Introducing Peer Assessment to a Reading Classroom: Expanding Thai University Students' Learning Boundaries beyond Alphabetical Symbols

Saowaluck Tepsuriwong

Thanis Bunsom

Department of Language Studies, School of Liberal Arts
King Mongkut's University of Technology Thonburi, Bangkok, Thailand
Email: saowaluck.tep@kmutt.ac.th, thanis.bun@kmutt.ac.th

Abstract

Liu and Carless (2006) propose that peer assessment through peer feedback giving has its potential for enhanced student learning. Introducing formative assessment, using valid and reliable rubrics, can be beneficial to both teachers and learners in the process of learner improvement. However, Carless's experiences in Hong Kong, a Confucius-based society where high-stake summative examinations are heavily prioritised, are solid proofs of how culture can interfere with education. In Asian societies, including Thailand, teachers having superior and absolute power in assessment are respected by students and hence, teacher voices take precedence over those of their peers. Implementing peer assessment in Asian classrooms, therefore, is problematic and to a certain extent, impossible. In this research, in order to expand our students' learning boundaries beyond familiar, traditional ways of teacher assessment of which an end result is a grade, we introduced peer assessment rubrics to our reading classroom to allow them to monitor their friends' reading progress. The peer evaluation results and comments were analyzed and compared with that of the teacher. An in-depth interview was also conducted with six subjects after the assessment process. The findings showed that as most of our students were engineering students and the texts they were reading were literary excerpts, employing rubrics was very helpful in the learning process. Nevertheless, as a result of long standing culture, final grades were still their priority. Our conclusive suggestions were that structural changes in assessment would be needed and they would certainly take time amidst the assessment culture in Thailand.

Keywords: Peer assessment, Rubrics, Assessment for learning, Reading class, Culture

1. Introduction

Since ancient times, assessment has always played a crucial role in education. Many people especially in Asian societies view assessment as a major criterion for success and failure. This belief is a result of long-standing practice initiated from Imperial China, regarded as "having introduced the first systematic assessment system for civil service appointments" (Min and Xiuwen, 2001, p.5). It can be said that the practice was imposed to determine people's positions in the country and separate the wise from the not so wise. Today, assessment in the West may be differently viewed as there have been attempts to emphasise the formative assessment, a type of assessment that places importance on feedback for learning improvement to help drive learning from the hearts and minds of students through formative processes. However, as Carless (2011) points out, the paradigm shift from summative assessment to formative one is almost impossible in the Eastern world thanks to cultural foundations of those countries heavily influenced by Confucian heritage.

It is not too far-fetched to claim that in Thailand, as well as in other Confucian heritage societies, summative assessment has always been used as a final evaluation to measure learning achievement. Since the national entrance into Thai tertiary education relies heavily on summative assessment, mainly multiple choices, most –if not all- Thai high schools deploy the curricula that will see as many students as possible pass the entrance examination into state-run universities. To achieve that end, teachers spend hours and hours training their students to choose the right answers for test questions whereas students ignore the learning process and pay more attention to the results.

Furthermore, in a culturally-influenced classroom setting, when results of examinations, either in terms of scores or grades, take precedence, teachers gain more authority as they have power to assign the results and therefore automatically determine their students' academic futures. In universities, the belief that teachers are to be respected and obeyed has continued to dominate the learning environment and that is the reason why summative assessment is still

prioritised, resulting in different washbacks on students' learning experiences.

2. Rationale of the Study

Aware of the prominent role of summative assessment in Thai society and its effects on our students, we initiated a research project that would see assessment as ways of improving our students' language learning abilities as well as our own methods of instruction instead of simply telling them whether they passed or failed the course. Assessment, in our context, should encourage students to develop themselves and put the spark back into their academic lives. Also, as we regarded assessment as means of improving our students, there was no reason why they should be excluded from the assessment processes. As a consequence, they were given opportunities to view their friends' assignments and give necessary feedback.

3. Research Questions

To expand Thai undergraduate students' learning boundaries beyond familiar, traditional ways of teacher assessment of which an end result is a grade, we introduced the deployment of rubrics with clear descriptors in a reading classroom at King Mongkut's University of Technology Thonburi and allowed students to assess and give feedback to one another. Reliability of peer evaluation and potential for implementing rubrics as learning-oriented assessment in a Thai context were critical concerns of this paper. In so doing, we attempted to answer the following research questions:
    1. *What are the peer evaluation results in comparison with those evaluated by the teacher?*
    2. *What are the students' perceptions on peer assessment?*

4. Literature Review

4.1 Three Types of Assessment

According to Ng (2008) and Airasian (2008), the types and definitions for assessments are as follows:
    i. Assessment of learning: this type of assessment is also known as summative assessment in which grades or scores are often assigned for the purposes of measuring learners' achievement against a standard. Traditional emphasis is given on national examinations and school examinations after the learning, usually at the end of the semester.
    ii. Assessment for learning: this is formative assessment that could take the form of diagnostic assessment or assessment that takes place throughout the learning such as feedback provision, progress monitoring forms and one-to-one consultation. The consistent monitoring of learning and teaching aims at improving students' performances and checking the effectiveness of teachers' instruction.
    iii. Assessment as learning: for this type of assessment, students are made aware of their learning goals to track their performance by themselves. If necessary, they will take remedial steps to improve their learning. As there are no fixed means of assessment, it should only be used with those who are independent and disciplined learners.
    For the purpose of this study, assessment for learning is the centre of our focus because the initial objectives of our research are to find out whether peer assessment with rubric sheets can help provide our students with better learning environment and enhance their learning experiences in an English reading class.

4.2 Rubrics

Rubrics are generally defined as assessment tools that help evaluate an assignment by indicating criteria that describe levels of qualities on a scale usually from excellent to poor (Andrade, 2010). Studies indicate that employing rubrics reinforce reliability in assessing performance (Hafner & Hafner, 2003). Raters can refer to explicit evaluation criteria designed as guidelines for interpreting students' performance thus reducing subjectivity in evaluation, and this leads to high agreement in the evaluation results.
    Essential features in rubrics include evaluation criteria and quality definition. The former are lists of factors that assessors considered when determining the quality of assignment. These factors may reflect either the process or the content required for successful task completion. The latter show detailed explanation of levels of qualities or proficiency

demonstrate at a particular level on a scale which could later be calculated into scores or grades.

These features provide both final scores that judge the quality of student's work and useful information to be fed into the learning process. That is to say rubrics can be used for both assessment and instructional purposes (Andrade, 2005). Their criterion-referenced nature provides a bench mark for students to know how well they perform on a particular skill or subskill demanded. They provide a roadmap for students to plan their route to success, indicating the goal and stages to move towards the goal. Using rubrics also provides formative feedback for students to improve their learning (Allen, 2010). Rubrics help learners realize the present levels of their performance and explicitly see the target of their learning, and therefore, understand what to do to achieve the target. Students reported that rubrics actively involved them in checking their work and reflecting upon the feedback obtained which guided them to monitor towards better results. They also perceived that rubric-referenced grading was fair and transparent. Moreover, a review of research studies in higher education points to positive directions in the use of rubrics with graduate and undergraduate students (Reddy & Andrade, 2010).

The values of rubrics in enhancing learning could not be taken for granted. It would be a great loss if their use is limited to only teachers who use them for summative procedures in grading the final products as in a traditional teacher-centred evaluation scheme. In fact, nowadays, potential of rubrics as learning-oriented assessment tools has been discussed and introduced in various contexts (Carless, 2007). Liu and Carless (2006), for example, encouraged students to give peer feedback in formative assessment in Hong Kong, using valid and reliable rubrics, which were reported to be beneficial to both teachers and learners in the process of learner improvement.

4.3 Problems with Peer Assessment

The introduction of peer assessment, however, is not without any problems. Learners may lack experience in providing effective, informative feedbacks or do not know how to avoid subjectivity. In addition, they could be biased based on individual preferences or find it difficult to trust or rely on their peers who are not in the authoritative status. Peer assessment can also be demanding because it requires 'double duty' of grading and learning, both of which involve complicated skills (Boud, 2000). Grading and evaluation involves measuring against a standard. It may also invoke comparisons between individuals to guarantee fairness and transparence. On top of that, learning demands adequate and appropriate suggestions and guidelines for further development. Culture and familiarity can also obstruct this form of assessment. A paradigm shift is, therefore, necessary to equip students especially in a context where students have long been taught to conform to teachers' instructions and absolute authority in a classroom. Carless's experiences in Hong Kong, a Confucius-based society where high-stake summative examinations are heavily prioritised, are solid proofs of how culture can interfere with education (Liu and Carless, 2006). In Asian societies, including Thailand, teachers having superior and absolute power in assessment are respected by students and hence, teacher voices take precedence over those of their peers. Implementing peer assessment in Asian classrooms, therefore, is very challenging.

5. Research Methodology

5.1 Research Settings

The study was conducted as classroom-based research in a natural classroom setting of two reading classes, each of which consisted of 20 engineering students taking the course as their electives. Despite differences in language abilities, these engineering students were enrolled in the course based on their interest in enhancing their English reading skills. The course was taught by the same teacher using the same instructional techniques and evaluation scheme. In class, students read literary excerpts and discussed issues outlined in a worksheet as reading evaluation. At the beginning of the semester, the worksheets were solely marked and evaluated by the teacher. Traditional teacher-centred assessment was clearly mirrored in this phase of the class. This evaluation scheme confined the learners in a dependent role relying on the teacher in judging levels of students' performance, and providing feedbacks and suggestions for their learning. The teacher, therefore, sought for an alternative means of assessment, designing and introducing rubrics to encourage active involvement and learning-oriented means of peer evaluation.

5.2 Research Tools

The concept of 'learning-oriented assessment' proposed by Carless (2007) was used as a guideline for designing a

rubric for peer assessment in this study. The rubric was designed by the teacher for students' peer evaluation of a worksheet on Licata and Chong's *Cocktail* (2009), a contemporary play that they individually read and reflected upon. The rubric consisted of seven criteria focusing on the following aspects:

1. Ability to understand the plot of the story.
2. Provision of logical explanations about the central conflict of the story.
3. Explanation on how the main character has developed from the beginning to the end of the story.
4. Identification of 'inauthenticity' of the play.
5. Recognition of certain stereotypical portrayals of Thailand and its people through the characters.
6. Understanding of a moral of the story.
7. Overall understanding of the play, the characters and the writer's intention.

These criteria were meant to measure learners' reading achievement after reading the play, which included demonstration of understanding the plot and the conflict in the play, the background of the characters and the settings, and critical reading: noticing inauthenticity and concluding the lesson learnt.

The quality definition used for measuring these seven evaluation criteria was the four-point rating scale that required them to judge whether they 'strongly agree', 'agree', 'disagree', or 'strongly disagree' that their peer could perform on each criterion. For the sample of the rubric sheet, refer to Appendix 1.

5.3 Steps of Conducting Research

The rubric was introduced and explained to the learners in the reading class. Then, after individually finishing the worksheet, they were randomly put together in pairs using the rubric to evaluate their partner's work. They were also encouraged to provide written feedback or extra comments to monitor their peer's reading progress.

A week after the use of the rubric, the learners were asked to work to reevaluate their peers' worksheet without using the rubric. Both peer evaluation results and comments were analysed and compared with those of the teacher to investigate the reliability of the assessment. As each reading task was accounted for 10% of the course evaluation, the scores from peer evaluation using the rubric, the teacher evaluation, and the peer evaluation without the rubric were calculated to the total score of 10 to ease the comparison.

Due to practical reasons, a complete set of data was available from 25 out of 40 students. Therefore, there were 25 subjects for the score analysis and six subjects from this group were called in for an in-depth interview on a voluntary basis to focus on their perceptions, attitudes, problems encountered, and suggestions to gage potential of rubric use.

5.4 Limitations of the Study

The results of the study may have been affected by the three following factors. The first factor is the duration of the research. As time was a major issue, different results could have been earned had the research been conducted for a longer period of time. The second limitation was caused by the students' lack of exposure to the use of rubric descriptors in peer assessment. The final limitation was that we did not further investigate the correlation values after comparing the three sets of scores, which could possibly yield further implications of the similarities and differences between peer evaluation and teacher evaluation.

6. Results and Discussion

6.1 Peer Evaluation Results

Employing the rubric in their evaluation process, the subjects gave the scores ranging from 5.71-9.64 with the mean score of 7.59, suggesting that most of them were satisfied with their peers' work. They seemed to agree that their friends understood the story well and satisfactorily achieved the task objectives.

However, when they were asked to reevaluate their peer's worksheet without using the rubric, it was noticed that the scores given were slightly lower. The range was from 0.83-9.17 with the mean score of 5.63. This time, ten subjects scored lower than 5, suggesting that they failed the task. Without using the rubrics, the scores were much lower.

When these scores were compared with those of the teacher, a difference in the scores was also noticed. The scores from the teacher ranged from 0 – 10 with the mean score of 6.07. It was also noticed that 8 students scored lower than 5. One of them even got a zero and one student got the full score. Remarkably, these extreme scores were not

observed in the peer evaluation especially when using the rubric.

Table 1: Scores from the Evaluation.

Subject	Peer Evaluation (Total = 10)		Teacher Evaluation (Total = 10)
	Using Rubric	Without Rubric	
R1	9.29	9.17	9.29
R2	9.29	9.17	8.21
R3	6.43	0.83	0.00
R4	5.71	3.33	6.07
R5	7.14	8.75	10.00
R6	8.57	8.33	8.93
R7	6.43	8.33	6.43
R8	6.79	4.79	2.86
R9	8.21	6.25	7.14
R10	9.64	7.50	7.14
R11	9.64	7.08	7.14
R12	7.14	5.00	5.36
R13	6.07	3.33	5.00
R14	8.57	5.83	6.43
R15	7.50	6.25	6.79
R16	6.79	6.25	6.43
R17	7.86	6.67	6.79
R18	6.43	2.50	3.57
R19	6.79	3.33	2.86
R20	8.57	5.42	6.43
R21	9.29	3.33	3.57
R22	6.43	4.17	2.86
R23	8.21	8.33	8.57
R24	5.71	2.50	3.57
R25	7.14	4.17	4.29
Mean score	7.59	5.63	6.07

To check the reliability of peer evaluation against teacher evaluation, we also ran the correlation value test. The correlation values were calculated from three pairs of scores and the results were as follows:

Table 2: Correlation Value

Pair of Scores	Correlation Value
1. Peer Evaluation with Rubric VS Peer Evaluation without Rubric	0.60
2. Peer Evaluation with Rubric VS Teacher Evaluation	0.53
3. Peer Evaluation without Rubric VS Teacher Evaluation	0.90

According to the table, the overall reliability of peer evaluation against teacher evaluation exists as all the correlation values are above 0.5, suggesting the substantial connection between the scores given by the subjects and those given by the teacher. What we found surprising, however, was the fact that the correlation value of the third pair was very high and that would in fact be worth further scrutiny to gain insight into what criteria the subjects were using and what factors were affecting them when they evaluated their peers' work.

6.2 Students' Perceptions on Peer Assessment

From the in-depth interview, the six subjects revealed that the concept of peer assessment was not new. Some of them had experienced it in their field. However, they were not familiar with the use of it in an English reading class. They said that in science and technology, their field of study, students' work or answers could be objectively evaluated as usually there were specific right or wrong answers. This divergent nature of the answers eases self- and peer evaluation. In a

reading class where comprehension and higher levels of cognition is targeted, however, the evaluation process is difficult and subjective even with the provision of rubrics. All of the subjects interviewed, therefore, stated that they were not certain whether their evaluation was valid. The process was also influenced by different factors as their own understanding of the excerpt, familiarity with their partner and the final grade or the score that they would obtain, as well as issues on 'trust' or beliefs on authority.

A. Personal understanding was a crucial reference that the subjects used in evaluating their peer. Even though the rubric was given and the subjects said that they understood the criteria used for peer evaluation, they were not certain how well their peer could perform as 'understanding' is a subjective construct. They explained that they were still not clear how well their friends could perform on each criterion. Therefore, they relied on their personal understanding and regard it as a benchmark to assess their peer, to see whether their answers were better, poorer, or in the same direction. Moreover, they expressed that they were not confident in their own ability to evaluate others, explaining that they were not even sure whether their understanding was 'right' so they would very much like the teacher to provide a definite answer key for them if peer evaluation was inevitable. All of them explicitly showed a clear preference for teacher evaluation. They regarded the teacher as a sage, an authoritative figure with an absolute power that they were not capable of sharing. This authoritative convention of teacher power is very common in Asian countries.

B. The intervention of Thai culture is clearly reflected in the subjects' evaluation process. The subjects were being compromising and extremely considerate. They did not want their peer to lose face or feel inferior and, thus, avoided giving negative comments or low scores. The problem was even more serious when pairing up with a senior. Their considerate nature deterred them from labeling 'strongly disagree' and they became more 'neutral' on the rating scale. The subjects also explained that the presence of the peer while evaluating influenced their evaluation. Noticeably, when they evaluated their peer using the rubric, they worked in pairs, so they did not dare to give their partner a low score. When doing the second evaluation, they focused only on the peer's worksheet and evaluated it without any peer pressure. This helped explain why the mean score of peer evaluation using the rubric was much higher than the second evaluation without the rubric. The subjects further suggested that if peer evaluation were to be used, they would prefer to do it anonymously.

C. Students' concern about scores and grades also strongly influenced their peer evaluation. The subjects revealed that they were concerned about getting a low grade and this might affect their application for future careers or higher education. Although the teacher tried to emphasise that they had better focus on learning rather than grades, they could not help worrying about getting low scores from the peers. Therefore, they compromised using a barter system, exchanging scores and expecting a good score in return.

Moreover, issues on 'trust' seemed to obstruct the subjects from evaluating their peers objectively. The subjects reported that they doubted whether their friends had any bias while evaluating, did not know their peer's background, and were not sure about their language proficiency and evaluation skills. Therefore, their friends might not be eligible to evaluate their work. They, instead, wholeheartedly trusted teacher evaluation.

7. Pedagogical Implications

7.1 Directions for Using Rubrics

Being asked about the possibility of using peer evaluation in the Thai context, five out of six interviewed subjects agreed that at present it would not be appropriate. They explained that main reason was because of the cultural background. Thai students get used to teacher evaluation and the change to this new evaluation paradigm of peer teaching requires more preparation and involvement from different sectors. It has to be done for the whole education system, at all levels, not just a single course. Only one student said that it should be used and could be used even with a single course as he/she realized the importance of active roles of learners in the whole learning process.

In fact, every subject agreed that peer evaluation using rubrics enhanced active roles of learners and it helped prepare them for the real world. Peer evaluation allows them to exercise their judgement and assessment skills which are important for work and they need to rely more on themselves. Rubrics, on the other hand, give them solid, clear-cut criteria to support their decision. However, in order to use the rubric more efficiently and effectively, we recommend that training is required and that teachers permit their students to negotiate the criteria and descriptors. Apart from showing teachers' willingness to make assessment accessible to their students, this can allow them to set learning goals to improve their learning from the early stage.

7.2 Directions for Applying Assessment for Learning

In accordance with the interview that we conducted with the six subjects, all of them agreed that making use of formative assessment had its advantages. While summative assessment is conducted at the end of learning, formative assessment is done consistently throughout the process of learning and it allows the students' active involvement. Our students enjoyed being part of assessing their peers' and their own learning. Also, the oral feedback they received from their friends was not as threatening as the teacher feedback. Above all, formative assessment had quite a positive impact on their motivation and self-esteem.

Hence, it is recommended for language teachers in Thailand and other Confucius-based countries to introduce formative assessment despite cultural baggage concerning teachers' supremacy and students' main concern over good grades. The implementation of formative assessment will result in a lot of challenges as it involves a lot of changes. Teachers need to take into account ways in which they can assert formative assessment in the scenario where summative assessment takes precedence. As Sahlberg (2012) proposes, the first thing teachers can do to create a change in assessment is to stop over-emphasising their teaching to the test which narrows the curriculum and lowers educational quality and standards.

8. Conclusion

Thailand, like other Asian countries, has heavily relied on summative assessment to provide measures and standards of learners' achievement. As long as the myth of good graders having more opportunities in life still perpetuates, teachers will still be regarded as grade providers, rather than knowledge providers. We strongly believe that this traditional, conservative idea of teaching and learning needs to be changed. Giving our students the opportunity to assess themselves and their friends is the first step in engaging them in their own learning and goal setting. To ensure the accomplishment of peer assessment as the basis of further formative assessment, teachers should train their students, provide necessary support such as solid assessment criteria and monitor them closely before permitting them to do it on their own. In spite of foreseen challenges and difficulties that may arise from strong cultural interventions and tradition of teaching and learning practice, we have to start the change for our students' sake. It is necessary to point out to them that the constant learning process is much more important and valuable in the long run than the alphabetical symbols they are given at the end of the course. In suggesting this, we are aware that it will take a lot of time before formative assessment is widely accepted and practised.

References

Airasian, P.W. (2008). *Classroom assessment: Concepts and applications* (4th edn.). Boston: McGraw-Hill.
Allen, M. J. (2010). The use of rubrics for assessment, grading, and encouraging student learning, paper presented at Atlantic Assessment Conference. April, 11, 2010.
Andrade, H. G. (2005). Teaching with rubrics, the good, the bad, and the ugly, *College Teaching, 53* (1), 27-30.
Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society, *Studies in Continuing Education, 22* (2), 151-167.
Carless, D. (2007). Learning-oriented assessment: conceptual bases and practical implications. *Innovations in Education and Teaching International, 44* (1), 57-66.
Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings.* New York: Routledge.
Hafner, J. C. & Hafner. P. M. (2003). Quantitative analysis of the rubrics as an assessment tool: An empirical study of student-peer group rating, *International Journal of Science Education, 25* (12), 1509-28.
Liu, N.F. & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279-290.
Min, H. & Xiuwen, Y. (2001). Educational Assessment in China: Lessons from history and future prospects, *Assessment in Education: Principles, Policy & Practice, 8*(1), 5-10.
Ng, S.M. (2008). Formative assessment and effective teaching and learning. In Cane, G. (Ed.) *Strategies in language learning and teaching.* (pp. 19-54). SEAMEO Regional Language Centre: Singapore.
Reddy, Y. M. & Andrade. H. 2010. A review of rubric use in higher education, *Assessment and Evaluation in Higher Education, 35*(4), 435-448.
Sahlberg, P. (2012). Finland: A non-competitive education for competitive economy. In OECD: *Strong performers and successful reformers—Lessons from PISA for Japan.* (pp.93-111). Paris: OECD.

Appendix A

Rubric for peer-assessment: *Cocktail*

Instructions: Discuss the questions about the play *Cocktail* with your friend. Then look at the table below and assess your friend accordingly.
Please add any extra comments that may be useful to him/her.
Peer's name _____Peer assessor's name _____

Criteria	strongly agree	agree	disagree	Strongly disagree
1. Your friend understands the plot of the story.				
2. Your friend can give logical explanations about the central conflict of the story.				
3. Your friend sees how the main character has developed from the beginning to the end of the story.				
4. Your friend can detect the inauthenticity of the play.				
5. Your friend recognises certain stereotypical portrayals of Thailand and its people through the characters.				
6. Your friend appreciates/understands a moral of the story.				
7. Overall, your friend understands the play, the characters and the writer's intention well.				