

Self-evaluation Strategies and Medical Students: Opportunities for Critical Thinking in a CLIL-based Classroom

Chad L. Godfrey, Saitama Medical University

ABSTRACT

John Hattie's (2012) meta-analysis on educational research is highly significant. One of Hattie's most important findings related to self-grading/ self-assessment. Based on his study, self-reporting can influence student achievement by an effect size of 1.44 years. By providing students with opportunities to create their own scoring criteria, which can be used to evaluate and reflect on their classwork, students are more critically engaged. In the end, they can have accountability for their own learning and potentially have richer learning outcomes.

In 2018-2019, students in a Japanese EFL classroom at Saitama Medical University (SMU) were involved with a leadership initiative to re-evaluate/re-author the presentation scoring benchmarks used in their CLIL classroom. After several sessions of brainstorming, discussion and drafting, students produced the criteria used to score both their peers as well as themselves. Through the process of creating and utilizing the scoring criteria, students' developed ownership of the evaluation criteria. This educator could witness a richer involvement in analyzing and achieving goals, and a greater awareness of the higher standards students had set for their own work. In this study, I will share the reasoning for giving students ownership in creating their own evaluation, and the outcomes of their efforts. I believe that any educator could benefit from this information, as it has implications for promoting achievement and critical thinking in any learning context.

KEYWORDS: self-evaluation, peer-evaluation, critical thinking

INTRODUCTION

Traditionally, when educators hear the word assessment, they may imagine an end-of-term exam, used to assess a student's achievement and to calculate the final scores for a subject; in other words, a summative assessment of assessment of learning (AoL). Basically, assessment can fall into the two groupings that have distinct purposes: summative and formative. As defined in Davies (2010), "Summative assessment – evaluation – comes at the end of learning, while formative assessment provides information and support during the learning." It goes without saying, both forms of assessment have their importance in the classroom, but are they both equally utilized by classroom teachers?

Japan is a country which has a rich testing culture and high-stakes exams that are used to advance students to the best universities. Although both formative and summative assessment are used, "summative testing still remains dominant in Japan, as it does in all Confucian-heritage cultures" (Carless, 2011, in Wicking, 2014). However, as an EFL educator in Japan, assessment *for* learning (AfL) has had a strong role in assessing my students. I believe this goes hand-in-hand with the nature of language education; how it is necessary to monitor progress and give feedback to students when they are using and/or acquiring language. As summarized by the OECD (2013), "while summative assessment and reporting remain essential at key stages of the education process, formative assessment frameworks tend to shift attention away from excessive focus on numerical marks, labelling and ranking of students, in order to focus on learning processes and individual progress."

Working with first year Japanese medical students in a CLIL classroom at SMU, I wished to explore two educational objectives for the 2018-2019 school year simultaneously: Allowing students to be involved in creating and carrying out assessment and to have students think more deeply about their learning. Based on past experiences, I recognized that there can be a lack of ownership for learning by students. Getting students to interact with one another and their teachers about the class content, and to think critically about their learning, are major hurdles to cross. Along these two points, I considered what would be the best ways to make learning meaningful for these students in the short time we will work together? By focusing AfL and offering

opportunities for deeper learning, a study was conducted to investigate better learning possibilities in a first-year CLIL classroom.

RATIONALE

John Hattie's book, *Visual Learning* (2008), has implications for all educators. Hattie's conducted a meta-study examining 50,000 research articles, and offered a list of influences on student achievement. Using the metadata from these studies, a numerical score was calculated (called an effect size). This numerical score is based on the learning impact of an item, with 0.40 being seen as significant. Some of the top items of the 2018 list included: (1.33) self-reported grades/student expectations; (0.75) evaluation and reflection; and (0.70) feedback. After first reading about this research in 2016, I carried out a small study on self/peer assessment in my CLIL classroom of first-year Japanese medical students in 2016-2017 (see Godfrey, 2018). The results of this study were promising. Therefore, I conducted an updated study in 2018-2019, expanding on what was explored in the previous study.

This present study set out to deepen my understanding of the following guiding questions: What are the students' past experiences in terms of using self- or peer-assessment? How can students think critically about presentations and create an assessment tool to be used by all members of the classroom? What will motivate students to move from a teacher-centered approach of assessment to a more student-centered approach? How will self-assessment and peer-assessment influence their learning outcomes?

STUDY

Assessment creation

In 2016-2017, I conducted a brief study where students explored creating assessment criteria to be used for presentations (see Godfrey, 2018). However, although the current study reexamined some of the past studies successes, it was the first year-long project, with several new elements added, including:

- Pre- and post-surveys about assessment and presentations
- More time was set aside to practice scoring using the student-authored criteria and to discuss its value
- Th reanalysis of criteria mid-year and adjusting certain scoring criteria

Twenty-four first year students completed a survey about learning preferences, assessment and presentations during the first class. This information was utilized to get a clearer picture of the students' background in using and creating assessment, as well as attitudes towards assessing others and self-assessment (see Table 1).

Evaluation	05/23/18		
	NO	YES	No Answer
1. I have scored my own work in class.	7	17	0
2. I have scored other students work in class.	13	11	0
3. I have made original scoresheets to score class activities.	19	5	0
4. I believe only teachers should score student work.	12	11	1

Table 1. Questionnaire class data about assessment

As for presentations, several questions were asked in regard to group vs solo presentations, the use of PowerPoint in their presentations, memorizing speeches, and their general attitude about presenting in English. Student answers were recorded numerically on a scale from 1-5. (see Table 2).

Presentation (scale of 1-5)	05/23/18
1. I like making solo presentations.	3.2
2. I like making team presentations.	3.7
3. I like using PowerPoint in my presentations.	3.5
4. I like memorizing my speech before my presentation.	3.0
5. I like making a presentation in English.	2.8

Table 2. Averaged questionnaire data about presentations

In the third week of the first term, students were placed in groups of two or three members and asked to brainstorm about the following two questions:

- (1) What is needed in the written part of a presentation?
- (2) What makes a presentation interesting? What doesn't make a presentation interesting?

Within the group, each person has a role. One student was a timekeeper, another was the note taker, and the final student was the leader/facilitator for the group. If a group had fewer than 3 members, then one person needed to perform two roles. Before each new question was introduced, the roles would shift, so each student had the chance to perform the roles at least once.

After each question, the brainstormed answers were shared on the whiteboard by the groups. Then, the groups would study the whiteboard results and update their list with any new ideas they thought were relevant. If any group's ideas seemed unclear, the group would be asked to clarify the meaning of their answer. The brainstorming's final question was the following:

- (3) Looking back at questions 1 and 2, which parts of a presentation do you think should be scored?

These final results were submitted to the classroom teacher on paper. In the end, I looked over all their whiteboard answers for questions 1 and 2, as well as their final answers for question 3. In the following week's lesson, a draft of the scoring criteria was shared with each group for them discuss and edit as needed. Afterward, the groups assigned points to each item of the scoring criteria. The rule was that the total amount of each score sheet criterion had to add up to one hundred points, and no criterion could have a score lower than five points. I averaged their answers, and in the following week, the groups looked over the final draft of the score sheet and voiced their opinions if there was something they wanted to adjust (see Appendix 1).

Classroom Practice

Following the creation of the scoresheet, students shared two short presentations during the second and third terms. One was a pair presentation focusing on a medical or health topic, while the other was a solo presentation about a health issue in Japan. Before students shared either presentation, they participated in several practice trails to test parts of the scoring criteria and their understanding on how to use it. The first trail focused on only one point of the physical message (eye contact) and later trails centered on combined points of the presentation's physical message (eye contact, posture, and gestures). Follow-up class discussion assisted students in sharing their views on scoring, attempting to target more objective scoring methods, as opposed to either giving completely perfect or completely poor scores to a student's effort. The

final trials included assessing a full practice presentation by classmates. This was the first time for students to focus on both peer-assessment and self-assessment as well as all of the assessment criteria on the scoresheet. In general, students found it challenging to peer assess, as there were several different kinds of points to assess. Following the practice presentation, students conferenced with their assessed pair, giving feedback in English or Japanese. As for self-assessment, students were often harsher in their own assessment compared to assessing a partner's work. However, by the second presentation, this tendency began to shift as they became more confident in using the scoring criteria, and more confident in their own abilities as well.

DISCUSSION

Students are concerned about how the teacher will evaluate them in the classroom. In higher education in Japan, evaluation standards may focus on attendance and the term exam(s), leaving students uninterested in other aspects of a lesson. However, what happens when students are critically involved with classroom assessment, shifting to having a more active role in the classroom?

Both self- and peer-assessment have important outcomes. As summarized by Sydney School of Education and Social Work (2019), the main aims of self- and peer-assessment are to:

- increase student responsibility and autonomy
- strive for a more advanced and deeper understanding of the subject matter, skills and processes
- lift the role and status of the student from passive learner to active learner and assessor (this also encourages a deeper approach to learning)
- involve students in critical reflection
- develop in students a better understanding of their own subjectivity and judgement.

From my experiences, students have accurate views about what is important to assess in terms of presentations and presentation skills, but they can lack responsibility in completing a finished presentation for class. Often, I find students busy preparing their work up to the moment of the presentation. I also find students only accept the final judgment of their work given by their teacher, with weakened sense of investment

reaching the goals outlined by their teacher.

Creating the assessment criteria was in itself a form of deeper learning. Students needed to think critically about what they believed was important in presentations, and be able to express these ideas with their classmates. The exercise itself was a chance to explore content critically and to take a greater role as an active participant in the classroom. By giving students the responsibility to author assessment, a greater sense of autonomy was fostered in my classroom.

But, based on the pre-questionnaire, for many of the students, this was the first time that they had made scoring criteria, and for nearly half of the students, it was the first time they assessed another classmate. Furthermore, some felt that the role of assessment resided solely with the classroom teacher. Some students needed time to move comfortably from a very teacher-centered approach to a more student-centered one.

After the criteria was completed, the students needed guidance in evaluating and putting the scoresheet into practice. Over time, as students began to display greater confidence, they were more active and offered deeper advice to their peers, internalizing the criteria's standards. This was also included their self-reflection skills. As Tanner (2011) states, "Once learners have practiced checking others people's work, they will be better able to look at their own work with the eyes of an outsider." Tanner further added, "Peer and self-assessment gives learners tools to monitor their own progress, which helps make them more independent." These points were true for my students, but I discovered with this study that for peer- and self-assessment to work successfully, time was needed. Several class sessions were needed to brainstorm, discuss, write and rewrite the criteria. Expanding the time needed for the students to become comfortable with understanding the criteria and using it for feedback was paramount.

However, the time investment resulted in positive outcomes. Most students by the end of the course had internalized the scoring criteria, and as a direct result, improved the quality of their work and the work of others. As stated by Althaus & Darnall, 2001; Tsai, Lin, & Yuan, 2002, as cited in Mogessie (2015), "It is also claimed that those students that provide their peers with high quality feedback tend to incorporate feedback from their peers effectively, raising their final grades in the process."

Students could think critically throughout this study because they were given the opportunity to reflect on their own learning. As stated in Nadri & Azhar (2016), "Indeed, in order for students to be able to monitor, assess, and improve their own performances

and their own thinking, they need to self-reflect regularly. Metacognitive reflection allows students to manage and assess their own thinking strategies.”

LIMITATIONS & CONCLUSION

Although there were many positive outcomes from this study, there were several limitations that need to be considered. This practice worked well with a small class of students, but working with a class of sixty or more may be more difficult to manage. In addition to this, some students needed the time and support to take peer- or self-scoring seriously. Some students were very modest, marking the quality of their work much lower than it actually was, and on the opposite end, there were those who thought their work was superior when it was average. Furthermore, time is a problem. Despite the positive learning outcomes, the entire project was time consuming, and I wished I had time to conference/negotiate with each student over their scores and my scores. Also, having numerical scores was difficult for some students. A simpler form of evaluation or a rubric may have helped students be more objective with their scores. Finally, peer- and self-assessment did not reduce my workload as you may imagine, but it did give me a clearer picture of my students’ efforts and opinions of their classwork.

The same questionnaire that was given on the first class was again given in the last class (see Table 3). Making solo presentations, using PPT and just a general liking for English presentations all increased for this class, whereas team presentations and memorizing a speech decreased. Even though students were initially fearful of the idea of solo presentations at the beginning of the year, working as a team was took more effort and planning for the students. Also, the content of their presentations was medically-based. Memorizing a speech was more of a challenge for some students depending on their English level.

Presentation (scale of 1-5)	05/23/18	01/09/19
1. I like making solo presentations.	3.2	3.6
2. I like making team presentations.	3.7	3.3
3. I like using PowerPoint in my presentations.	3.5	4.0
4. I like memorizing my speech before my presentation.	3.0	2.9
5. I like making a presentation in English.	2.8	3.6

Table 3. Questionnaire data about presentations (first and last classes)

Despite the any of the challenges of creating scoring criteria in the classroom, the results of this study had a great effect on student learning and achievement. On a final note, students continued to assess themselves after their work was finished, making the scoring criteria what I considered a living set of learning objective that their carried with them after my class was over.

REFERENCES

Chan C. (2010). *Assessment: Self and Peer Assessment*, Assessment Resources@HKU, University of Hong Kong. <http://ar.cetl.hku.hk> Accessed: June, 30th, 2019.

Dale, L., Van der Es, W. & Tanner, R. (2011) *CLIL Skills*. Haarlem: European Platform, Internationalising Education, 2011.

Godfrey, C. (2018). *Medical Students' Evaluation Strategies in the CLIL Classroom*. Journal of Medical English Education, Vol.17, No.1, p.22-23.

Hattie, J. (2012). *Visible Learning for Teachers*. Routledge Publishers.

Luoma, S. (2004). *Assessing Speaking Skills*. Cambridge: Cambridge University Press. https://www.academia.edu/36242047/Sari_Luoma-Assessing_Speaking_Cambridge_Language_Assessment_2004_.pdf Accessed: June, 30th, 2019

Mogessie, Michael. (2015). *Peer-assessment in higher education – twenty-first century practices, challenges and the way forward*. Assessment & Evaluation in Higher Education. https://www.researchgate.net/publication/283212439_Peer-assessment_in_higher_education_-_twenty-first_century_practices_challenges_and_the_way_forward. Accessed: July 1, 2019.

Nadri, Y. & Azhar, A. (2016). *Self-Assessment of Critical Thinking Skills in EFL Writing Courses at the University Level: Reconsideration of the Critical Thinking Construct*. pp57-71. Arab World Journal, December 2016. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2895540. Accessed: July 1, 2019.

OECD Reviews of Evaluation and Assessment in Education. <https://doi.org/10.1787/22230955> . Accessed: July 1, 2019.

https://sydney.edu.au/education_social_work/groupwork/docs/SelfPeerAssessment.pdf

Accessed: June, 30th, 2019

Wicking, P. (2016). *The Role of Formative Assessment in Global Human Resource Development*. JALT Journal, Vol. 38, No. 1, May 2016. <https://jalt-publications.org/files/pdf-article/jj2016a-art2.pdf> .Accessed: June, 30th, 2019

APPENDIX 1

<input type="checkbox"/> Peer <input type="checkbox"/> Self <input type="checkbox"/> Teacher		Title:	
		Name:	
Written Message	Introduction	/	+greeting
	Body	10	+examples
	Conclusion	12	+summary
Visual Message	PPT Appearance	/	(includes fonts, text size, amount of text, color choice, etc.)
	English Usage	11	(correct spelling, grammar and punctuation, plain English usage, etc.)
	Use of Visuals & Media	9	(use of pictures, graphs, videos, music, etc.)
Physical Message	Eye Contact	/	
	Voice	9	+ clear, loud and with good tempo
	Posture & Gestures	11	
	Q/A	/	
		8	
		7	
Total Points = 100pts		/	
		100	