A reflection on the use of self-assessments in Japanese universities
Nicholas Carr

Abstract

Many Japanese universities run courses designed with the primary goal of improving students' oral proficiency. In line with the recent studies which highlight that language learning is not a linear process (Nunan, 2001), the importance of formative and sustainable assessment rather than a single summative type assessment has been emphasized in previous research (Brown & Hudson, 1998; Everhard, 2015; Kissling & O'Donnell, 2015). This short paper discusses the issues that utilizing self-assessments in a Japanese university English class raised, with a focus on how this type of assessment influenced not only testing itself, but also pedagogy and the whole language learning process. The potential advantages and limitations of self-assessments are discussed. This is followed by a reflection on how self-assessments were implemented in a second year Reading and Discussion class for English major students at Gifu Shotoku Gakuen University.

Assessment in self-assessment

When considering self-assessments, the term assessment itself requires to be defined in order to be distinguished from more traditional notions of testing. Everhard (2015) describes testing as an instrument used at the end of a course, or learning period, to measure how much a learner can reproduce under exam type conditions. Everhard contends that our understanding of assessment needs to shift away from the traditional notion of "testing" and move towards a notion of assessment being part of the learning process. Based on this, assessments can be broadly categorised as: summative, formative and sustainable assessment. Everhard provides "working" definitions of these as follows:

Summative assessment – determines how much a student has learnt over a set period of time.
Formative assessment – assesses what needs to be learnt. Teacher and students make decisions together about possible learning pathways based on the assessments. In addition to Everhard's definition, I argue that this also includes an assessment of what has been learnt.
Sustainable assessment – assessment is part of learning. Students participate in on-going activities that enable them to become autonomous learners.

As Everhard (2015) points out, these categories should not be thought of as completely discrete, but rather they are to be viewed on a continuum. Due to self-assessments requiring learners to assess their own language use (Brown & Hudson, 1998) , it can be argued they fall into the sustainable assessment area of the continuum. Therefore, when in the realm of self-assessments, the term assessment should be more thought of as part of the learning process rather than a measure of what a learner produces under certain conditions.

nickca5@yahoo.com.au

Brown and Hudson (1998) provide three sub-categories of self-assessments:

Performance self-assessment – the learner evaluates how well they would perform in a situation.
Comprehension self-assessment – the learner assesses how well they comprehended a situation.
Observed self-assessment – learners view a recording of themselves in a situation, such as a role play, and evaluate their performance. (In the absence of a "recording", an observed self-assessment could be achieved based on reflection rather than viewing a video or listening to audio).

Potential advantages of self-assessment

A common theme throughout the literature that advocates the use of self-assessments is the emphasis of the positive washback effects of such assessment instruments. Everhard (2015) argues that summative assessments have the washback effect of reducing 'learner-centeredness' in the classroom and push language teachers away from the communicative language teaching paradigm and back towards a more traditional paradigm. Nunan (as cited by Little, 2005, p.321) goes to on claim that if a class is learner-centered, then it requires learners to be involved in assessment. Therefore, it can be surmised that involving learners in assessment tasks – such as the use of self-assessments – the learner-centeredness of the classroom increases.

One of the goals of a learner-centered classroom is for learners to become active agents in their own learning processes and set their own goals. Kissling and O'Donnell (2015) argue that in order for this to be achieved, learners require an understanding of what successful communication entails. They contend that self-assessment is a vehicle that can help students gain this understanding – thus assisting learners to become active in their own learning processes. Another potential positive washback effect of self-assessments is that they shift the responsibility of learning back onto the learners (Everhard, 2015; Kvale, 2007; Little, 2005) and increase learner autonomy (Brown & Hudson, 1998). For many TESOL educators, especially in EFL contexts such as Japan, learner autonomy is critical due to class time not being enough to enable progress in students' second language development – with autonomous learners having more motivation and spending more time outside of class developing their language skills (Harmer, 2007). The uprise of SALCs (Self-Access Learning Centers) in Japan is further evidence of the recognition that class-time only is often not sufficient exposure to English and of the need for learners to become more autonomous. It has also been argued that self-assessments increase learners' language awareness, which Marsh (as cited in Kissling & O'Donnell, 2015, p.284) argues in turn increases students' knowledge of how language can be used to achieve communicative goals.

Self-assessment has been presumed beneficial due to promoting self-regulated learning and autonomy (Oscarson, 1989). Goto-Butler and Lee (2010) argue that self-assessments are beneficial because they increase learners' awareness of their own learning and performance, which in turn enables students to become more proficient learners. Goto-Butler and Lee also contend that self-assessments enable learners

to understand how much assistance is needed to reach their individual goals, and develop and employ strategies that will enable them to achieve these goals. It has also been argued that this self-reflection gives students a sense of control over their learning, which in turn increases motivation (Paris & Paris, 2001). Finally, and I argue most importantly, self-assessment promotes learning long after the class ends (Oscarson, 1989). For language learners this is paramount, as the skill of evaluating how they performed in any communicative situation, what they can learn from it and how to improve on it next time is essential for second language learning to continue after formal training ends. It is proposed that as a learner moves away from a dependence on external assessment and moves towards independent assessment, the learning becomes more profound (Von Wright 1993, as cited by Everhard, 2015, p.20). It could also be argued that a learner's ability to provide feedback for themselves would also act as a recovery device when negotiating meaning between interlocutors, i.e. as a type of self-correction.

Potential limitations of self-assessment

Self-assessments are often rejected due to a lack of validity and reliability (Everhard, 2015). Blanche (1988) found that students' accuracy of assessment differed depending on the skills required to complete the assessment and the materials used during evaluation. Blanche also highlighted that subjective errors often influenced scores due to external factors such as career objectives and parental expectations. Therefore, even if construct validity is high in a test design, a potential disadvantage of using self-assessments is students compromising validity due to a misunderstanding of how to assess the task or other external factors.

Self-assessments also raise issues regarding reliability. Studies have found that highly proficient learners tend to under evaluate themselves (Blanche, 1998). Cultural and social factors can also cause some learners to underrate themselves in self-assessments due to the cultural desire to save face by appearing humble (Matsuno, 2009).

Another area of concern for some regarding the reliability of self-assessments is how strongly they correlate with the teacher's assessment of the same task. If a strong correlation is found, then it could be argued that self-assessments could be used in place of, at least to some degree, teacher assessments. However, many studies have found there to weak correlation coefficients between teacher assessments and self-assessments (Matsuno, 2009; Wilkes, 1995).

Educational and cultural background also has the potential to cause self-assessments to be viewed negatively – and thus have the disadvantage of negatively influencing learners. Sengupta (1998) found English second language learners in Hong Kong to be unable to see any of the claimed benefits of self-assessments. Sengupta argued the biggest obstacle for students was the perceived notion that only native or near-native level speakers of English were able to judge a student's work and that the only "real" reader of a written text is the teacher. Sengupta highlights that these perceptions are engrained in the educational system in Hong Kong. Therefore, a significant issue when considering the use of self-assessments is that,

even if a teacher believes in their benefits, in many TESOL educational settings breaking down these perceptions will require significant work by the teacher.

Reflection on implementation of self-assessments in a Reading and Discussion Class

The following section reflects upon the implementation of self-assessment tasks in two separate Reading and Discussion classes at Gifu Shotoku Gakuen University. The reflection is based on not only classroom observations but also a review of how the students' self-assessment scores correlated with the teacher's assessment – the course required students to be assessed on their participation in discussions on a weekly basis.

The primary goals of the classes were to develop speaking skills and to develop the ability to participate in discussions. Classes designed to improve oral proficiency present many challenges, one of which is the need to increase learners' accountability for the success of their learning (Kissling & O'Donnell, 2015). As outlined above, one of the benefits of self-assessments is the shift of responsibility of learning from the teacher back to the students. Accordingly, it was judged that self-assessments would offer many benefits for students participating in these Reading and Discussion classes. The classes were of a pre-intermediate and an intermediate level. Each unit of the course was completed over two 90 minute lessons, during which students learn vocabulary related to a specific topic, read a short text on the topic, learn helpful phrases related to the topic and learn how to respond to discussion questions.

The self-assessment tool's[1] criteria was limited to three categories: fluency, vocabulary and task focus. The rationale behind just three criteria was ease of use, i.e. too many criterions may result in students finding the act of self-assessing burdensome or unenjoyable. The criteria were kept broad for the purpose of being able to be used for a variety of speaking tasks. The tool's design allowed students to track the progress of one criterion on multiple occasions, with the aim of assisting students to gain confidence due to visual evidence of improvements in their performance and feel an increase of responsibility in their learning.

In line with the argument that students benefit from repeated self-assessments (Kissling & O'Donnell, 2015), students were given the opportunity to assess themselves on a weekly basis. To relieve any angst students may have felt, all self-assessed tasks were practiced at least twice before being "assessed" – with students being reminded that they would reflect on their performance before performing the task for the purpose of self-assessment. After completing the task for the final time, students were reminded how to use the criteria to assess their performance. Students were also aware that, as per course requirements, the teacher was officially assessing them simultaneously on a weekly basis.

During the semester, it was evident that the self-assessments provided several positive washback effects but also had limitations. The interpretation of the first criterion – fluency – varied considerably

---

[1] See Appendix 1 for a copy of the self-assessment tool

among students and often had a weak correlation with the teacher's assessment of the same task. When considering fluency, it seems that students had trouble conceptualising micro skills such as flow, pauses, reformations and self-corrections. It can be argued that this was largely due to students requiring more knowledge of these micro skills to be able to more effectively assess themselves. Upon reflection, I argue that for the whole semester the self-assessments would have benefited from assessing only one of the micro skills of fluency. This would have developed a deeper understanding of one of the facets of fluency but also most likely increased the accuracy of students' assessment of this criterion. In turn, this would have potentially created more positive washback effects, such as increasing students' language awareness, which then helps students to understand how language can be used to achieve their communicative goals.

The second criterion of the self-assessment tool – vocabulary – was consistently successfully assessed by students and had a strong correlation with the teacher's assessment. Due to each unit having key vocabulary that was introduced, students had something concrete to measure their performance by – namely, to judge if they had been able to successfully incorporate the target words and phrases when the opportunity arose. The benefits of students having something concrete to latch on to when assessing themselves has been found in other self-assessment studies (Kissling & O'Donnell, 2015).

The task focus criterion was also easily understood by students and, more often than not, had a strong correlation with the teacher's assessment. Whilst this criterion was not visible like the vocabulary criterion, students clearly understood that this was a judgement on whether or not they felt they had stayed on task and actively participated in the discussion. This seemed to have significant positive washback effects, with students feeling an increased sense of responsibility to ensure they fully participated in the tasks and maximise the learning opportunities for both themselves and their peers.

Higher achieving students of each class often rated themselves lower than students with lower proficiency. One reason for this could be that the higher achieving students had a higher level of language awareness, which enabled them to notice the gap between their current level and their language goals. This notion is further substantiated by the intermediate class self-assessments having a stronger correlation with the teacher's assessment than the pre-intermediate class.

Despite the limitations, I still argue that self-assessment in these Reading and Discussion classes is justified due to the increased learner-centeredness and learner autonomy it initiated. Many students who developed the ability to self-assess also displayed an increase in self-efficacy, which was displayed in the learning techniques utilized in class during vocabulary practice activities. If self-assessments were to be utilized again in future semesters, I would recommend the following to maximise the learning opportunities:

● The fluency criterion be broken down into one or more of its micro skills to enable a clearer understanding of the criterion

● Students to show their score to another classmate and explain the rationale behind their self-assessed score

● Students to be consistently reminded that progress in oral proficiency takes time and that it may take more than one semester for them to see improvement

● For some tasks to be audio or video recorded so that self-assessment does not always rely on students' memory of the task

● Due to higher achieving students rating themselves lower than lower proficient students, self-assessments should not make up any significant, if any, part of official grades (they did not count towards official grades in this case)

Conclusions

I contend that it is difficult to argue against the literature that highlights the benefits self-assessments in the language classroom. However, as both the literature and the reflection presented in this paper illustrates, self-assessment tasks do have potential limitations with both reliability and validity. Furthermore, for many students the idea of taking responsibility for their own learning appeared foreign and a period of adjustment is required. Despite this, the significant influence any type of assessment has on the whole learning process cannot be ignored, and neither can the advantages of a learner-centered classroom – which is clearly best served by involving learners in assessment.

As Everhard (2015) argues, the issue at hand is not whether or not students can perform the assessments while overcoming the aforementioned problems. Rather the point is the process of doing these assessments is learning itself and that this process is more important than the potential disadvantages. Consequently, this reflective paper is not arguing that we need to find a better way to assess, but that teachers should be more concerned with the influence testing has on the whole teaching process.

Finally, I argue that perhaps language teachers need to consider that different types of assessment play different roles. As Devenney (1989) argues, in certain circumstances reliability and validity may not be the primary concern. In other words, the learning benefits outweigh potential validity and reliability concerns. Practical solutions around this could be, as suggested above, self-assessments not counting towards official grades, or a degree of teacher supervision be included in the self-assessments.
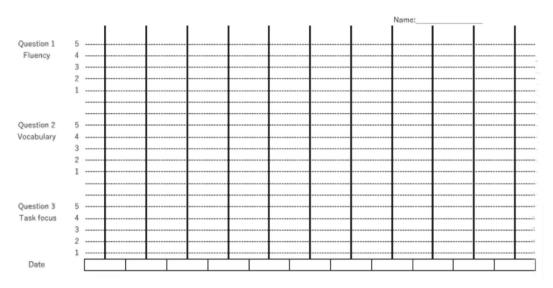
References

Blanche, P. (1988). Self-assessment of foreign language skills: Implications for teachers and researchers. *RELC Journal,* 19(1), 75-83. doi: 10.1177/003368828801900105

Brown, J., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly,* 32(4), 653-675. Retrieved from http://www.jstor.org/stable/3587999

Devenney, R. (1989) How ESL teachers and peers evaluate and respond to student writing. *RECL Journal: A journal of language teaching and research,* 20(1), 77-90. doi: 10.1177/003368828902000106

Everhard, C. (2015). The assessment-autonomy relationship. In C. Everhard & L.

Murphy (Eds.), *Assessment and autonomy in language learning* (pp.8-34). [Adobe digital editions]. Retrieved from http://deakin.eblib.com.au/patron/FullRecord.aspx?p=2006602

Goto Butler, Y., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing,* 27(1), 5-31. doi: 10.1177/0265532209346370

Harmer, J. (2007). *The practice of English language teaching* (4th ed.). Essex: Pearson Education Limited.

Kissling, E., & O'Donnell, M. (2015) Increasing language awareness and self-efficacy of FL students using self-assessment and the ACTFL proficiency guidelines. *Language Awareness,* 24(4), 283-302, doi:10.1080/09658416.2015.1099659

Kvale, S. (2007). Contradictions of assessment for learning in institutions of higher learning. In D. Boud &

N. Falchikov (Eds.), *Rethinking assessment in higher education* (pp.57-71). Retrieved from http://deakin.eblib.com.au/patron/FullRecord.aspx?p=292915

Little, D. (2005). The common European framework and the European language portfolio: involving learners and their judgements in the assessment process. *Language Testing,* 22(3), 321-336. doi: 10.1191/0265532205lt311oa

Matsuno, S. (2009). Self-, peer-, and teacher assessments in Japanese university EFL writing classrooms. *Language Testing,* 26(1), 75-100. doi: 10.1177/0265532208097337

Nunan, D. (2001). Teaching grammar in context. In C. Candlin & N. Mercer (Eds.),   English language teaching in its social context (pp. 191–199). Abingdon, Oxon: Routledge

Oscarson, M. (1989). Self-assessment of language proficiency: rationale and applications. *Language Testing,* 6(1), 1-13. doi: 10.1177/026553228900600103

Paris, S., Paris, A. (2001). Classroom applications of research on self-regulated learning. Educational Psychologist, 36(2), 89–101.

Sengupta, S. (1998). Peer evaluation ʻI am not the teacherʼ. *ELT Journal,* 52(1), 19-28. Retrieved from http://eltj.oxfordjournals.org.ezproxy-b.deakin.edu.au/content/52/1/19

Wilkes, M. (1995). Learning pathways and the assessment process: A replication of a study on self- and peer- assessment I the ESL classroom. In G Brindley (Ed.), *Language assessment in action* (pp.307-321). Retrieved from http://apps.deakin.edu.au/ereadings/equella/download/unitcode/ECL776_TRI-2_2015/item/711e2763-7825-6e6b-5844-33bb7a6eaae4/version/1/attachment/scan-learningpathways-wilkes-1995.pdf.

Name:_____

Question 1 Fluency	5										
	4										
	3										
	2										
	1										
Question 2 Vocabulary	5										
	4										
	3										
	2										
	1										
Question 3 Task focus	5										
	4										
	3										
	2										
	1										
Date											

Appendix 1