

EFL Writing Assessment Practices: Teachers' Perspectives

Somruedee Khongput, Prince of Songkla University, Thailand
somruedee.kh@psu.ac.th

Abstract

In classroom contexts, teachers as raters are required to assess students' performance-based writing tasks. What criteria they use and how they rate students' writing do not seem well understood (Lumley, 2002). Research into rating processes has been conducted widely in English as a second language (ESL) contexts (e.g. Connor-Linton, 1995; Cumming, 1990; Lumley, 2002). However, only a few studies of EFL contexts (e.g. Shi, 2001) can be found, especially in Thailand (Chinda, 2009). The purpose of this study is to explore assessment practices of teachers of writing for English major students in Thai public universities across the country. Teachers' views on assessment practices and their actual practices were investigated. A questionnaire was used to gather data about teachers' personal and professional backgrounds as well as their views on effective writing and on marking writing. The teachers of writing were asked to complete the questionnaire online during the semester of a writing course they were teaching at the time. This presentation will focus on findings from the questionnaire. Teachers' views on effective writing and how to mark writing will be discussed.

Keywords: performance-based writing, writing assessment practice, online questionnaire, teachers' perspectives

Introduction

Assessment of performance-based writing is not an easy task for teachers of writing. It can be seen as a time-consuming and complex activity. When assessing writing tasks where human raters are required, raters' subjectivity may play an important role. Raters' biases towards student performances, their different perceptions of good writing and their cultural and professional backgrounds are all factors that can influence the rating (Cumming et al., 2002; Kobayashi, 1992; Shi, 2001; Wood, 1993).

Though rating criteria are employed in the rating process, some variability between practices still occurs. Wood (1993) and McNamara (1996) posited that there are a lot of variables involved in rating writing, such as between-writer variation, within-writer variation, within-rater variation and between-rater variation or even physical features such as the neatness of handwriting.

As Lumley (2002) stated, raters are the most significant component of the rating process because they can make decisions about scale features they will focus on, how to adapt scale wording to suit their situations, and how they justify the written texts according to their educational contexts and requirements. Thus, raters need to keep many things in mind while they mark in order to maintain rating reliability and validity of the rating.

A number of research studies have contributed to the study of rating process. Cohen (1994) explored several studies of both first and second language writing (e.g. Connor & Carrell, 1993; Huot, 1990; McNamara, 1990; Rafoth & Rubin, 1984) and found that raters are likely to focus on grammar and mechanics more than they realize. Also, raters tend to employ criteria different from the central guidelines they get (Cohen, 1994; McNamara, 1996).

Moreover, the score given to the same piece of writing by different raters and that given to different pieces of writing by the same raters may be different. Besides, some issues such as students' handwriting, time of marking in a day, or teachers' preferences towards students may also cause bias in assessing writing (Wood, 1993).

There are tendencies that teachers with different backgrounds will have different perceptions towards good writing and thus tend to focus more on some specific features. Shi (2001) conducted a study to investigate how different native speaker English teachers and nonnative speaker English teachers rate their Chinese university students' writing. The teachers were asked to rate writing samples holistically using their own criteria and to provide three reasons based on the rank of importance to support their judgment. The results showed that though both groups of raters gave similar scores to the writing, they weighted writing features differently in their rating. The native

English-speaking teachers focused more positively on content and language while Chinese teachers stressed more negatively on organization and length of the writing.

In Cumming et al. (2002)'s study which aimed at investigating features of TOEFL essays rated by experienced raters, they found that the qualities perceived as good writing in the examination context varied among raters. Ten raters were asked to identify the three most significant features of writing from their point of view. It was discovered that the most frequently mentioned features were rhetorical organization (nine raters), expression of ideas (nine raters), accuracy and fluency of English grammar (seven raters) and vocabulary, and the length of writing (two raters).

Teachers can also have different perceptions and practices regarding rating criteria. In Chinda's (2009) pilot study exploring Thai teachers' perspectives in writing assessment practices, he found that the teachers had different views towards criteria and employed them differently. Even though they had central criteria to follow, they applied them in individual ways. Some teachers tried to follow the criteria though they did not agree with them. Some added their own criteria when marking students' work.

To understand more about writing assessment practices, it is worth exploring raters' perceptions concerning good writing and writing assessment as well as investigating how they actually mark their students' writing. As Connor-Linton (1995: p. 763) stated "if we do not know what raters are doing (and why they are doing it), then we do not know what their ratings mean."

The study

This paper reports the first part of a bigger study which aims at exploring assessment practices of teachers of writing for English major students in Thai public universities. Teachers' views on assessment practices and their actual practices will be investigated. In the first part of this project, the research questions are as follows:

1. What do teachers of writing in Thai public universities perceive as good writing?
2. What criteria do they use in marking students' writing?
3. What marking practices do they have?

Methods

In the first part of the study, a questionnaire was developed. It was piloted many times with various groups of people including teachers who have experience in teaching writing in Thailand before finalising it. The questionnaire comprised seven parts with both closed and open questions. The first part is devoted to demographic and general information about the respondents. The second and third parts are for gathering information about respondents' professional background. The part on views on good writing was designed to collect participants' opinions about good writing in general while the next two parts are for collecting data about the current writing courses the participants teach and their opinions towards marking students' writing in a particular course. The last part is for respondents to provide comments on writing assessment in Thailand.

An online software 'Survey Monkey' (<http://www.surveymonkey.com/>) was used to construct the questionnaire and gather data. An email message was sent out to heads of English department in public universities in Thailand from mid-June requesting them to forward the message to their colleagues. Over a month of sending out the questionnaire, 40 respondents started the questionnaire and 21 of them completed it. The collection of responses is ongoing. This paper reports the preliminary findings from the 21 respondents mentioned.

Results

This section provides preliminary assessment of data collected from the online questionnaire. It starts with general information on the respondents including their demographic information and professional experience. Then, it is followed by the results on teachers' views on good writing and writing assessment.

1. Description of respondents

From 21 respondents who completed the questionnaire, the number of male writing teachers (14) is twice the number of female teachers (7). Most respondents are in 31-40 age range (8), followed by those in the age ranges of 51-60 (6), 41-50 (4) and 26-30 (3), respectively. For their educational

qualifications, 12 out of 21 respondents possess a master’s degree, followed by a doctoral degree (7 respondents) and a bachelor’s degree (2 respondents).

The majority of the respondents are Thai teachers of writing (17 out of 21 respondents) and the rest are native speakers of English. Of 17 Thai teachers, 13 reported they had at least one year’s experience in studying or living in English speaking countries.

When asked about their teaching experience, 20 respondents reported they had at least four years of English teaching experience. Only one respondent reported having less than one year of teaching English. For their experience in teaching writing, it is found that five respondents have more than 10 years of teaching writing experience while three of the respondents have less than two years of experience in teaching writing.

2. Teachers’ perspectives on good writing

Regarding views on good writing, the teachers were asked to rate the degree of importance of seven writing features as shown in Table 1. The rating is a 5-point numerical scale ranging from 1 (not important) to 5 (very important). The results show that most teachers perceived all features as very important. However, when considering the ranks of importance, the three most important writing features perceived in good writing are relevant development of ideas, organisation/overall structure of the text and cohesion. Mechanics are perceived as the least important.

Table 1: Rating average of writing features

Writing features	Not important (%)				Very important (%)	Rating average
Relevant development of ideas	0.0	0.0	4.8	9.5	85.7	4.81
Organisation/Overall structure of the text	0.0	0.0	4.8	23.8	71.4	4.67
Cohesion	0.0	0.0	9.5	23.8	66.7	4.57
Task completion	0.0	0.0	9.5	28.6	61.9	4.52
Appropriate vocabulary use	0.0	0.0	19.0	38.1	42.9	4.24
Appropriate grammar	0.0	4.8	14.3	38.1	42.9	4.19
Mechanics (e.g. spelling, punctuation, capitalisation)	0.0	9.5	23.8	33.3	33.3	3.90

The results also show that some teachers considered that grammar (4.8%) and mechanics (9.5%) are not very important components in good writing compared to other features. 23.81% of the teachers provided other writing features they perceived as a component in good writing. Most of the features mentioned concern the impression of the writing. They include genre, originality/creativity of ideas, rhetoric, aesthetic and enjoyability of the writing. Also, students’ ability in transferring one language to another language is taken into account as mentioned as “conscious suppression of L1 negative transfers from Thai.”

3. Teachers’ perspectives towards marking writing

The teachers were provided with 13 statements of assessment practices and asked to indicate the level of frequency each statement applies to them (see Table 2). The statements can be grouped into two broad categories, namely, kinds of criteria used (statements 1-5) and teachers’ practices in marking writing (statements 6-13).

Regarding marking criteria, it is found that teachers employed different types of criteria. A third (33.30%) of the teachers reported they give a single score to the overall quality of writing most of the time. 28.6% often compare the quality of writing in the same task and give scores according to rank orders of all students’ writing against each other. A third (33.3%) indicated that they also often give scores to distinct writing features. Kinds of writing text may be an important indicator of what criteria to use since almost half (42.9%) of the teachers reported they employ different sets of criteria

when marking writing in different text types most of the time. The results also reveal that the teachers hardly ever assign equal weighting to every feature of writing (42.9%).

Table 2: Level of frequency of writing assessment practices

Teachers' perspectives regarding marking writing	hardly ever (%)	Occasionally (%)	Often (%)	most of the time (%)	Always (%)	N/A (%)
1. I mark compositions/essays by giving a single score to the overall quality of writing.	9.50	19.00	19.00	33.30	9.50	9.50
2. I compare the quality of all compositions/essays in the same task and give scores according to their rank orders.	9.50	19.00	28.60	23.80	19.00	0.00
3. I mark compositions/essays by giving scores to distinct features of writing (e.g. task completion, cohesion, mechanics).	9.50	9.50	33.30	28.60	14.30	4.80
4. I mark different types of compositions/essays by using different sets of criteria.	19.00	9.50	14.30	42.90	9.50	4.80
5. I assign equal weighting to every feature of writing.	42.90	23.80	14.30	4.80	9.50	4.80
6. Before marking compositions/essays, my co-teacher(s) and I discuss marking criteria and mark the compositions/essays based on our agreement.	4.80	19.00	19.00	4.80	33.30	19.00
7. Before marking compositions/essays, my co-teacher(s) and I select and mark a sample of compositions/essays to ensure score reliability.	33.30	9.50	14.30	4.80	9.50	28.60
8. I adjust the criteria during the marking period without consulting my co-teacher(s).	28.60	28.60	14.30	0.00	4.80	23.80
9. I compare different pieces of writing to ensure the score consistency.	4.80	9.50	38.10	19.00	28.60	0.00
10. I read compositions/essays more than once before giving scores.	4.80	4.80	38.10	28.60	23.80	0.00
11. The final score I give to all compositions/essays is based on the criteria.	0.00	4.80	14.30	14.30	61.90	4.80
12. The quality of students' handwriting affects the grade I give their compositions/essays.	61.90	23.80	0.00	4.80	0.00	9.50
13. Knowing who the writer is affects my decision of what score to give.	66.70	23.80	4.80	0.00	0.00	4.80

Considering their answers regarding their marking practices, the results show that the teachers varied in the degree of cooperation with their co-teacher(s). Most of the teachers cooperate with their co-teacher(s) at some degrees while some teachers may not have to cooperate at all.

The results reveal that the teachers try to adhere to the criteria they have. There are tendencies that most teachers seem to maintain reliability of the marking. Larger proportion of teachers answered on the higher frequency side of the scale when asked about the score consistency. Moreover, the teachers seem not to have biases while they mark writing. Most of the teachers reported that the scores they give are not affected by the quality of students' handwriting (61.90%) and by the fact that they know who the writer is (66.70%).

4. Teachers' voices about writing assessment practices in Thailand

Comments about writing assessment practices in Thailand were gathered from open-ended questions in the questionnaire. The results reveal some of their concerns, their practices and their suggestions about writing assessment.

Regarding concerns over students' writing ability, the role of grammar can lead to low quality of writing outcome. Two English native teachers who have ten and nine years of teaching writing experience addressed their concerns about this. While Terry¹ stressed that his students have problems with grammar, vocabulary and sentence structure, Michael¹ stated the lack of grammar course is the cause of students' low ability in writing. They commented:

As a native speaker, we often hear that we should give high marks for the idea of the paper. However, I usually find that my students have very poor grammar and very poor sentence structure and very poor and incorrect vocabulary usage which severely limits [sic] their ability to complete assignments at a higher level. [Terry]

There are no grammar-course prerequisites to the writing classes. This is a big mistake. It places the cart before the horse. Courses in composition writing or even paragraph writing become reduced to coursework -review in basic sentence writing and basic grammar review. Therefore, the term "assessment practices" sometimes begs the question: What exactly is being assessed? [Michael]

Writing is seen as a skill that not only requires linguistic ability but also some other abilities beyond the language itself. This may raise a question about the capability of criteria to measure writing. Anong and Daniel¹ who have more than ten years of experience in teaching writing seem not to believe in the objectivity of the marking. They tend to favor more on subjective way of marking. They elaborated:

I don't believe that the grads [sic] as a result of work evaluation can explain students' competence. We have assessed students' learning by using some methods that we called 'scientific ways', not realizing that human elements are so complex that they cannot be evaluated scientifically as scientists do in their scientific experiments. I feel apathetic for the students who sit in classes whose teachers are grammar-oriented, who never appreciate nothing but correct, but dead, grammar. these [sic] students have well organized, well written pieces, but these pieces are useless in terms of minds and souls. [Anong]

I think there is inconsistency, but I do not believe any more than anywhere else. The problem with criteria is that they tend to invite a mechanical application, as if grading papers is a science, whereas I think it is more of an art. Criteria give the illusion of greater fairness and reliability, but I have reservations. As I pointed out, when I apply criteria, students tend to get lower grades than the overall grade I believe they deserve. We have had cases where teachers apply the same criteria very differently. I think the best probable solution is a (somewhat messy) mixture of criteria, judgment and experience. [Daniel]

The difficulty in teaching and assessing writing seems to affect the assignment of teachers to teach writing. A teacher who has less than one year experience in teaching English expressed her

¹ These are pseudonyms of the teachers who gave comments on writing assessment practices in Thailand.

point of view that at her workplace new lecturers having good writing ability are assigned to teach writing because no one else would want to teach. She pointed out:

Most lecturers here do not want to teach any writing course. The newest lecturers, with no or little experience, have to teach writing to English major students and have to keep teaching it until they are not the new faces anymore. Then, the new lecturer(s) can take over. The department might think that a person's ability to write well (or even all right) can be equated to the ability to teach writing.

Apart from this, some teachers provided further information about their marking practice. In terms of correction, a teacher mentioned that he used correction symbols to help students practise correcting their own work. Bias against handwriting is also taken care of since a teacher explained he asked the students to type their writing using a font format he assigned.

Another teacher also suggested that double marking method helps maintain score reliability. He explained:

For the group of 15 students and below, after you mark students' work, it is a good idea to ask one of your writing teachers to mark again. Then compare your marks and the marks given by your colleague to see if the marking is reliable or not.

Though these results were gathered from a small number of respondents, they reveal many noteworthy issues in marking writing practices in Thailand and these will be discussed below.

Discussion and conclusion

The questionnaire reveals that most of the respondents are experienced teachers. They used to study or live in English speaking countries and many are experienced teachers of English with some experience in teaching writing for EFL students. Their experience as a language user and learner may influence their preference of good writing and how to teach and mark writing. As language users, they can be exposed to various kinds of writing in real life such as news articles, letters and e-mail correspondence. This experience may help build up their perceptions towards what features constitute good writing. As language learners who are required to write assignments or articles in English, they may learn how to write better or how to mark writing from their teachers or reviewers' marking practices.

What the teachers perceived as important in good writing shows that to some extent they have similar perceptions towards component of good writing. They tend to focus more on a discourse level than on grammatical level.

Regarding the criteria employed in marking writing, the results from the questionnaire suggest that in general teachers employ various kinds of criteria and their choice of the criteria may depend on types of writing they have to mark. When applying a set of criteria in the marking process, it seems that the teachers agree to use central criteria to mark students' writing even though they do not have any practice in applying the criteria to mark the writing before the actual period of assessment. The results reveal that they can have full authority over how to score their students' writing without consulting their co-teacher(s).

However, they try not to adjust the criteria they have while marking. This may show that to some extent they consider between-rater reliability and within-rater reliability though they seem to stress the latter more. The teachers tend to maintain within-rater reliability by sticking to the central criteria they have and avoiding being biased towards physical features of writing and their realisation of the writer. This finding seems not surprising because it was gathered from questionnaire which is a respondents' self-report. The respondents may answer the questionnaire based on what they think or believe they do but not what they actually do.

The results also show that the teachers perceived writing as a difficult skill to learn since it requires more than language skill to construct a good piece of writing. For example, students may be required to use their creative thinking in order to make readers enjoy reading their work by giving "good attention-getting beginning and good memory-ending" as mentioned by one of the teachers. As Hamp-Lyons (2003) mentioned, "writing is a very complex activity involving thinking, planning,

organizing, and linking as well as several levels of language manipulation (sentence and clause levels as well as word and phrase level, plus spelling, punctuation, etc.)". This may lead to difficulty in assessing writing; some teachers mentioned writing as "an art" which requires "the mixture of criteria, judgment and experience" as an assessment method.

It seems that writing cannot be purely assessed objectively. While there have been attempts to promote objectivity of marking by employing marking criteria and rater training (e.g. Chinda, 2009; East, 2009; Lumley, 2002), subjectivity may also have to play a role. Some features such as creativity and originality of writing may need to be judged based on a rater's personal point of view.

Even though the findings show that grammar is perceived not as important as writing features in a discourse level, lack of appropriate grammar may obstruct readers' understanding of the writing and can cause low marks in writing. Generally, grammar is among the first language skills taught in schools in Thailand. When students start learning English, they also learn grammar. Thus, students have knowledge of grammar at some degree when they enter university. Nevertheless, teachers' concern about grammar in writing seems to indicate that students may not apply or may not know how to apply their knowledge of grammar into writing. As a result, they cannot perform well in writing.

As mentioned earlier that writing is a complex activity, it can be a big burden for teachers to guide their students to become a good writer. Because of this, some universities may put basic sentence writing or grammar review into their writing courses in order to help their students to write better. A lot of time needs to be devoted to teaching students who can write poorly to become good writers as well as to marking their writing. With teachers' heavy load of teaching, this big burden may make teaching writing become a hot potato to some teachers. Thus, junior teachers may have to take responsibility over writing courses instead of experienced teachers.

These findings from the questionnaire can shed some light to what teachers of writing in Thai universities think and believe in terms of writing assessment. However, there are limitations of this part of the study. Firstly, with a small number of questionnaire respondents these findings cannot be generalised to all writing teachers in the universities across Thailand. The low response rate of this questionnaire may have resulted from various reasons. One reason may be the use of web-based questionnaire itself that needs follow-up reminders to increase the response rate (Sue & Ritter, 2007). The period of data collection may also play a role. The email requesting teachers to participate in the questionnaire was sent out in mid-June when the first semester in Thailand had just started. Teachers may have been too busy teaching and managing their courses to complete the questionnaire. The contrast of the number of teachers starting the questionnaire and those completing it may suggest that some respondents might have tried to answer it before realising later that they were not potential respondents. So they may quit answering the questionnaire at a certain part of it. As mentioned earlier, collecting data through this questionnaire is therefore still ongoing.

Secondly, only general views of writing assessment practices drawn from the questionnaire as descriptive analysis was used in this preliminary assessment of data. More interesting findings may be found if inferential analysis is used to compare groups of teachers based on their personal and professional backgrounds such as their first language, and their teaching and marking training experiences. Lastly, this questionnaire can only reveal what teachers think and believe they do but cannot reveal what they actually do. The actual rating process can be further investigated in the second part of this study. Its finding can also be compared with what teachers think and believe to provide a clearer view of writing assessment practices in Thailand.

This paper has drawn an overview of teachers' perspectives in writing assessment practices in Thailand from a small number of respondents. However, it can show a direction of the second phase of the study.

Acknowledgement

I would like to thank Professor Anne Burns and a group of PhD colleagues including Sara Cotterall, Apiwan Nuangpolmark, Siri Bounchan, Fumiko Murase and Theresa Wannisch for their useful comments on the initial draft of this article.

References

- Chinda, B. (2009). *Professional development in language testing and assessment: a case study of supporting change in assessment practice in in-service EFL teachers in Thailand*. PhD thesis, University of Nottingham.
- Cohen, A. (1994). *Assessing language ability in the classroom* (2nd ed.). Massachusetts: Heinle & Heinle Publishers.
- Connor, U., & Carrell, P. (1983). The interpretation of tasks by writers and readers in holistically rated direct assessment of writing. In J. Carson & I. Lexi (Eds.), *Reading in the composition classroom : Second language perspectives*. Boston: Heinle.
- Connor-Linton, J. (1995). Looking behind the curtain: what do L2 composition ratings really mean? *TESOL Quarterly*, 29(4), 762-765.
- Cumming, A., Kantor, R., & Powers, D. (2002). Decision making while rating ESL/EFL writing tasks: a descriptive framework. *The Modern Language Journal*, 86, 67-96.
- East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing*, 14, 88-115.
- Hamp-Lyons, L. (2003). Writing teachers as assessors of writing. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 162-189). Cambridge ; New York: Cambridge University Press.
- Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, 60(2), 237-263.
- Kobayashi, T. (1992). Native and nonnative reactions to ESL compositions. *TESOL Quarterly*, 26(1), 81-111.
- Lumley, T. (2002). Assessment criteria in a large-scale writing test: what do they really mean to the raters? *Language Testing*, 19(3), 246-276.
- McNamara, T. (1990). Item response theory and the validation of an ESP test for health professionals. *Language Testing*, 7(1), 52-76.
- McNamara, T. (1996). *Measuring second language performance*. New York: Longman.
- Rafoth, B., & Rubin, D. (1984). The impact of content and mechanics on judgments of writing quality. *Written Communication*, 1(4), 446-458.
- Shi, L. (2001). Native- and nonnative-speaking EFL teachers' evaluation of Chinese students' English writing. *Language Testing*, 18(3), 303-325.
- Sue, V. M., & Ritter, L. A. (2007). *Conducting online surveys*. Thousand Oaks, California: Sage Publications.
- Wood, R. (1993). *Assessment and testing: A survey of research*. Cambridge: Cambridge University Press.