The Assessment Beliefs and Practices of English Teachers in Japanese Universities

Paul WICKING
Nagoya University Graduate School of International Development

Abstract

Any discussion about English education in Japan is invariably bound up with a discussion about assessment. All too often, such discussions have focused on entrance examinations and other high stakes summative tests. However, language testing and assessment do not take place in a vacuum, but are deeply affected by broader social and cultural contexts, as well as individual features of the school and the classroom. In particular, perhaps more than any other factor, it is the teachers who have great influence over how tests are created, conducted and interpreted. In order for universities to foster an environment in which reliable assessment can take place, it is not enough to provide well-designed, rigorous tests. It is also imperative that teachers’ beliefs and practices concerning assessment be taken into account. This paper seeks to answer the question, what are the assessment beliefs and practices of EFL teachers working in Japanese universities? To answer this question, survey responses were gathered from English language teachers working in Japanese higher education. The results indicated some slight differences in belief and practice between native speaker teachers and non-native speaker teachers, as well as between full-time and part-time teachers. Despite these differences, it seems that most teachers have a learning-oriented approach to assessment. The findings provide support and direction for policy-makers and educational leaders seeking to promote better testing practice.

Keywords: assessment, teacher belief, teacher practice, university EFL

The processes of teaching, learning and assessment are deeply embedded in the social, cultural and historical contexts in which they occur. Teachers, students, and other people involved in assessment make meaning of their experiences according to social norms and personal beliefs that have been conditioned by the surrounding culture. The practice of assessment does not occur in a sterile classroom bubble, with teachers, students and curriculum interacting together independently of the wider social, cultural and political contexts. Rather, the sociocultural context and the individual histories of the teachers and learners exert a profound influence on how assessment is carried out (Elwood & Murphy, 2015; Inbar-Lourie, 2008); and, in turn, the ramifications of that assessment extend into complex social, economic and philosophical spheres (Ross, 2008). A logical consequence is
that assessment will be enacted differently in Japan as compared with how it is enacted in other settings. However, while there has been much research in Japan into language test design and implementation, there has been relatively little research into the surrounding context within which this testing takes place. The aim of this paper is to focus upon one aspect of that context: the beliefs and practices of English language teachers working in Japanese higher education.

**Literature Review**

The assessment belief of teachers working in Canada, Beijing and Hong Kong higher education contexts was investigated by Rogers, Cheng and Hu (2007). They used a survey questionnaire to measure beliefs about assessment in a number of dimensions, including: assessment procedures, the value of assessment, the time needed to engage in assessment, and the place of standardized testing. Overall, the results were quite disparate, as “the beliefs expressed by the instructors in all three contexts were somewhat mixed, uncertain, and, at times, contradictory” (Rogers et al., 2007, p. 52). Research with teachers in Kuwait also found a significant gap between teachers’ philosophy and practice concerning assessment (Troudi, Coombe, & Al-Hamly, 2009).

Within Japan, conflicting views regarding testing have been found concerning the issue of the university entrance examination system. Cook (2013a) found that expatriate teachers, many of whom have completed their training in integrationist academic cultures, believe that the primary purpose of entrance examinations should be pedagogical. However, when faced with the system in Japan, they find that “examinations serve a much wider range of purposes” (Cook, 2013a, p. 14). These purposes include local economic, social and political purposes, as well as showcasing a university’s status and indirectly testing a student’s diligence and intelligence.

Concerning the purposes of assessment in general, a comparative study of English teachers in Japan and Japanese teachers in the Philippines was undertaken by Gonzales and Aliponga (2012). Assessment purposes were categorized into three general domains: (a) *assessment for learning*, (b) *assessment as learning*, and (c) *assessment of learning*. These three domains are roughly equivalent to the formative, metacognitive and summative purposes of assessment. Results from the study suggested that teachers from both countries most preferred assessment practices that were aligned with the notion of *assessment as learning*. The results also indicated that English teachers in Japan were more concerned with *assessment of learning* than the teachers in the Philippines. Gonzales and Aliponga suggested that this may be due to the strong influence that standardized language examinations exert in Japanese society. Their study provides some helpful insights into the general trends surrounding the assessment preferences of teachers, but the small sample size and lack of qualitative data mean that there is still much more to be explored. In particular, it was noted, “From the respondents’ viewpoints, there are some conditions that act as barriers to put their preferences into action and practice” (Gonzales & Aliponga, 2012, p. 12). What exactly these
barriers are, and how teachers could attempt to overcome them, has been suggested by Cook (2013b). She found that teachers believed beneficial change was inhibited by hierarchical structures, face, fear, concerns about continuity and the fact that many professors simply don’t see the need for change.

This resistance to change is not just a feature of Japanese teachers, but teachers everywhere. As Pajares (1992) has noted, beliefs in general are rooted deep within personal identities and conceptions of self, and so persist often in spite of conflicting evidence and new information. Teachers have formulated their educational beliefs based not only upon their teaching experience, but also upon their experiences as students, and so beliefs concerning education have been built up over many years from a very young age. Unlike other professions, teachers come to the classroom as ‘insiders’, in that they have knowledge and experience of how schooling is conducted. Pajares argued that these commitments to prior beliefs make changing conceptions and accommodating new information “nearly impossible” (p. 323).

However, although adjusting beliefs may be nearly impossible, it can still happen. The first step is to identify and shed light on those beliefs, to bring underlying and perhaps unconscious attitudes into the light of examination. As Freeman and Johnson (1998) wrote, “We believe that teachers must understand their own beliefs and knowledge about learning and teaching and be thoroughly aware of the certain impact of such knowledge and beliefs on their classrooms and the language learners in them” (p. 412). Although we may be able to construct tests that are valid and reliable and accurately measure declarative or procedural knowledge, that is not enough. The implementation of any new testing policy or practice must take into account the complex nature of teachers’ belief and practice if it is to be successful (Brown, 2004). It is hoped that the present study would make a contribution towards this end, in order to promote better testing and assessment practice.

Method

In order to investigate the assessment beliefs of teachers, a questionnaire survey was constructed, based loosely on Cheng, Rogers and Hu (2004) and Rogers et al. (2007). Part I of the survey asked teachers about their practices of assessment and evaluation. Part II of the survey consisted of 32 items addressing beliefs about assessment and evaluation. Participants were asked to indicate their agreement to each statement on a Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Part III of the questionnaire focused on demographic information. It asked participants about their gender, age, educational qualifications, teaching experience, current teaching conditions, and any courses they had completed in assessment and evaluation.

Once a draft of the survey was written, it was prepared in two formats: a paper-based version and an online version. Participants were chosen by two strategies: an exponential non-discriminative snowball sampling method and a stratified purposeful sampling method. Effort was made to ensure that non-native speaker (NNS) and native speaker (NS) teachers
were represented in the sample. From the 144 paper surveys that were distributed, 35 were
returned, making a return rate of 24%. The online survey was completed by 113 respondents.
Overall, 148 responses were successfully gathered from English as a foreign language (EFL)
teachers in Japan. Of the respondents, 73% were NS teachers and 27% were NNS teachers,
69% were male and 31% were female, and 69% were full-time teachers while 31% were
part-time. All returned responses were answered completely.

**Results**

The first part of the survey aimed to gather information about the purposes, methods,
and procedures used by teachers when assessing students. Altogether there were 31 items,
and respondents were asked to check all that apply. The most common purpose reported by
teachers for assessing students was to determine their final grades \( (n = 134) \), followed by,
respectively, to obtain information on students’ progress \( (n = 123) \) and to provide feedback to
students as they progress \( (n = 118) \). The least reported purpose was to prepare students for a
standardized test \( (n = 27) \).

Teachers reported using a variety of methods when assessing students. The most
common method was performance-based assessment tasks, which were reported to be used
more often than paper-and-pencil assessment tasks. The least common method of assessment
was found to be standardized tests. Less than half the respondents reported using student
peer-assessment and student self-assessment (see Table 1).

Table 1

<table>
<thead>
<tr>
<th>Methods of Assessing Students</th>
<th>‘Yes’ responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student peer assessment</td>
<td>58 (39.2%)</td>
</tr>
<tr>
<td>Student self-assessment</td>
<td>63 (42.6%)</td>
</tr>
<tr>
<td>Tests that are provided with the textbook</td>
<td>40 (27.0%)</td>
</tr>
<tr>
<td>Standardized tests</td>
<td>29 (19.6%)</td>
</tr>
<tr>
<td>Paper-and-pencil assessment tasks made by a teacher</td>
<td>106 (71.6%)</td>
</tr>
<tr>
<td>Performance-based assessment tasks made by a teacher</td>
<td>118 (79.7%)</td>
</tr>
</tbody>
</table>

When asked to indicate their primary source(s) of material for creating assessment
items and procedures, most teachers reported developing test items themselves. The next
most popular source of assessment material was that from published textbooks. Items from
mandated syllabuses and other published test items were reported as not commonly used (see
Table 2).

Pearson’s Chi-square test was used to determine whether the categorical variables of testing
practice were independent from the two categorical variables of English speaker status
(native speaker or non-native speaker) and teaching appointment (part-time or full-time). The
two important assumptions of the chi-square test, being independence and the absence of
expected values below 5 in the contingency table, were both met.

Overall, there were significant differences found between the way in which native-speaker teachers practice assessment and the way in which non-native speaker teachers practice assessment. 88% of NS teachers reported assessing students for the purpose of providing feedback to students as they progress, compared to 57.5% for NNS teachers, representing a significant difference ($\chi^2(1) = 16.76, p < .001, \phi = -.34$). Based on the odds ratio, the odds of assessing students for the purpose of providing feedback were 5.40 times higher for NS teachers. Another significant difference was found with the purpose of formally documenting growth in learning ($\chi^2(1) = 7.48, p < .01, \phi = -.23$), which was cited more frequently by NS teachers (41.7%) than NNS teachers (17.5%).

Table 2
Primary Sources of Test Items

<table>
<thead>
<tr>
<th>Primary sources of test items</th>
<th>‘Yes’ responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items developed by yourself</td>
<td>123 (83.1%)</td>
</tr>
<tr>
<td>Items prepared together with other teachers</td>
<td>36 (24.3%)</td>
</tr>
<tr>
<td>Items from published textbooks</td>
<td>68 (45.9%)</td>
</tr>
<tr>
<td>Items from mandated syllabuses / curricula</td>
<td>13 (8.8%)</td>
</tr>
<tr>
<td>Items found on the Internet</td>
<td>23 (15.5%)</td>
</tr>
<tr>
<td>Other published test items</td>
<td>17 (11.5%)</td>
</tr>
<tr>
<td>Other</td>
<td>10 (6.8%)</td>
</tr>
</tbody>
</table>

In keeping with Rogers et al. (2007), the items in Part II were divided into four clusters for analysis. These were: beliefs about enhancing instruction and student learning (11 items); beliefs about item formats, classroom assessment procedures, and time to prepare assessments (14 items); beliefs about standardized testing (4 items); and beliefs about instructor understanding of and preparation for assessment (3 items).

Table 3 shows respondents’ beliefs about instruction and learning. The responses indicated a positive attitude towards assessment, with the mean scores indicating a high level of agreement that assessment results are important for instruction ($M = 4.12$) and assessment tasks provide a valuable learning experience for students ($M = 4.04$). Teachers also reported that they have enough time to properly prepare assessment and feel confident in their assessment abilities.

As can be seen in Table 4, a high level of agreement was elicited for the statement, “I need a variety of assessment methods to assess my students” ($M = 4.31$). Respondents also indicated that students should be assessed on their actual abilities, rather than in comparison with their classmates ($M = 3.99$), and that there was more value for learning in having frequent, shorter assessments, rather than less frequent, longer assessments ($M = 3.97$).
Table 3

Beliefs About Enhancing Instruction and Student Learning (N = 148)

<table>
<thead>
<tr>
<th>Item</th>
<th>Content</th>
<th>Mean</th>
<th>SD</th>
<th>Skew</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Assessment results are important for instruction.</td>
<td>4.12</td>
<td>0.83</td>
<td>-0.87</td>
<td>0.81</td>
</tr>
<tr>
<td>2</td>
<td>Assessment tasks provide a valuable learning experience for students.</td>
<td>4.04</td>
<td>0.82</td>
<td>-0.96</td>
<td>1.61</td>
</tr>
<tr>
<td>22</td>
<td>Assessment makes my students work harder.</td>
<td>3.85</td>
<td>0.85</td>
<td>-0.91</td>
<td>1.12</td>
</tr>
<tr>
<td>5</td>
<td>Assessment improves EFL learning.</td>
<td>3.84</td>
<td>0.84</td>
<td>-0.51</td>
<td>-0.15</td>
</tr>
<tr>
<td>29</td>
<td>Assessment results have an important effect on the way students see themselves.</td>
<td>3.82</td>
<td>0.71</td>
<td>-0.79</td>
<td>1.70</td>
</tr>
<tr>
<td>16</td>
<td>Assessment helps me to focus my teaching.</td>
<td>3.66</td>
<td>0.80</td>
<td>-0.38</td>
<td>-0.19</td>
</tr>
<tr>
<td>25</td>
<td>The way I teach my classes is tied closely to my student assessment results.</td>
<td>3.56</td>
<td>0.88</td>
<td>-0.25</td>
<td>-0.32</td>
</tr>
<tr>
<td>31</td>
<td>Assessment results tell me how well I have taught my students.</td>
<td>3.44</td>
<td>0.95</td>
<td>-0.79</td>
<td>0.56</td>
</tr>
<tr>
<td>15</td>
<td>Almost all my students try to achieve their best when they are assessed.</td>
<td>3.41</td>
<td>0.93</td>
<td>-0.50</td>
<td>-0.35</td>
</tr>
<tr>
<td>27</td>
<td>Assessment creates competition among students.</td>
<td>2.93</td>
<td>0.87</td>
<td>0.13</td>
<td>-0.78</td>
</tr>
<tr>
<td>8</td>
<td>Students dislike being assessed.</td>
<td>2.84</td>
<td>1.02</td>
<td>0.33</td>
<td>-0.75</td>
</tr>
</tbody>
</table>

In order to measure how beliefs differ according to English status (native speaker vs. non-native speaker) and teaching appointment (part-time vs. full-time), the Mann-Whitney’s U test was used. This non-parametric test was employed in the analysis because the data failed the Kolmogorov-Smirnov test for normal distribution, which precluded the use of a t-test. Of the 32 questions which asked about assessment belief, 28 elicited responses in which no significant differences were found between native speakers and non-native speakers. The four responses which revealed a significant difference are shown in Table 5.
Table 4
Beliefs About Item Formats, Assessment Procedures, and Time to Prepare Assessments (N = 148)

<table>
<thead>
<tr>
<th>Item</th>
<th>Content</th>
<th>Mean</th>
<th>SD</th>
<th>Skew</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>I need a variety of assessment methods to assess my students.</td>
<td>4.31</td>
<td>0.81</td>
<td>-1.40</td>
<td>2.45</td>
</tr>
<tr>
<td>24</td>
<td>Students should be assessed based on what they can do, rather than based on how they compare with their peers.</td>
<td>3.99</td>
<td>0.95</td>
<td>-0.75</td>
<td>-0.26</td>
</tr>
<tr>
<td>10</td>
<td>More frequent shorter assessments are more effective in encouraging learning than less frequent longer assessments.</td>
<td>3.97</td>
<td>0.77</td>
<td>-0.58</td>
<td>0.73</td>
</tr>
<tr>
<td>30</td>
<td>Assessment methods that are similar to real life situations are better than paper-and-pencil procedures.</td>
<td>3.76</td>
<td>0.82</td>
<td>-0.44</td>
<td>0.59</td>
</tr>
<tr>
<td>13</td>
<td>I am happy about the quality of my assessment.</td>
<td>3.43</td>
<td>0.87</td>
<td>-0.65</td>
<td>-0.26</td>
</tr>
<tr>
<td>6</td>
<td>I assess my students more than other EFL teachers.</td>
<td>3.18</td>
<td>0.90</td>
<td>-0.02</td>
<td>-0.17</td>
</tr>
<tr>
<td>12</td>
<td>I have sufficient time to develop the methods I use to assess my students.</td>
<td>3.08</td>
<td>1.04</td>
<td>-0.09</td>
<td>-0.80</td>
</tr>
<tr>
<td>9</td>
<td>Speaking and listening assessments are better for assessing EFL learning than paper-and-pencil assessments.</td>
<td>2.95</td>
<td>0.95</td>
<td>0.11</td>
<td>-0.05</td>
</tr>
<tr>
<td>11</td>
<td>Paper-and-pencil assessments provide the primary basis for the grades I assign to my students.</td>
<td>2.94</td>
<td>1.27</td>
<td>-0.03</td>
<td>-1.18</td>
</tr>
<tr>
<td>3</td>
<td>Other commitments do not allow me sufficient time to properly prepare assessment.</td>
<td>2.80</td>
<td>1.24</td>
<td>-0.01</td>
<td>-1.17</td>
</tr>
<tr>
<td>4</td>
<td>Formal assessments (e.g. tests) provide for a better evaluation of EFL students than do informal assessments (e.g. casual observation).</td>
<td>2.75</td>
<td>1.01</td>
<td>0.28</td>
<td>-0.60</td>
</tr>
<tr>
<td>19</td>
<td>In terms of student learning, assessments take up more time and effort than they are worth.</td>
<td>2.47</td>
<td>1.09</td>
<td>0.60</td>
<td>-0.24</td>
</tr>
<tr>
<td>20</td>
<td>All things considered, paper-and-pencil assessments are the best methods for determining what each student has learned.</td>
<td>2.44</td>
<td>1.00</td>
<td>0.54</td>
<td>-0.10</td>
</tr>
<tr>
<td>26</td>
<td>I use published assessment methods (e.g., in textbooks) as my primary basis for assigning grades to my students.</td>
<td>2.03</td>
<td>1.10</td>
<td>0.92</td>
<td>-0.04</td>
</tr>
</tbody>
</table>
There were significant differences in the beliefs of NS teachers and NNS teachers concerning paper-and-pencil assessments. NS teachers believed that speaking and listening assessments were better for assessing learning than paper-and-pencil assessments ($\bar{x} = 3.08$) to a greater extent than NNS teachers ($\bar{x} = 2.58$), $U = 1,509.00$, $z = 1,509.00$, $p = .003$. Similarly, NS agreement ($\bar{x} = 3.94$) to the statement that assessment methods related to real life were better than paper-and-pencil procedures was significantly higher than NNS agreement ($\bar{x} = 3.30$), $U = 1,219.50$, $z = -4.38$, $p = .000$. In accordance with these results, NNS teachers agreed much more that paper-and-pencil assessments were the best method for determining what each student has learned, and also that they provide the primary basis for assigning grades.

A very few number of divergent beliefs were found between part-time (PT) teachers and full-time (FT) teachers. Only two questions elicited responses in which significant differences were found. The statement that assessment tasks provide a valuable learning experience for students elicited more agreement from FT teachers ($\bar{x} = 4.20$) than PT teachers ($\bar{x} = 3.70$), $U = 3,064.50$, $z = 3.26$, $p = .001$. This more negative view of assessment held by PT teachers is possibly linked to their practice of using published assessment methods as the primary basis of assigning grades to students. PT teachers ($\bar{x} = 2.46$) reported doing this to a greater degree than FT teachers ($\bar{x} = 1.83$), $U = 1,660.00$, $z = -3.00$, $p = .003$. 
Discussion

The most common reason given by teachers for assessing students was administrative in nature, namely, to determine final grades. Following on from that, the next four most important purposes of assessment were all student centered: to obtain information on student progress, to provide feedback, to diagnose student strengths and weaknesses, and to motivate students to work harder. These results are consistent with the most common purposes of assessment for teachers in Hong Kong, China and Canada (Cheng et al., 2004). Teachers seem to place most value on assessment practices for encouraging student learning and guiding growth. In contrast, assessment was considered as less important for professional development and lesson planning.

The most common methods of assessment were reported as being teacher constructed, with less assessment being student-conducted and even less as standardized assessment. This indicates that assessment is still firmly in the hands of teachers, who construct assessment tasks and evaluate the results largely independently. There is very little in the way of top-down pressure from the university to use standardized tests, and little bottom-up input from students in terms of peer and self-assessment. EFL teachers in Japanese universities are at the front and center of assessment practice in their classrooms, maintaining a high level of independence in the way assessment is undertaken and judgments made. Only 18.2% of teachers reported using assessment for the purpose of preparing students for a standardized test (such as TOEIC), 19.6% reported using such standardized tests in their classes, and 8.8% sourced their test items from a mandated syllabus or curricula. This seems to indicate that standardized and high-stakes testing is not of great concern to university teachers, although it may be of concern to university students. This was also evident from the way in which teachers reported sourcing test items. The most popular reported source of test items was that developed by teachers themselves (n = 123), which was almost two times greater than the second most popular source of items, that from published textbooks (n = 68), which further adds to the picture being formed of teachers exhibiting a high degree of autonomy over the assessment process.

It has been noted that there is a belief in Confucian Heritage Cultures that educational testing and assessment leads to competition and hard work, which brings out the best in people (Carless, 2011). Within Japan, however, it has been argued that the over-importance placed on testing has led to many detrimental effects on genuine education (Kerr, 2001; McVeigh, 2006; Sugimoto, 2014). Contrary to this, the EFL teachers in this survey generally supported the view that assessment is useful in promoting student learning. Most agreed that assessment results are important for instruction, that assessment tasks are a valuable learning experience and that assessment improves EFL learning. If teachers were obligated to implement high-stakes tests and standardized assessments, they would probably have a much weaker belief in the value of assessment, as indicated by their negative views of formal testing and paper-and-pencil assessments.

Teachers were not so strong, however, in their beliefs about assessment as a tool for
enhancing instruction. There was weak agreement to the statement, “assessment helps to focus my teaching”, and teachers were unsure whether “assessment results tell me how well I have taught my students”. It appears that teachers are more interested in using assessment to give feedback to students rather than to get feedback for themselves. It could be that teachers feel confident in their teaching and instruction skills, but are unsure about using assessment results for self-improvement and professional development.

Native speaker teachers were more concerned about giving feedback to students as they progressed than were non-native speaker teachers. The NS teachers reported using checklists, giving written comments, holding student conferences and assigning letter grades to a much greater extent than NNS teachers. Instead, NNS teachers reported providing test scores as a method of feedback to a greater extent. It was the NNS teachers who held a higher regard for paper-and-pencil assessments, which provided the primary basis for the grades they assigned to students. Perhaps the NNS teachers, presumably all of whom were Japanese nationals, were raised in a culture where test scores were the predominant means of getting feedback concerning learning progress, and this cultural background has carried over into their own teaching practice.

It was also found that NS teachers used performance-based assessment tasks more often than NNS teachers. This could be due to the fact that in Japanese higher education it is the NS teachers who more often take the speaking and communication classes, which lend themselves more easily to performance-based assessment. The NNS teachers, on the other hand, reported using published test items more often than NS teachers. This is understandable if NNS teachers are mainly focusing on reading and listening skills. The creation of a reading assessment that is appropriately graded and correctly worded requires a considerable investment of time and effort. This is even more true of listening assessments, which also require audio recording and editing skills. If in fact NNS teachers were mainly teaching receptive skills, this would also help explain why they commonly give feedback in terms of test scores rather than written comments or student conferences, as this is more suitable for listening and reading. However, in the current survey, only 37.83% of NNS teachers reported that they were thinking of listening or reading classes when they completed the survey. Most NNS teachers were responding according to classes they taught in writing, speaking and general English skill areas. The reason why NNS teachers reported using published test items to such a high degree is still unclear, and invites further research.

In a similar way, it was found that part-time teachers used items from published textbooks and tests provided with textbooks roughly twice as often as full-time teachers. As part-time teachers often divide their time between a number of different universities, the ease of using existing tests would present itself as an efficient alternative to developing a number of different assessments across varying English programs. Tests that are provided with the textbook also provide a certain degree of confidence that what is being assessed is that which has been covered during the course.

Although some other statistically significant differences were found between full-time
and part-time teachers, in actuality these didn’t amount to a great deal of variation. The similarities were far more striking. In particular, for all teachers the main purposes for assessment practice are student-focused, while the influence of high-stakes standardized tests is minimal. Teachers reported that they felt no real pressure to teach toward standardized examinations and the general benefits of assessment for enhancing learning were affirmed.

Consideration of the current beliefs and practices of EFL teachers provides a guide by which action can be taken to foster a culture of assessment that promotes productive student learning. The main encouraging result from this survey is that at the level of teacher beliefs, there is broad acceptance of the helpful role assessment plays in enhancing learning. Assessing student ability and reporting the results is not viewed as a mere bureaucratic burden that lies extraneous to the real business of teaching and learning, but it is believed to be an important part of EFL education and to provide a valuable learning experience for students. These beliefs are present across the varying cultural and contextual backgrounds of teachers working in Japanese EFL, and provide a foundation upon which learning-oriented assessment practices can be built.

It has been observed that high-stakes standardized testing, based upon psychometric models of assessment and focusing students’ attention on achieving a passing grade rather than actual achievement, is often at odds with assessment that has student growth as its primary concern (James, 2012; McVeigh, 2002). It is therefore reassuring to note that teachers in Japanese EFL do not feel much pressure to prepare students for high-stakes tests. Critics of Japanese higher education may attribute this to the mindset that once the hard work of passing a university entrance exam is over, high-stakes tests are not that important and students can have a “holiday” for four years (Aspinall, 2010). On the other hand, newspapers report that a good TOEIC score remains highly important for job seekers when they graduate from university (Murai, 2016), and so students and university administrators no doubt feel pressure to expend time and effort in preparing students for that test. This pressure does not seem to affect teachers, though, who are able to focus their assessments on encouraging learning, should they so desire.

In fact, teachers reported using assessment methods that were oriented toward a learning outcome. Teachers agree that they need a variety of tasks to assess their students, and that frequent shorter assessments are more effective than less frequent longer assessments. There is also a belief that the value of performance-based tests related to the real world is greater than the value of traditional paper-and-pencil tests. These beliefs are completely in line with assessment practices that promote learning (Abrams, 2015; Finch, 2002; Lopez-Pastor & Sicilia-Camacho, 2015; Sambell, McDowell, & Montgomery, 2013).

However, when we consider how teachers reported actually conducting assessment, we find a greater disconnect with good practices theory. The value of peer and self-assessment for learning has been overwhelmingly supported by the literature (Asaba & Marlowe, 2011; Klenowski, 2009; Sato, 2013; Taferner, 2008), and yet only 42.6% of teachers reported using self-assessment and 39.2% reported using peer assessment. Teachers did report giving a
certain amount of feedback as verbal comments during class, as well as written comments and test scores. But only a small proportion (36.5%) reported using checklists or other assessment rubrics when providing feedback. Without such rubrics, a test score or letter grade would hold minimal value for encouraging learning. With a rubric, the often hidden goals of the curriculum are made clearer, and students are able to evaluate their own achievement and make plans to progress toward the next learning goals. While it appears that teachers have a productive belief system that aligns closely with learning-oriented assessment principles, their actual practice would benefit from giving a greater importance to developing students’ evaluative expertise and deepening students’ engagement with feedback.

Conclusion

Although some scholars and social commentators have painted a bleak picture of the ways in which high-stakes testing exerts a powerful and debilitating effect on Japanese education in general (Clark, 2010; Kerr, 2001; McVeigh, 2006), the situation of EFL in higher education appears somewhat rosier. Teachers, who are closest to the front-lines of EFL learning, report having freedom from the institutional power exerted by high-stakes summative tests. They also have confidence in assessment as a positive force for improving learning outcomes and empowering learners. Once these positive beliefs get channeled into more effective processes and practices of assessment, the benefits for higher education throughout Japan will be keenly felt.

Acknowledgements

This study was conducted as part of a much larger doctoral research project into learning-oriented assessment, under the supervision of Toru Kinoshita. His valuable advice and support is gratefully acknowledged, along with the helpful suggestions of two anonymous reviewers.

References


