writing process: the feelings and thoughts in your mind and occurrences in your writing?
13	14	3	Are you able to discover issues or sticking points amidst your writing and determine the causes of these issues?
12	15	3	Are you able to include changes in your writing process on the basis of noticed things?
9	19	2	Are you able to alter your writing process based on the situation: the readers, subject, writing or type etc.?

Table 7 indicates that 8, 14 to 20 respondents are aware of the writing process as well as the control. More than nine respondents sometimes can vary the style of their writing, render details and notice problems in composition. Less than three of the respondents are not aware, cannot do changes and cannot express their feelings and thoughts in composition.

How do different FA writing procedures contribute to professional practice development?

The essays were conveyed to the contributing teachers who assessed them via a 10-point scale and submitted three reasons or comments which justified their marks. There were no pre-determined indices for the 10-point scale, hence the teachers used personal experience for guidance to conduct qualitative and quantitative assessments. The researcher was mindful of the fact that with no guiding indices, it would be arduous to understand the specific meaning of the scores when applied by various raters. Although, prior researchers e.g. Cumin (1990) did not implement any indices or evaluation categories for rating scales for the purpose of determining how raters themselves defined the indices.

For Essay 1, the experienced teachers exhibited more support compared to the writing teachers of most experience when marking the general essay organization. The positive and negative comments did not entail any substantial differences between the three categories. One of the other essays started with an introduction to this topic: “*traveling lonely or with friends or family*”. *Nowadays fewer and fewer people are traveling with families...* It then

commented on the advantages and disadvantages of traveling lonely, with friends and with family. Such introductory essay organization varied from the development of traditional English essays that normally begin with a thesis statement prior to supporting statements. The comments and attitudes of the more experienced teachers who were less positive suggested their higher sensitivity towards similar breaches of English expository conventions in comparison to the least experienced teachers. In regard to the significantly higher marks given by the least experienced writing teachers for Essay 1, the positive comments given by them for the overall organization may explain their higher marks. Alternately, the writing teachers with the most experience may be traced to attitudes that are less positive.

Concerning the second essay, a substantial difference was evident in negative comments on language fluency among the less experienced teachers and the group of most experienced teachers. The writing teachers of less experience provided substantially higher marks for this essay. They justified their higher fluency marks with less number of negative comments compared to the teachers with the most experience. An initial reading of essay 2 showed the similarity of the syntax used by students throughout the essay. The majority of sentences i.e. 4 from 10, contained the "I/we/want/will/it can ___" structure. When this type of syntax is repeated it hinders the smoothness when reading and it was more notable for experienced teachers.

Various attitudes regarding fluency quality in essay 2 suggested the impact of teaching experience in evaluating L2 writing. In regard to essay 4, there was one substantial difference which concerned the positive comments submitted for the ideas aspect. The teachers of less experience submitted more positive comments for the ideas aspect compared to the more experienced and most experienced teachers. In terms of ideas, essay 4 argued that paper-based newspaper is better than online internet-based news since the newspapers provide comprehensive information which enables critical reflections for readers.

Writers' preferences concerning written communication vary from the preferences of the general public due to the images and sound options conveyed via television which has apparently led to positive comments by the teachers of less experience. Explaining the reasons that more experienced teachers have not placed value on personal and individual ideas amidst their assessments have been difficult. The qualities pertaining to the overall organization of essay 1, the fluency pertaining to essay 2 and the ideas quality for essay 4 showed the ways in which teachers with different experience levels in teaching EFL writing reacted to a specific pupil's writing. In regard to essay 10, the less experienced teachers were the only teachers to submit positive comments for the quality of general language.

According to comparisons between individual essays related to the experience of teachers, the current research suggested that EFL writing teachers of the most experienced submitted significantly fewer scores compared to the teachers of less experienced for essays 1, 2, 4 and 10. Such outcome contributed to the concept related to the impact of teaching reputation concerning L2 writing assessment. Moreover, it echoed Cummings (1990) findings that beginner teachers exhibited higher ratings for different sample composition aspects compared to teachers of more

experience. This research also suggested that similar to tutoring general English, the experience of teachers in EFL writing influences the strictness of teachers/raters in assessing writings. Furthermore, current findings which show the teachers of L2 writing exhibited less leniency in their assessment as they gained more experience or became more cognizant of pupil weaknesses has discrepancies compared to the findings of Song and Caruso (1996) in a way that faculty members with prior ESL teaching experience exhibited more sympathy to ESL student writing issues and thus provided lenient and more holistic assessments. On the other hand, the other findings additionally demonstrate that neither novice teachers nor years of experience affect their appraisal inclinations. Moved on from an ELT division or from another English related one, they have similar beliefs about the uses of appraisal in teaching English. Likewise, the instructors' accounted for similarity on appraisal are not influenced by their long stretches of experience whether novice or experienced. These outcomes are in accordance with the discoveries of a comparable report directed by Calveric (2010). The differences in conclusions in the current research imply that tutoring L2 experience differs more than only exposure to L2 students and relevant texts which several instructors experience. Various student readings are a significant concern for researchers in terms of L2 writing evaluation fairness (Connor-Linton, 1995b, Silva, 1997). This research suggested that utilizing raters of similar experience levels in teaching may appease some potential variances.

Evaluations of qualitative comments pertaining to essays 1, 2, 4 and 10 indicate variances among different teacher categories. As previously mentioned, the teachers of most experience submitted significantly fewer scores when assessing more negatively or less positively traits for the fluency, general language, ideas, and general organization. The teachers of less experience who were either less negative or more positive concerning the aforementioned traits subsequently submitted higher marks for these essays. The most experienced teachers of EFL writing were less positive and stricter in regard to the aforementioned essay aspects which implies that specific assessment attitudes are correspondent to the experience level of teachers/raters.

This study isolates years of EFL teaching as the prominent focal point of the research and provides contributions to the researchers' perception of the ways in which teaching L2 experience may assist in predicting and explain the responses of some readers in regard to student writings.

The differences determined in qualitative and holistic assessments among the EFL writing teachers experience levels were not without similarities between participants, as discovered by this study. Since substantial differences were discovered in four of ten essays, it can be stated that the most and less experienced teachers had similar opinions concerning other essays. The differences and similarities between the teacher categories were an indication of other variables apart from teaching experiencing L2 writing e.g. academic status, educational background, age and L1 as vital parameters in teaching English in Iran. Essentially, despite different experience levels of teaching writing, the teacher categories are different or similar in terms of various background variables that may contribute to the differences and similarities in assessments. This

research also denotes encouraging pilot research for a more comprehensive study with more intricate management and control of variables to outline consequential patterns in the real world teaching contexts like Iran.

Conclusion

This study examined the manner and nature of differences exhibited by teachers of varying experience levels in EFL writing in terms of qualitative and quantitative assessments pertaining to the same circle of pupil essays. Categorically, it investigated the ways in which 46 teachers assessed 10 essays by conducting comparisons between self-generated assessment indices pertaining to the number of years they were teaching EFL writing. The results showed the EFL teachers of the most experience submitted significantly lower marks compared to the less and least experienced teachers when assessing essays 1, 2, 4 and 10. In providing the reasons for their lower marks, the experienced EFL/ESL teachers provided more negative or less positive comments for general language terminology, ideas, language fluency, and general organization.

Even though the current research assists us to see the influence of teaching experience levels with respect to assessing L2 writing, it bears a minimum of two restrictions. Initially, despite the necessity of focusing on variables of real-world context to distinguish the ways teaching operates within a defined context, we are cognizant that outcomes of this field might be restricted upon the escape of a vital variable from a specific control level. As an example, both non-native and native English speaking tutors/raters were sampled to denote two categories of writing teachers in Iranian institutions; expatriates and locals. Moreover, variances in the backgrounds of L1 raters is considered an impactful variable on the ways that raters evaluated L2 texts (Shi, 2000 as an example). He implied that the experience of English teachers of ESL may affect the assessment of texts. Further research can examine the experience of teachers in L2 writing in terms of a variable which may find variances in rating student compositions. As the majority of the least experienced teachers in the research were local whilst the most experienced tutors were expatriates, the variances among these categories may originate from their different L1 backgrounds as well as their diverse teaching experience.

References

- Alhareth, Y., & Ibtisam, A. (2014). The assessment process of pupils' learning in the Saudi education system: A literature review. *American Journal of Educational Research*, 2(10), 883-891.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191.
- Beck, S. W., & Jeffery, J. V. (2007). Genres of high-stakes writing assessments and the construct of writing competence. *Assessing Writing*, 12(1), 60-79.
- Begum, M., & Farooqui, S. (2008). School-Based Assessment: Will It Really Change the Education Scenario in Bangladesh?. *International education studies*, 1(2), 45-53.
- Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science education*, 85(5), 536-553.

- Black, P., & Wiliam, D. (2005). *Inside the black box: Raising standards through classroom assessment*. Granada Learning.
- Calveric, S.B. (2010). Elementary teachers' assessment beliefs and practices. *Unpublished doctoral dissertation thesis*, Virginia Commonwealth University, Virginia, USA.
- Crooks, T. (2001, September). The validity of formative assessments. In *British Educational Research Association Annual Conference, the University of Leeds* (pp. 13-15).
- Brown, Jr, J. (1999). *Assessment matters in higher education*. McGraw-Hill Education (UK).
- Brown, G. T. L., & Gao, L. (2015). Chinese teachers' conceptions of assessment for and of learning: Six competing and complementary purposes. *Cogent Education*, 2(1), 993836. 1-19.
- Callahan, S. (1999). All done with the best of intentions: One Kentucky high school after six years of state portfolio tests. *Assessing Writing*, 6(1), 5-40.
- CONNOR- LINTON, J. E. F. F. (1995). Looking behind the curtain: what do L2 composition ratings really mean?. *Tesol Quarterly*, 29(4), 762-765.
- CONNOR- LINTON, J. E. F. F. (1995). Cross-cultural comparison of writing standards: American ESL and Japanese EFL. *World Englishes*, 14(1), 99-115.
- Cumming, A. (1990). Expertise in evaluating second language compositions. *Language Testing*, 7(1), 31-51.
- Dappen, L., Isernhagen, J., & Anderson, S. (2008). A statewide writing assessment model: Student proficiency and future implications. *Assessing Writing*, 13(1), 45-60.
- Eckes, T., Ellis, M., Kalnberzina, V., Pižorn, K., Springer, C., Szollás, K., & Tsagari, C. (2005). Progress and problems in reforming public language examinations in Europe: cameos from the Baltic States, Greece, Hungary, Poland, Slovenia, France, and Germany. *Language Testing*, 22(3), 355-377.
- Evers, W. J., Brouwers, A., & Tomic, W. (2002). Burnout and self-efficacy: A study on teachers' beliefs when implementing an innovative educational system in the Netherlands. *British Journal of educational psychology*, 72(2), 227-243.
- Gilbert, J., & Graham, S. (2010). Teaching writing to elementary students in grades 4–6: A national survey. *The Elementary School Journal*, 110(4), 494-518.
- Hughes, A., & Lascaratou, C. (1982). Competing criteria for error gravity. *ELT Journal*, 36(3), 175-182.
- Jeffery, J. V. (2009). Constructs of writing proficiency in US state and national writing assessments: Exploring variability. *Assessing Writing*, 14(1), 3-24.
- Koh, K., Burke, L. E. C. A., Luke, A., Gong, W., & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*, 22(3), 264-288.
- Leung, C., & Mohan, B. (2004). Teacher formative assessment and talk in classroom contexts: Assessment as discourse and assessment of discourse. *Language Testing*, 21(3), 335-359.
- Lee, I. (2007). Feedback in Hong Kong secondary writing classrooms: Assessment for learning or assessment of learning?. *Assessing Writing*, 12(3), 180-198.
- Moon, T. R., & Hughes, K. R. (2002). Training and Scoring Issues Involved in Large-Scale Writing Assessments. *Educational Measurement: Issues and Practice*, 21(2), 15-19.
- Nemati, M., Alavi, S. M., Mohebbi, H., & Masjedlou, A. P. (2017). Teachers' writing proficiency and assessment ability: the missing link in teachers' written corrective feedback practice in an Iranian EFL context. *Language Testing in Asia*, 7(1), 21.
- Önalın, O., & Karagül, A.E. (2018). A study on Turkish EFL teachers' beliefs about assessment and its different uses in teaching English. *Journal of Language and Linguistic Studies*, 14(3), 190-201.

- Riggs, I. M., & Enochs, L. G. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education*, 74(6), 625-637.
- Rotter, J. B. (1990). Internal versus external control of reinforcement: A case history of a variable. *American Psychologist*, 45(4), 489.
- Song, B., & Caruso, I. (1996). Do English and ESL faculty differ in evaluating the essays of native English-speaking and ESL students?. *Journal of Second Language Writing*, 5(2), 163-182.
- Stiggins, R. J. (1991). *Relevant Classroom Assessment Training for Teachers*. Educational Measurement: Issues and Practice, 10 (1), 7-12.
- Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of educational research*, 68(2), 202-248.
- Van Driel, J. H., Verloop, N., & De Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 35(6), 673-695.
- Wininger, S. R. (2005). Using your tests to teach: Formative summative assessment. *The teaching of Psychology*, 32(3), 164-166.
- Zou, P. X. (2008). Designing effective assessment in postgraduate construction project management studies. *Journal for Education in the Built Environment*, 3(1), 80-94.