Hale *Language Testing in Asia* (2015) 5:1
DOI 10.1186/s40468-014-0010-0

Language Testing in Asia
a SpringerOpen Journal

RESEARCH Open Access

Self-assessment as academic community building: a study from a Japanese liberal arts university

Chris C Hale

Correspondence: cch2@nyu.edu
American Language Institute, Tokyo Center, NYU School of Professional Studies, 2-15-1 Konan, Minato-ku 108-6022, Tokyo

Abstract

Student self-assessment has been heralded as a way of increasing student ownership of the learning process, enhancing metacognative awareness of their learning progress as well as promoting learner autonomy. In a university setting, where a major aim is to promote critical thinking and attentiveness to one's responsibility in an academic community, the temptation to implement self-assessment is undeniable. In the liberal arts university context, the activity seems to speak directly to the values and belief systems inherent to the liberal arts ideals, not the least of which is the preparation of young minds in the pursuit of recognizing and expanding their intellectual abilities, and then applying that knowledge to solving the world's problems for the betterment of humankind. As self-assessment on high-stakes testing is uncommon in East Asia, where the education system is still heavily teacher-centered and controlled (particularly in relation to grading and assessment), there is added interest and novelty in incorporating an assessment mechanism in which the students have a direct voice in their own grading. The purpose of this study was to to determine what value, if any, students placed in the practice as members of a university academic community. Following the general inductive approach to qualitative research, in which "the findings arise directly from the analysis of the raw data, not from a priori expectations or models" (AJE 27:237-246, 2006), student responses to open-ended questions related to their perception of the value of self-assessment were coded and analyzed. Qualitative findings indicate that students appear to find value in the exercise consistent with the overall aims of a liberal arts education, indicating that it makes them feel *trusted* by instructors, develops their sense of *responsibility* and forces them to look at their writing more *objectively*. These findings suggest that there is more value to self-assessment on high-stakes essay tests than simply achieving inter-rater reliability with expert teacher raters.

Keywords: Self-assessment; Higher education; Academic writing; East Asia; Japan

Background

Self-assessment: accuracy versus value

There are several inconsistencies in the literature regarding the utility of self-assessment with several authors praising the practice (Blanche & Merino, 1989; Gardner, 2000; McDonald & Boud, 2003; Patri, 2002), while others reporting serious concerns as to the reliability and validity of it (Blue, 1994; Huang, 2010; Matsuno, 2009; Oldfield & MacAlpine, 1995; Sullivan & Hall, 1997). One major perceived benefit of self-assessment is that it is assumed to enhance learners ability to self-monitor their

Springer

© 2015 Hale; licensee Springer. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

learning, which leads to autonomy (Paris & Paris, 2001). The inherent subjectivity of self-assessment has led to questioning of its validity as a measure of student learning, particularly in relation to foreign language proficiency, and several studies have looked at ways of validating the process through attempts to correlate self-assessment scores and scores of other assessments, such as course grades, tests and teacher or expert ratings (Butler & Lee, 2010). Ross (1998), looking at second-language assessment, performed a meta-analysis of validation studies and concluded that accuracy in self-assessment depended on the skill being evaluated, and found that writing was among the more difficult skills to self-assess reliably. In addition to variation of the skill under assessment, other researchers have found that other factors can influence the accuracy in self -assessment, such as the learners' proficiency level (Heilenman, 1990, Patri, 2002), second-language anxiety (MacIntyre et al. 1997), and motivation (Dornyei, 2001). Butler and Lee (2010) studied the effects of self-assessment on English language learners in South Korea and found that improvements in accuracy were only marginal over time, and that training in the process was critical, as well as providing sustained feedback to the self-assessors. One teacher in Sato's study reflected that if self-assessment were ever to be taken seriously by students and teachers, then it needed to be associated with formal grading, rather than remain an ungraded exercise.

Perhaps the most controversial aspect of self-assessment is that students appear under-qualified to accurately assess their own learning when compared to expert (teacher) raters.

Matsuno (2009) for instance, looking at Japanese learners on writing tests, used multifaceted Rasch measurement to find that self-assessors consistently under-rated themselves in comparison to teacher raters, while rating their peers more highly. It was speculated that this was perhaps a result of their Japanese cultural conditioning to appear individually modest while reverential to peers. Matsuno therefore conclude that self-assessment was less accurate and therefore less valuable than other assessments, such as teacher- and peer-assessment.

While there appears to be much evidence that student and teacher assessment rarely show high correlation, there is nonetheless a substantial body of literature promoting self-assessment for other reasons. Bedore and O'Sullivan (2011), for instance, discuss the importance of "removing the instructor from the position of sole authority" (p. 13) while Blanche and Merino (1989), in reviewing the literature on self-assessment, point out a number of studies addressing the increased learner motivation associated with including students in their own assessment. Similarly, Harris (1997) and Gardner (2000) highlight the relationship between self-assessment and increased learner motivation and autonomy. Sadler (1989) suggests that learners can move beyond becoming consumers of education, and places them at the center of their learning. He encapsulates the pedagogical rationale for incorporating self-assessment in the formal grading process by emphasizing the metacognitive benefits inherent to the practice, while emphasizing the increased sense of community it promotes:

> Providing guided but direct and authentic evaluative experience for students enables them to develop *their* evaluative knowledge, thereby bringing them within the guild of people who are able to determine quality using multiple criteria. It also enables transfer of some of the responsibility for making decisions from teacher to learner.

> In this way, students are gradually exposed to the full set of criteria and the rules for using them and so build up a body of evaluative knowledge. (p. 135)

Further adding importance to self-assessment, Hattie (2008) conducted an extensive quantitative study of over 100 factors that influence student learning (from teacher quality to curriculum design) and found that the single most salient indicator of student learning was their ability to accurately self-assign grades. Even with such evidence of the efficacy of the practice, the reality is that while "teachers embrace the theoretical promise of self-assessment, few devote much time to its practice" (Hilgers et al. 2000, p. 9) indicating that, perhaps, the lack of accuracy in student-assessment outweighs the other pedagogical benefits.

However, Boud (1990, 2000) reminds us that assessment in higher education is often at odds with the purported values espoused by universities, and that self-assessment is one way to encourage critical thinking and responsibility—traits that will serve students well in their lives after graduation: "Assessment therefore needs to be seen as an indispensable accompaniment to lifelong learning. This means that it has to move from the exclusive domain of assessors into the hands of learners" (Boud, 2000, p. 151). Boud (1990) laments the lack of student participation in decision-making at universities claiming that there is an "unhealthy dominance of a situation where staff are always both an authority and in authority. The challenge is to find a place for significant student responsibility in this context" (p. 106). While all assessments need to in some way lead to learning, self-assessment of academic writing in particular can actually raise students' metacognative awareness of their own capabilities in ways that teacher-only assessment cannot.

Raider-Roth (2005), while not explicitly addressing self-assessment, makes a compelling case for including students in decision-making processes that impact their lives. Like other student-centered theorists such as John Dewey (1903), who likewise advocated for the promotion of democratic environments even within the formal institutional settings of schools, Raider-Roth (2005) implores educators to provide conditions where students feel safe enough to "challenge teachers' authority" (p. 34). Such environments allow students to be "dangerous and to take risks, to voice that which had not been said" before (p. 34). Empowering students, it would seem, can have a lasting influence on their lives long after they leave the college classroom. However, such a philosophy is not prominent in the Confucian-based education models of East Asia, including Japan (Marginson, 2011). Therefore, introducing student-liberating pedagogies to East Asian contexts may be met with confusion or resistance, at least initially, even at a Japanese liberal arts university based on the western model that promotes developing "adventurous minds capable of critical thinking and sensitivity to questions of meaning and value" (*Citation removed for blind review.*)

Much of the research in self-assessment uses student self-reporting on reflective-type diagnostic questionnaires (such as a series of "can do" statements) in order to ascertain perceived student competence (e.g. Blanche & Merino, 1989; Harris, 1997; Heilenman, 1990). These questionnaire responses are then compared to teacher evaluations of the students. The fundamental problem with such a system is that it is far too broad. It attempts to address overall student competence, often across skill areas, using different rating rubrics (e.g. a formalized scoring rubric for the teachers that is used in grading,

and a reflective one for the students that is not.) Butler & Lee (2010), did not look at graded self-assessments, while Matsuno (2009), who did in fact look at self-assessment on graded tasks, used a "simplified" version of a teacher-rating rubric because it was believed that students were too inexperienced to effectively apply the same rubric teachers use. This two-tier system of assessment may indicate to students that there is one grading procedure for teachers that is "real," and another one for students that is not. The current study overcomes this limitation in that both student self-raters and expert-raters utilized the exact same points-based grading rubric in order to assign scores to four timed essay tests over the course of two ten-week terms. Reflective questionnaires here were used to elicit student reactions to the self-assessment process, not to self-assess their own perceived writing proficiency.

Method

This study involved advanced-level second language learners of English in the assessment of four high-stakes writing tests over two terms: two mid-terms and two finals. Combined, these assessments accounted for ten-percent of students' final course grade per term.

Participants

Participants in the study were freshman university students at a selective, private university in Japan. The university followed the American liberal arts model, in that students did not declare their major until their junior year, and much of the curriculum consisted of humanities courses. All freshman students at the university enroll in a semi-intensive English language program designed to prepare them for university content courses taught in English, as well as to acclimate them to a western, liberal-arts style of education very much unlike the traditional, exam-oriented traditional education they experienced in their secondary schools. The 100 students were divided into four classes of approximately 20 students each. Students' English proficiency level was considered "advanced" according to placement score results prior to entering the university. Their average paper-based TOEFL Test score (Test of English as a Foreign Language) was 580. Participants took two terms (ten-weeks each) of semi-intensive English language courses totaling five contact hours per-week. The course from which the data was collected was an academic writing course totaling three contact hours per-week, per-term. The course consisted of presenting academic writing as a genre, placing emphasis on strong thesis statement creation, evaluation of sources, and logical essay development. Students produced two fully referenced essays per term, and took two timed essay tests in which students were given a prompt and required to take a position and argue that position making reference to course readings as well as their own original supporting examples. Prior to this course, students had no experience with academic writing of this nature (in English or Japanese). The participants were approximately 65% female.

Self-assessment procedure

Prior assessment procedures consisted of two teacher-raters blindly rating each essay test on a 15-point scale allotting 10 possible points for writing content, and the other 5

for mechanics (see Table 1). Teachers were expected to be within two points in their assessment, and the scores were then summed for a total possible score of 30. When teachers were three or more points apart, a third teacher-rater was called and the two closest scores (within two points) were kept. Prior to assessing the tests, teachers engaged in a "norming session" with all other teacher-raters in which several example essays were presented and discussed. Teachers discussed their ratings for the tests and negotiated in order to fall within the necessary two-point range. When consistency was reached, teachers each took 40 tests to rate individually, placing their scores on a separate paper. They then exchanged the tests with another teacher to rate who could not see the first rater's score. After completing their rating, the second rater gave the tests to a testing coordinator who examined the tests to determine how many third ratings were needed. Historically, the inter-rater reliability among teacher raters has been 80% or greater, which is widely considered the acceptable minimum for achieving inter-rater reliability on written assessments.

For this research project, the second rater was substituted by the actual student who wrote the essay test. All other aspects of the procedures remained the same. Prior to students assessing their own tests, they conducted a norming session with their classroom teacher on selected tests the teachers viewed as beingexemplars of exceptionally well-written test responses (after securing permission from the students who produced the tests and removing their names). This norming session lasted for one class period, and the students assessed their own tests at the end of the class session. This norming session also included a rationale for self-assessment given by the classroom teacher which included the notion of an "academic community" and reiteration of the principles of a college education (Boud 1990, 2000, Boud and Brew 1995). The four essay-test prompts were based on course readings where students were expected to apply concepts from the readings to a real-world problem (see Table 1 for example essay prompt).

Table 1 Test prompt and scoring rubric

Quality of understanding & response	Writing	Total
9-10 Strong understanding of core concepts, and strong, well-developed support for your opinion.	**4-5** Strong paragraph unity and coherence with topic sentences (including key words from prompt) and transitions.	
7-8 Satisfactory understanding of core concepts, with good development and support for your opinion.	**3** Minor problems with paragraph unity or coherence such as topic sentences or transitions or minor grammar or word choice errors.	
5-6 Basic understanding of core concepts and basic development and support for your opinion.	**2** Some problems with paragraph unity or coherence such as topic sentences or transitions or minor grammar or word choice errors.	
3-4 Only partial understanding of core concepts and much more development and support needed for your opinion.	**1** Major problems with paragraph unity and coherence, lack of topic sentences or transitions and many grammar or word choice errors.	
0-2 Little or no understanding of core concepts with little or no development or support for your opinion.	**0** Little or no paragraph unity or coherence and pervasive grammar or word choice errors.	
First Rater (Teacher) /10	First Rater (Teacher) /5	/15
Second Rater (Student) /10	Second Rater (Student) /5	/15
Third Rater (If needed) /10	Third Rater (If needed) /5	/15
Final Score /20	**Final Score /10**	**/30**

Prompt: What ethical theory do you think could best be applied to guide decision making for an issue related to either race or gender?

Data collection

Data collection followed an inductive qualitative design (Creswell, 2009; Thomas, 2006). Following the fourth treatment, students were asked to take an online, anonymous questionnaire designed to elicit reactions to and impressions of the self-assessment procedure (see Figure 1). Of the 100 students in the study, 85 responded to the survey (85% response rate). Questionnaire items contained both closed (Likert-type) and open-ended questions where students could write as much as they wished about a particular prompt. The focus of this study is the responses to the open-ended questions asking, 1) "What was your overall impression of self-assessment?" And 2) "Did you find any value in it?" Data collection followed the general inductive approach to qualitative research, that is, "although the findings are influenced by the evaluation objectives or questions outlined by the researcher, the findings arise directly from the analysis of the raw data, not from a priori expectations or models" (Thomas, 2006, p. 269). Responses were analyzed and coded inductively (Thomas, 2006) using the qualitative coding software HyperRESEARCH version 3.0. Themes were then generated from the initial codes and these themes were used to inform the findings (Saldaña, 2009). It was hoped that responses to the questionnaire would add insights into how students perceived the activity and whether it should be continued, altered, or abandoned (see Figure 1).

Research question

The following research question framed the analysis of data: How do students at one Japanese university perceive the effectiveness of self-assessment on high-stakes writing tests?

Results

Inductive coding of student statements on the open-ended questions asking them to share their impressions of the self-assessment activity revealed three salient themes: 1) Objectivity, 2) Responsibility and 3) Trust.

Objectivity

The most prevalent codes in the data transcripts were terms associated with the notion of *objectivity.* Students indicated that participating in their own assessment led to a greater objective awareness of their own writing in relation to expert raters and the grading rubric. Patri (2002) makes reference to the importance of objectivity in student ratings by stressing that when "learners are placed in a situation where they can access information on the quality and level of their own performances, or those of their peers, then it is possible to clarify their own understandings of the assessment criteria and, more importantly, what is required of them" (p. 111). Students made comments such as, "It was a good experience for me because it enabled me to participate in grading and evaluate my answer objectively."

Some students made explicit reference to aligning their assessments with the expectations of teachers: "With self-assessment, I could see what the teachers are looking for in the essay. Of course, there are rubrics for the essays and teachers say what they are looking for in the test, but with self-assessment I can see more precisely about what is required." Another wrote how he or she could "improve my answers on next test or

The questionnaire consisted of the following items, with 9 and 10 comprising the open-ended questions from which the data was compiled:

1. What was your feeling about being a rater on your own exam?
    a. Liked it
    b. Did not like it
    c. No opinion
2. Do you think the resulting grade was a fair assessment of your learning?
    a. Yes
    b. No
    c. Not sure
3. Do you agree that part of a student's responsibility in an academic community is to be involved in assessment (grading)?
    a. Yes
    b. No
    c. Not sure
4. What do you think would be the most valuable in determining grades?
    a. Self-assessment
    b. Peer-assessment
    c. Combination of self and peer-assessment
    d. Teachers should determine grades alone
5. Would you like the opportunity to take part in your own assessment (grading) in the future?
    a. Yes
    b. No
    c. Not sure
    d. Other (please explain)
6. The assessment practice (norming session) with sample essay tests was helpful (rate from 1-5, where 5 is very helpful).
7. The explanation of the rationale for self-assessment in a liberal arts community by the teacher was helpful (rate from 1-5, where 5 is very helpful).
8. Consulting with my classmates and teacher on the criteria prior to assessing my test was helpful (rate from 1-5, where 5 is very helpful).
9. Give your overall impression of the self-assessment activity, including ideas for improving it, or reasons for ending it. Please explain.
10. Did you find value in participating in self-assessment? Please explain.

Figure 1 Questionnaire items. The questionnaire consisted of the following items, with 9 and 10 comprising the open-ended questions from which the data was compiled.

essays because I know what the teachers are asking me to do in the test." Another concurred, saying that through self-assessment, "students will be made to consider how teachers would grade their works and that would lead students to see their works from teachers' point of view objectively."

Several students made connections between enhanced objectivity and improving their writing for future tests. One student claimed, "I saw my essays from a different point of view and could have the time to determine what exactly I had done wrong and use the knowledge for my next essay. I felt that my essays became more logical and critical." Another referred to this metacognitive awareness-raising by saying that "this self-scoring activity is useful to know my own ability and to not make the same mistakes on the next test."

Still other students made reference to the applicability of enhanced objectivity to other areas of their lives. They highlighted the wider applicability to the skills acquired through self-assessment making statements such as these:

> I think the self-assessment activity plays a valuable role in my life after I graduate. In real life, no one would provide us any assessment. We have to objectively figure out what is being asked, what the problem is, and how can we fix it. That kind of practice should be done in university and self-assessment is one of the best ways to create that skill.

And:

> From the Self-Assessment activity, I learned to critically read my own paper from the reader's point of view. This technique can be applied to anything from daily life activities to academic activities. Self-Assessment activity allows students to look at themselves from different perspectives and spot flaws and weakness.

Responsibility

In line with Harris (1997), who stated that "self-assessment can help to make learners more active, to realize that they have the ultimate responsibility for learning" (p, 13), and Boud (1990) who claimed that the "common goal of higher education [is] that students should become autonomous learners who can take responsibility for their learning" (p. 104), the second most salient theme to emerge from the transcript data was that of *responsibility.* Students continually made comments that indicated they were cognizant of the added responsibility they were given as graders of their own work. This statement by one student was representative of many: "I think it was a good idea to grade our own exams because by doing that, I felt I was totally responsible for my work."

Many students made direct reference to teachers in their comments relating to responsibility. Some believed that the act of handing in a test or assignment acted as a cognitive ending to the assignment, whereas under the self-assessment process, they acutely felt the added responsibility required of them. One stated: "Just depending on the grades by teachers makes me feel like it's not my own exam." Another concurred saying that with self-assessment, students "won't finish the test and say, 'Oh that's it I have nothing to worry about now- it's up to teachers' but instead think back on what

they've written." Another also made reference to a lack of finality: "Checking my own paper made me realize that the learning process doesn't just stop when we finish our [tests] and hand them in." One student stated that the experience had profound effects on how she or he viewed grading:

> Self-assessment completely changed my attitude toward my assignments. I used to consider my work was done when I had submitted the assignment. However, I realized that being involved in the grading process, and understanding the reason for why the grade I was given are absolutely my responsibility.

Another student said that, after participating in the grading process, he or she found it quite natural as students are most familiar with their own work and their abilities:

> It was surprising to know that we had to grade ourselves, but when I took part in it, I felt it was quite reasonable. Since you know how much effort you put in and the ability of your English writing skills, you actually are the one who knows what is the suitable grade for your piece of work. Also, you are not grading it by yourself so you won't always put a high score for your work. Knowing that your teacher will also give a grade, you are likely to be more honest, and you would be more responsible for what you have handed in.

Trust

Battistich et al. (1997) found that "mutual trust seems to be characteristic of schools that are felt to be communities" and that a "sense of community among students was strongly correlated with student achievement and inductive reasoning skill" (p. 143). Likewise, Raider-Roth (2005) advocates for educators to create "an environment where trust can prevail" (p. 30) and cautions that for students "to develop trustworthy knowledge, they must learn in the context of trustworthy relationships" (p. 18). In other words, for students to trust their own knowledge as well as their teachers, they need to feel trusted by their teachers to be ratified members of their learning community.

The students in this study made several references to the notion of *trust*. One student echoed many saying that "the self-assessment activity made a lot of sense after acknowledging our responsibilities in our academic community. I felt trusted by the teachers, knowing that they were letting us grade our own tests." Another expressed her or his feelings of mental confusion at first, as she or he had never been a part of assessment before: "I was confused because this was my first experience with self-assessment. However, I was happy to know teachers trust us. Having responsibility for my grade made me study harder for the next exam." This comment also underscores the fact that when students feel trusted by their teachers, it has a concomitant effect on their sense of responsibility.

This sense of trust was not limited only to feeling trusted by teachers, but also to feelings of some students trusting their own judgment. One student highlights this, at first being skeptical of the notion that students can be trusted to assess themselves, yet comes to the conclusion that students are in fact trustworthy:

> The idea of self-assessment was totally new to me and I was suspicious if it works effectively because I did not trust in students' ability to justly assess themselves. Now

> I think differently because it did work successfully. I was happy when I realized that my grading sense became a little closer to teachers each time.

Some students made comments indicating that the activity raised their awareness of the importance of feeling trusted, and began to wonder why other areas of the university did not employ a similar system. This student made reference to the discrepancy between feeling trusted in one area of the university (the courses where self-assessment was employed), and not others:

> Overall I understood the reason why self-assessment is important and why we shouldn't just have the teachers grade our essay tests. Then it made me wonder about the other tests. If this way of marking tests is going to be used in English class, could other tests besides English change as well? Actually I thought they should because they are all within this academic community. Otherwise it would be weird to have this way of rating in only one area of study.

Discussion

The themes found in the transcript data fit well with the academic environment some faculty at this Japanese liberal arts university are trying to create: one in which students feel trusted and respected enough to be part of the formal assessment process, and responsible enough to look at their writing objectively. Clearly, as some students indicated, this feature needs to be expanded to other areas of the university community, though convincing disparate departments, many still following the tradition of teacher-centered, knowledge transference, will be a challenge.

Unlike Matsuno (2009) who found self-assessment to be of "limited utility as part of formal assessment" (p. 1), it is posited here, on the contrary, that the "utility" lies not in a student's ability to assess themselves with great accuracy, but rather in the concomitant socio- and metacognative benefits they experience as a result of the activity. Though accuracy is of course important in the longer term, and having the narrow two-point inter-rater threshold acts as a check for student-initiated "grade inflation." In addition, unlike previous research into self-assessment, the students in this study tended to *over-rate* themselves. Further research needs to be conducted to determine if the high-stakes nature of the assessments influenced how students self-assessed their performance.

In a university setting, where critical thought and self-exploration are fundamental educative outcomes, self-assessment of essay tests fits perfectly within the ethos of higher-order thinking skills the university attempts to foster in its students. Some student comments indicated that they seemed to suffer "cognitive dissonance" (Song et al. 2007) when initially asked to participate in the formal assessment process, and their reactions were somewhat reluctant. However, it is reasonable to expect student "buy in" to increase with further self-assessment tasks. Anecdotally, this view is supported by a Likert-type prompt on the questionnaire asking if students believed their resulting grades to be "a fair assessment of their learning" which garnered 65% agreement after the first assessment (with many students answering "I'm not sure"), and 85% by the fourth. This suggests that while some students are still more comfortable with teachers being the sole assessors, the longer students are exposed to self-assessment, the stronger the "buy-in" becomes.

Perhaps most important to educators considering implementing self-assessment is making the case for it even if the quantitative results initially indicate it is not particularly effective in terms of student-teacher evaluation agreement. In fact, it is quite unreasonable to expect students to perform on par with trained raters after only limited experience, and sustained exposure to the practice, including training, is paramount to success (McDonald & Boud, 2003). Within the context of a university, engaging students in their academic community cannot be more enhanced than by inviting them to contribute to perhaps the most important artifact associated with their academic work: grades. Being situated in East Asia, where teacher-centered learning is still the overwhelming norm, making the case for student participation will not be without challenges. If, however, the aim of university educators is to promote student self-awareness and inclusion in a democratic community of practice, then showing students are respected and trusted enough to be a part of their own learning assessment can only enhance these aims. In fact, these aims are not inconsistent with the broader aims of higher education in Japan; rather, self-assessment on high-stakes testing is, as of yet, an under-utilized means for achieving them.

Conclusions

As with any study, there are certainly limitations in the present study beyond its limited scope. In particular is the fact that student autonomy was not fully respected (in that students were not party to the creation of the criteria for the assessment, and instead followed a teacher-created rating rubric). However, students first need to be aware of what constitutes proficiency in the domain before they can be expected to recognize deficiencies in it. This is especially true of academic, argumentative writing in English, a genre of writing with which students coming from the Japanese secondary school system are not familiar. When students have internalized the particular features embedded in the genre, it is conceivable that they could be given more autonomy in devising criteria for assessing it (including assessing logical essay development, critical thinking, and quality of supporting claims). This is certainly an area for further research, particularly in relation to high-stakes testing. In addition, quantitative analysis of intra-rater reliability, inter-rater reliability and correlation with expert raters should be evaluated in order to determine if the student perceptions of the task matched with their actual ratings. While this study has shown that students could value the process of self-assessment, accuracy is undeniable important as well, as the preponderance of research in this area focuses on this aspect. Though the focus of this study was on the perceptions of students as members of an academic community, further research is needed to determine the extent to which the value participants found in the process correlates with their accuracy as raters.

As Boud implores us, the learning aims of higher education need to be represented in the assessment mechanisms we employ (Boud & Brew, 1995; Boud, 1990, 2000). The teaching of academic writing is unique in that it can help students recognize flawed arguments, deficiencies in logic, as well as enhance students' own critical thinking development—all important to productive, engaged members of a democratic society. As the findings of this limited study indicate, when students are invited to participate in the formal, summative assessment of their own writing, the metacognitive gains they experience are directly applicable to life beyond the course and even the university. In

this way, the summative assessment process leads to formative assessment, that is, it becomes another important learning opportunity for students—one with lasting implications.

Competing interests

The authors declare that they have no competing interests.

Received: 2 October 2014 Accepted: 30 November 2014
Published online: 03 February 2015

References

Battistich, V, Solomon, D, Watson, M, & Schaps, E. (1997). Caring school communities. *Educational Psychologist, 32*(3), 137–151.
Bedore, P, & O'Sullivan, B. (2011). Addressing instructor ambivalence about peer review and self-assessment. *Writing Program Administration, 34*(2), 11–26.
Blanche, P, & Merino, BJ. (1989). Self-assessment of foreign-language skills: implications for teachers and researchers. *Language Learning, 39*(3), 313–338.
Blue, GM. (1994). Self-assessment of foreign language skills: does it work? *CLE Working Papers, 3*, 20.
Boud, D. (1990). Assessment and the promotion of academic values. *Studies in Higher Education, 15*(1), 101–111.
Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151–167.
Boud, D, & Brew, A. (1995). *Enhancing Learning Through Self Assessment*. London: Kogan Page.
Butler, YG, & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing, 27*(1), 5–31.
Creswell, J. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks California: Sage Publications, Inc.
Dewey, J. (1903). Democracy in education. *The Elementary School Teacher, 4*(4), 193–204.
Dornyei, Z. (2001). *Motivational Strategies in the Language Classroom*. Cambridge: Cambridge University Press.
Gardner, D. (2000). Self-assessment for autonomous language learners. *Links & Letters (Universitat Autónoma de Barcelona), 7*, 49–60.
Harris, M. (1997). Self-assessment of language learning in formal settings. *ELT Journal, 51*(1), 12–20.
Hattie, J. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement* (1st ed.). New York: Routledge.
Heilenman, KL. (1990). Self-assessment of second language ability: the role of response effects. *Language Testing, 7*(2), 174–201.
Hilgers, TL, Hussey, EL, & Stitt-Bergh, M. (2000). *The Case for Prompted Self-Assessment in the Writing Classroom*. New Jersey: Hampton Press.
Huang, L-S. (2010). Seeing eye to eye? The academic writing needs of graduate and undergraduate students from students' and instructors' perspectives. *Language Teaching Research, 14*(4), 517–539.
MacIntyre, P, Noels, K, & Clément, R. (1997). Biases in self-ratings of second language proficiency: the role of language anxiety. *Language Learning, 47*(2), 265–287.
Marginson, S. (2011). The Confucian model of higher education in East Asia and Singapore. In S Kaur, E Sawir, & Marginson (Eds.), *Higher Education in the Asia-Pacific: Strategic Responses to Globalization* (1st ed., pp. 53–75). New York: Springer.
Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing, 26*(1), 075–100.
McDonald, B, & Boud, D. (2003). The impact of self-assessment on achievement: the effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice, 10*(2), 209–220.
Oldfield, KA, & MacAlpine, JMK. (1995). Peer and self-assessment at tertiary level–an experiential report. *Assessment and Evaluation in Higher Education, 20*(1), 125–132.
Paris, SG, & Paris, AH. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*(2), 89–101.
Patri, M. (2002). The influence of peer feedback on self-and peer-assessment of oral skills. *Language Testing, 19*(2), 109–131.
Raider-Roth, M. (2005). *Trusting What you Know: The High Stakes of Classroom Relationships*. San Francisco: Jossey-Bass.
Ross, S. (1998). Self-assessment in second language testing: a meta-analysis and analysis of experimental factors. *Language Testing, 15*(1), 1–19.
Sadler, RD. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144.
Saldaña, J. (2009). *The Coding Manual for Qualitative Researchers*. Thousand Oaks California: Sage Publications Ltd.
Song, L, Hannafin, MJ, & Hill, JR. (2007). Reconciling beliefs and practices in teaching and learning. *Educational Technology Research and Development, 55*(1), 27–50.
Sullivan, K, & Hall, C. (1997). Introducing students to self-assessment. *Assessment & Evaluation in Higher Education, 22*(3), 289–305.
Thomas, DR. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation, 27*(2), 237–246.