



#JALT2019 • NAGOYA 11.1-11.4

Teachers' Beliefs About Classroom-Based Assessment

Adam Murray

University of the Ryukyus

Yuji Nakamura

Keio University

Kahoko Matsumoto

Tokai University

Yasuko Ito

Kanda University of International Studies

Taiko Tsuchihira

Seitoku University

Reference Data:

Murray, A., Nakamura, Y., Matsumoto, K., Ito, Y. & Tsuchihira, T. (2020). Teachers' beliefs about classroom-based assessment. In P. Clements, A. Krause, & R. Gentry (Eds.), *Teacher efficacy, learner agency*. Tokyo: JALT. <https://doi.org/10.37546/JALTPCP2019-62>

In this paper, we report on the early stages of the development of the Classroom-Based Assessment Self-Efficacy Scale (CBA-SES), an instrument designed to examine how Japanese

Teachers of English (JTEs) feel about classroom-based assessment. The questionnaire (31 items) consists of three sections: (a) teachers' beliefs, (b) teachers' self-efficacy, and (c) their own teaching practice. We pilot tested this instrument with 30 JTEs in order to assess its appropriateness and to get a better understanding of the tendencies and characteristics of JTEs. We found that the belief statements are suitable, but revision along with additional statements will be needed for self-efficacy and practice for the next version of the instrument. The participants believed language tests should resemble real-life language use. Notably, they were able to make such tests and were doing so in their teaching contexts. They also felt that effective feedback and the use of clear learning targets were important.

教室内評価の重要性が増しつつある中、日本人英語教師がそのような評価に対してどのように感じ、また実践しているのかを探るため、本研究では教師の自己評価を測る実験的な質問紙を作成し、現役の英語教師に回答を依頼した。質問紙は、信条 (Belief)、自己効力感 (Self-efficacy)、実践 (Practice) の3セクションから成り、それぞれ、「評価はこうあるべき」、「このような評価を行うことができる」、「このような評価を実際に行っている」という側面の自己評価を測定した。その結果、信条に関する項目は適切であったが、自己効力感と実践に関するものについては次の本格的な実施に向けて修正や追加が必要であることが判明した。殆どの回答者が現実の言語使用を反映したテストの作成が重要と考え実践しようとしていること、また、明確な指導目標の設定と効果的なフィードバックの重要性を感じていることは注目すべきであった。

As a result of the recent changes in the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Course of Study placing a greater emphasis on the input-output connection and the use of CEFR-based can do-type criteria, teachers are now expected to effectively implement classroom-based assessment (CBA). Actually, the new so-called "Core Curriculum for Teacher Education" (MEXT, 2016), for the first time, explicitly stresses the development of teachers' ability to carry out valid and proper evaluation. It is unclear how Japanese teachers of English (JTEs) at both public and private schools around Japan are coping with this fundamental change to classroom practices. In this paper, we report on an attempt to develop an instrument to measure JTEs' attitudes and beliefs about CBA, and actual implementation of CBA in their classrooms. We conclude with suggestions for teacher education programs for pre- and inservice language educators.



#JALT2019 • NAGOYA 11.1-11.4

Murray, Nakamura, Matsumoto, Ito, & Tsuchihira: *Teachers' Beliefs About Classroom-Based Assessment*

Background

Although there are many descriptions and definitions of classroom-based assessment (CBA), we use Hill and McNamara's: "any reflection by teachers (and/or learners) on the qualities of a learner's (or group of learners') work and the use of that information by teachers (and/or learners) for teaching, learning (feedback), reporting, management or socialization purposes" (2012, p. 396). This broad definition not only covers all forms and uses of assessment, but also the characteristics of teachers and learners.

In addition to defining assessment, it is important to describe assessment literacy. Fulcher (2012) defines language testing and assessment literacy as:

The knowledge, skills and abilities required to design, develop, maintain or evaluate, large-scale standardized and/or classroom-based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. (p. 125)

Because of the importance placed by MEXT on classroom-based assessment, we are interested in teachers' beliefs about CBA, the teachers' levels of efficacy, and the extent to which it is implemented in their classrooms.

Although teacher self-efficacy (TSE) is a fairly well-established area in general education, relatively little research has been published about language teachers self-efficacy (LTSE) (see Klassen et al., 2011; Wyatt, 2018). TSE beliefs can be defined as "teachers' beliefs in their abilities to support learning in task-, domain- and context-specific cognitive, metacognitive, affective and social ways" (Wyatt, 2018, p. 93).

There are some studies on TSE, for instance, a cross-cultural study comparing Japanese and Finish primary and lower secondary school teachers (Yada et al., 2019) and Japanese high school English teachers' present state of TSE (Thompson & Woodman, 2018). However, the former does not focus on language teachers while the latter deals more with teaching practice than assessment. At the same time, many existing studies focus on agents-ends (beliefs about one's ability to cause specific outcomes) rather than agents-means (a belief in the ability to take action). Also, some of the studies utilize questionnaire items such as "teach speaking skills" that are insufficiently task-specific. One important element of classroom teaching is assessment. Without appropriate assessment, it is difficult, if not impossible, to evaluate the effectiveness of instruction. In order to better understand JTEs, we need to know more about their levels of assessment literacy in terms of the three dimensions: practice, concepts, and context (Fulcher, 2012). Also, because the teacher is primarily responsible for classroom-based assessment (CBA),

we decided to place our focus on that rather than other forms of assessment. Classroom-based assessment consists of four phases of assessment: planning, framing, conducting, and using (Hill & McNamara, 2012):

- Planning – The kinds of planned assessment tasks and how they are related/connected to classroom instruction.
- Framing – If and how students are informed that a classroom activity is for assessment.
- Conducting – Types of assessment used in the foreign language classroom (formal: tests, assignments; planned: activities used for assessment; incidental: unstructured observation).
- Using – Teachers can use assessment for a variety of purposes: teaching, learning (feedback), reporting, management, and socialization.

We set out to develop an instrument to measure JTEs' beliefs about classroom-based assessment, their levels of self-efficacy, and their actual assessment practices. In other words, we aimed to develop an instrument that incorporated the four phases of assessment: planning, framing, conducting, and using (Hill & McNamara, 2012) in three essential dimensions of teacher assessment literacy (TAL): practice, concepts, and context (Fulcher, 2012). With this in mind, we developed a preliminary version of our instrument which was heavily influenced by the Michigan Assessment Consortium's *Assessment Literacy Standards* (2017). Additional influences were two studies that took place in Japan. The first study (Nishino, 2012) used an instrument named the *Teacher Beliefs Questionnaire* which consisted of 11 variables such as L2 self-confidence and CLT self-efficacy. The second study (Thompson & Woodman, 2018) used an instrument named *The Japanese Teacher of English Teacher Efficacy Scale (JTE-TES)* which had questions such as "How confident are you that you can develop appropriate assessments for evaluating your students' English ability?"

Our Objectives

The first objective of this pilot study was to assess the appropriateness of our instrument for JTEs. First, we wanted to check the applicability of the items selected for three areas (belief, self-efficacy, and practice). Also, we wanted to identify tendencies and defining characteristics in order to create standards and criteria to be used for future assessment literacy education in Japan, and for this purpose, we aimed to collect the responses from teachers teaching at different levels of education.

Methods

Participants

To collect responses, we created an online questionnaire (Google Form) and personally invited colleagues by email and in person to complete it. The questionnaire had a brief overview of the study and provided contact information (email) so that the participants could contact the authors with any questions or concerns. However, we were unable to get as many responses as anticipated. Fortunately, one of the authors was able to ask 12 teachers who had gathered for a meeting for a different research project to respond to the questionnaire. She also took the opportunity to collect verbal open-ended comments and feedback. In total, 30 participants, English teachers in different teaching contexts, responded to the preliminary version of our instrument. Figure 1 shows the ratio of the school settings where they were teaching.

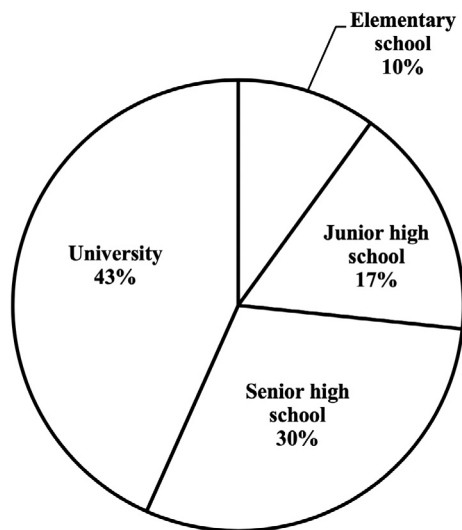
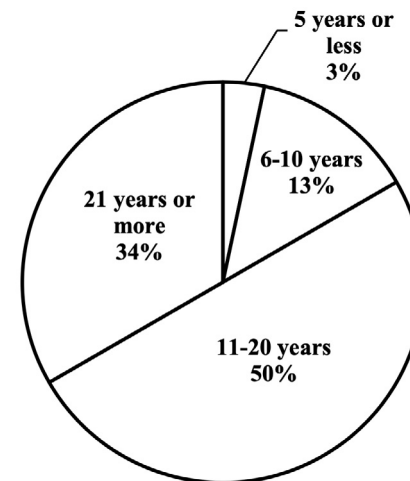


Figure 1. Participants' school settings.

A large percentage of the participants (43%) were teaching at the university level and nearly half of participants (47%) were teaching at the secondary school level. There were only three participants teaching at elementary schools. Figure 2 provides the data on

number of years of teaching.

Figure 2. Participants' number of years of teaching.



The majority of the participants (84%) were experienced teachers, teaching for more than 10 years. Many of them could be considered veteran teachers with 21 or more years of experience. Only one teacher had five years or less of teaching experience.

It is probably worth noting that these participants were likely highly motivated teachers because they were willing to respond to our online questionnaire. Keeping this in mind, the interpretation of the results must be regarded with caution. In fact, approximately one-third of the data were collected at one time by one of the authors. This enabled us to gather qualitative information (reactions and spontaneous comments) which were helpful in interpreting the quantitative information collected by the items on the questionnaire.

Questionnaire

The short questionnaire created for the present study consisted of 31 items organized into three sections: teacher beliefs (11 items), self-efficacy (10 items), and teaching

practice (10 items). Belief refers to what teachers think they should do—in other words, their teaching ideals, while self-efficacy indicates what they think they are capable of doing. The last section, practice, is what they think they are actually doing in their teaching. All items utilized 5-point Likert scales, and are based on the *Assessment Literacy Standards* of the Michigan Assessment Consortium (2017). The *Assessment Literacy Standards* (2017) were created by Michigan educators and national experts and endorsed by the Michigan State Board of Education in 2016. They are divided into sections depending on the targets (e.g., students and their families, teachers, administrators, and policy makers), and the section for teachers was referred to in order to create the items used in the present study.

The items on teachers' beliefs were about what teachers think they should do. For example:

- B1. Teachers should understand and be able to use tests.
- B2. Multiple assessments provide a more balanced picture of a student.
- B3. When done correctly, the resulting data can be used to make sound educational decisions.

(See Appendix for all the items along with their Japanese translations)

On the other hand, the items on teacher self-efficacy were about what teachers think they can do. For example:

- E1. I can select and use various assessment methods appropriate to assessment purposes and learning targets.
- E2. I can implement the 5-step process (plan, develop, review, field test, review and revise) for assessment development.
- E3. I can use learning targets aligned to the standards to guide instruction.

Finally, there were items about teaching practice: how teachers implemented assessment in their classrooms and how often it was done. For example:

- P1. I use various assessment methods.
- P2. I use the 5-step process (plan, develop, review, field test, review and revise) for assessment development.
- P3. I use learning targets to guide instruction.

On the online questionnaire, each question was displayed on the screen and the participants were to answer about themselves by choosing from 5 options such as a) very much agree, b) agree, c) neutral, d) disagree, e) very much disagree. The data were gathered online and analyzed using Microsoft Excel by the authors.

Results

The results will be first given for all three sections, followed by comparisons of some items. Figures 3, 4, and 5 show the results of teachers' ratings on all items related to belief, efficacy, and practice, respectively.

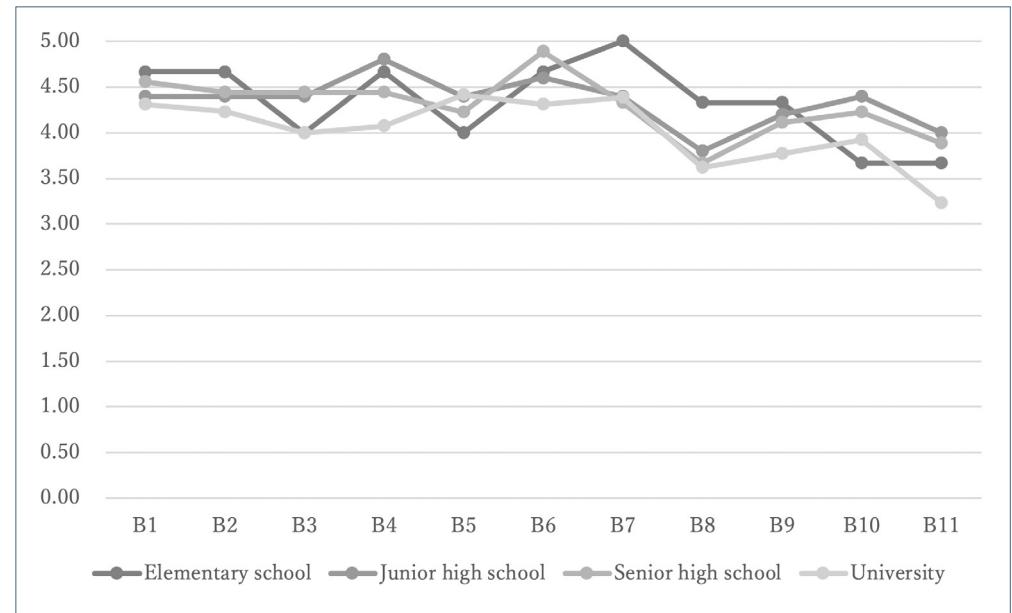


Figure 3. Average responses on belief items.

There were relatively low ratings for B8 (“students should learn how to use assessment results to improve their learning”). This seems to suggest that the participants think that student-centered learning or autonomy is not essential. Also, there were lower ratings for B11 (“language tests should resemble real-life language use”). This may be the result of the difficulty of making such tests as well as a lack of awareness of the importance of connecting teaching, which has become increasingly communicative, with assessment.

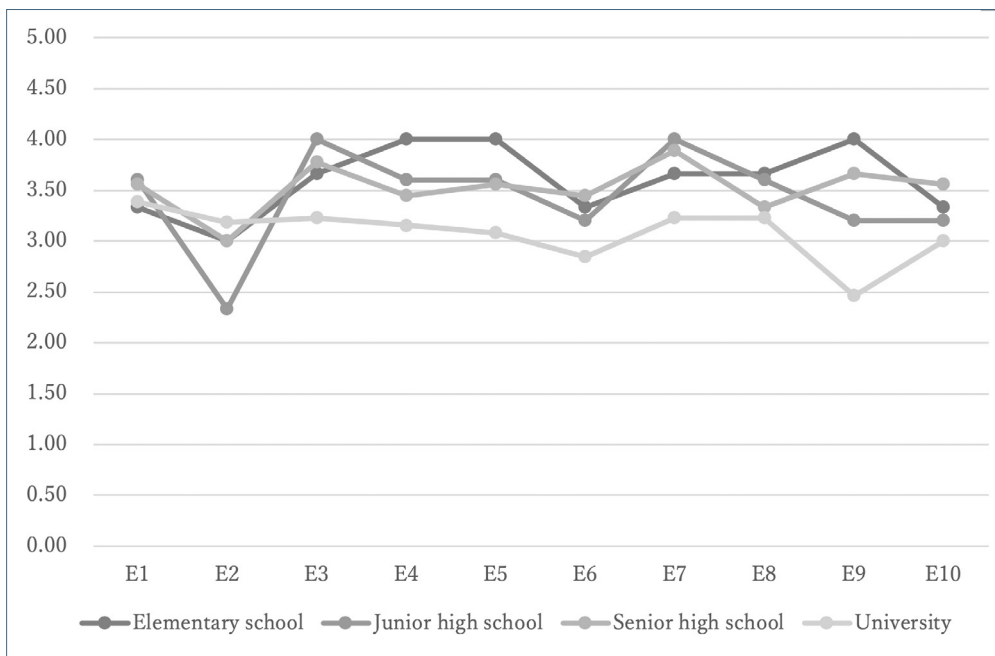


Figure 4. Average responses on efficacy items.

As can be seen in Figure 4, participants, particularly junior high school teachers, rated themselves lower on E2 (“I can implement the 5-step process [plan, develop, review, field test, review and revise] for assessment development”), compared to the other items. For E9 (“I can communicate effectively with students, parents/guardians, other teachers, administrators and community stakeholders about student learning”), they gave themselves lower ratings, particularly the university teachers. This can be attributed to the fact that university teachers are unlikely to communicate with stakeholders other than students.

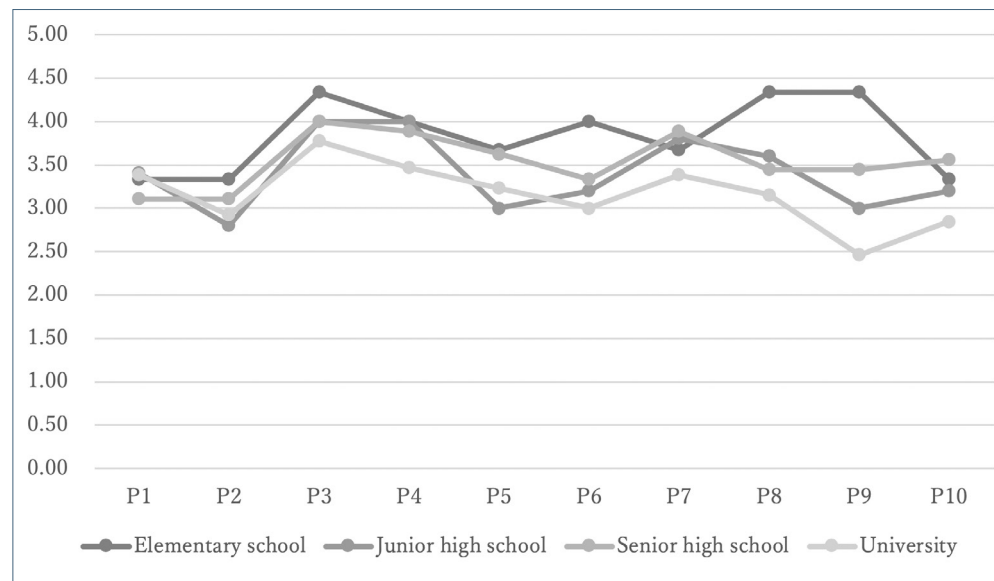


Figure 5. Average responses on practice items.

P2 (“I use the 5-step process (plan, develop, review, field test, review and revise) for assessment development”) and P9 (“I communicate effectively with students, parents/guardians, other teachers, administrators and community stakeholders about student learning”) were rated lower than other items, which is a similar pattern to the efficacy items, as illustrated in Figure 4.

We have examined belief, efficacy, and practice separately, but it is also crucial to examine the relationship between pairs of elements as well as among the three. Comparing efficacy and practice enables us to explore whether or not the teachers think they are doing what they think they are capable of. Table 1 shows the overall average scores for efficacy and practice items.

Table 1. Efficacy and Practice Average Scores

Item	Efficacy	Practice
1	3.47	3.30
2	3.00	2.85
3	3.57	3.93
4	3.40	3.73
5	3.38	3.34
6	3.13	3.23
7	3.60	3.63
8	3.37	3.43
9	3.10	3.03
10	3.23	3.17

We can observe that there is a greater difference between efficacy and practice scores with Items 3 and 4 compared to the other items. Statements for E3, P3, E4, and P4 are given in Table 2.

Table 2. Statements for Items No. 3 and 4

Item	Element	Statement
3	Efficacy	I can use learning targets aligned to the standards to guide instruction.
	Practice	I use learning targets to guide instruction.
4	Efficacy	I can use assessment results to make appropriate instructional decisions for individual students and groups of students.
	Practice	I use assessment results to make appropriate instructional decisions.

The results of the statistical analyses for the differences between each item for Efficacy and Practice can be found in Table 3.

Table 3. Results of T-Tests for Mean Efficacy and Practice Scores

		t	pf	p-value
Pair 1	E1 - P1	0.817	29	0.420
Pair 2	E2 - P2	1.143	19	0.267
Pair 3	E3 - P3	-2.483	29	0.019
Pair 4	E4 - P4	-3.340	29	0.002
Pair 5	E5 - P5	0.226	28	0.823
Pair 6	E6 - P6	-1.000	29	0.326
Pair 7	E7 - P7	-0.254	29	0.801
Pair 8	E8 - P8	-0.528	29	0.601
Pair 9	E9 - P9	0.571	29	0.573
Pair 10	E10 - P10	1.000	29	0.326

The difference between E3 and P3 are statistically significant. The same can be observed with the difference between E4 and P4. What the former indicates is that although the teachers thought they were not capable of using learning targets to guide instruction, they did think they were actually trying to do so in their classrooms. Similarly, the latter can be interpreted to mean that they thought they could not use assessment results to make appropriate instructional decisions, but that they were aiming to do so in their teaching.

A comparison between efficacy and practice was also done according to the participants' school setting, particularly for junior and senior high schools, as shown in Figures 6 and 7.

Murray, Nakamura, Matsumoto, Ito, & Tsuchihira: *Teachers' Beliefs About Classroom-Based Assessment*

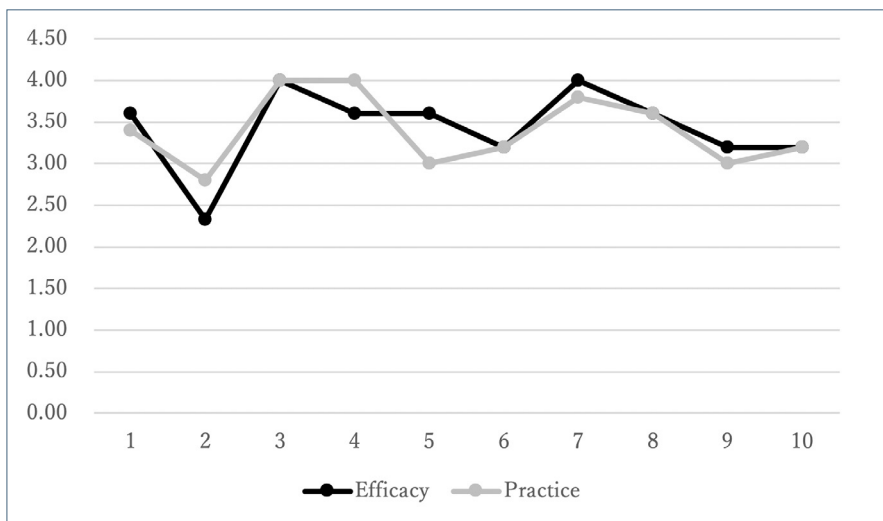


Figure 6. Junior high school teachers' average responses on efficacy and practice items.

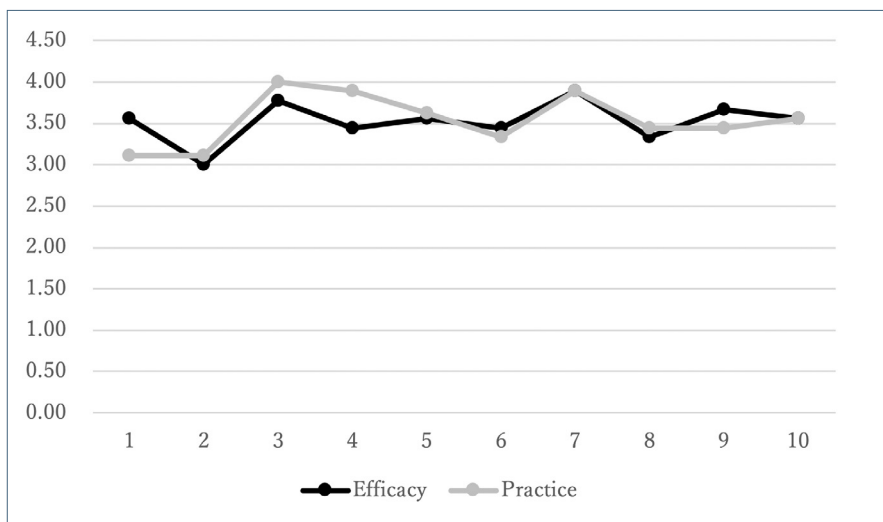


Figure 7. Senior high school teachers' average responses on efficacy and practice items.

It is likely that there is a bigger discrepancy with junior high school teachers than senior high school teachers between what the teachers thought they were doing and what they thought they were capable of.

Next, let us examine the relationships between belief and efficacy. Table 4 is a summary of the correlation between these two elements.

Table 4. Correlations between Belief and Efficacy

		Efficacy									
		E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
Belief	B1	●			●	●					
	B2										
	B3	●				●					●
	B4	●			●	●					
	B5					●					●
	B6	●		●	●	●		●			●
	B7		●			●					●
	B8										●
	B9				●	●		●			
	B10										
	B11				●		●			●	●

Note. The black mark means the two items are highly related at a significance level of < 0.001.

Among those significantly correlated, the following three pairs of items were most highly correlated: B11 and E10, B7 and E5, and B6 and E3. Thus, a closer examination of these three pairs and their corresponding practice items was carried out.

Table 5. Correlations among B11, E10, and P10

	E10 I can make tests that resemble real-life language use.	P10 I make tests that resemble real-life language use.
B11 Language tests should resemble real-life language use.	.717**	.632**
	0.000	0.000

Table 6. Correlations among B7, E5, and P5

	E5 I can provide timely, descriptive, and actionable feedback to students based on assessment results.	P5 I provide timely, descriptive, and actionable feedback to students.
B7 Effective feedback is necessary for learning.	.510**	.531**
	0.004	0.003

Table 7. Correlations among B6, E3, and P3

	E3 I can use learning targets aligned to the standards to guide instruction.	P3 I use learning targets to guide instruction.
B6 Clear learning targets are necessary for learning and assessment.	.485**	.570**
	0.007	0.001

Statistically significant correlations for B11, E10, and P10 indicate that the teachers believed that language tests should resemble real-life language use and they thought they could make those tests and were doing so in their classroom assessment. A similar pattern was found for the provision of effective feedback (Table 6) and the use of learning targets (Table 7).

Finally, the data were examined according to the number of years of teaching experience. Figures 8, 9, and 10 illustrate the results of responses for belief, efficacy, and practice respectively, according to the amount of teaching experience. It should be noted here, however, that there was only one participant belonging to the “5 years or less” group; therefore, the results for this group do not tell us much about any tendencies.

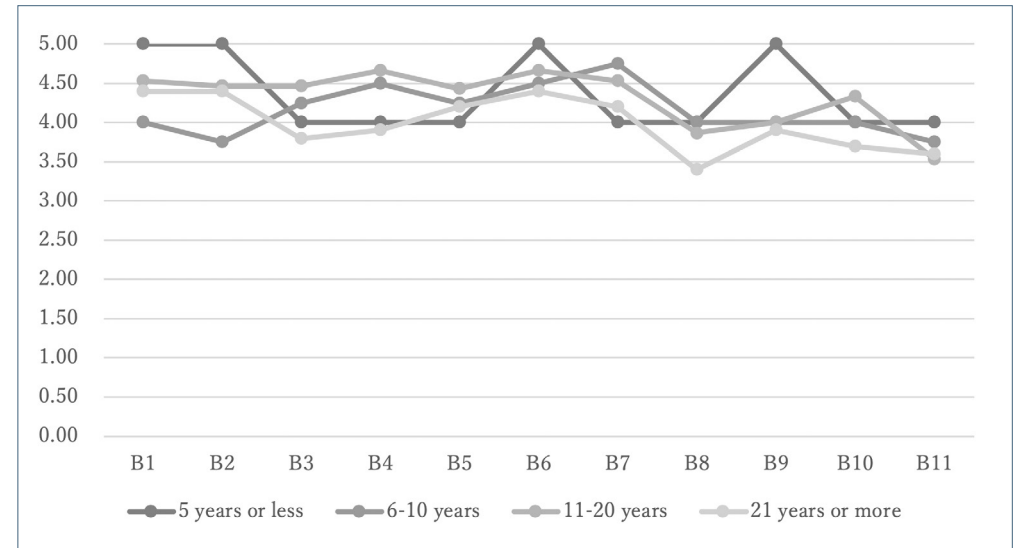


Figure 8. Average responses on belief items.

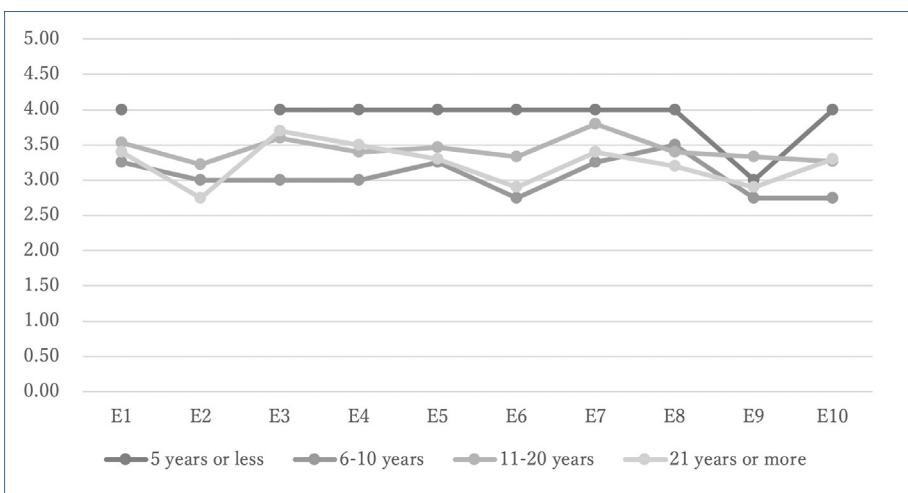


Figure 9. Average responses on efficacy items. Note. The participant for the “5 years or less” did not answer item E2 and the data was not available for the item.

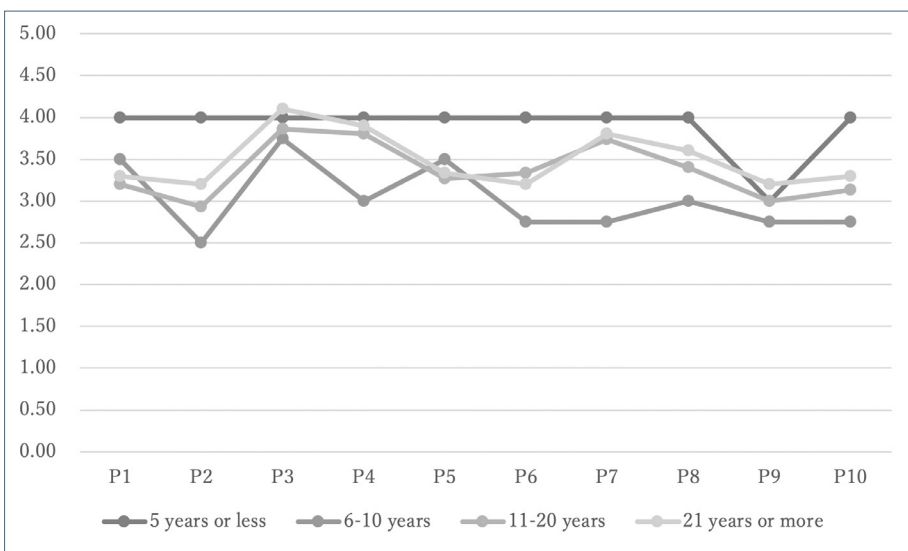


Figure 10. Average responses on practice items.

Due to the small sample size of each group, no statistical tests were run for the comparisons between the groups with different amounts of teaching experience. However, while less-experienced teachers (6-10 years of teaching experience) were less likely to feel that they were practicing what they thought they should compared with more-experienced teachers (11 years or more), both groups of teachers thought that they should be able to handle most aspects of classroom-based assessment.

Feedback from the Teachers

As previously described, 12 teachers were able to tell one of the authors their thoughts about the questionnaire items after completing it. These brief informal interviews were conducted to help us determine the applicability of the items and improve the instrument for future use. Though the teachers agreed to most of the items in the belief section to varying degrees, some of them reported having difficulty responding to these items because self-analysis and reflection were required. A typical comment was that they lacked time to think about these general beliefs and principles as their days were packed with the work-related pressures involved in dealing with a variety of people, including students, colleagues, supervisors and other superiors, and parents. In responding to the sections on efficacy and practice, it was pointed out that answers may vary depending on the kinds of evaluation (formative or summative) and how much of the school curriculum is controlled by entrance exam preparations. In other words, the teaching context (private vs. public schools, competitive vs. regular schools, etc.) has a large influence on teachers that cannot be overlooked. Therefore, items will need to be revised to better reflect the variety of teaching contexts that JTEs work in.

Conclusion

The first objective of this pilot study was to assess the appropriateness of our instrument for JTEs. In terms of the belief statements, all of the respondents (at all four teaching contexts) showed relatively high levels of agreement. They believed that tests and their results were important for making educational decisions. For this reason, these statements are suitable for Japanese teaching contexts. However, the JTEs did not feel as strongly about student use of test results, subjective interpretation of results, and tests resembling real-life use. These three statements require further refinement.

Next, the majority of the statements about efficacy (E1-E10) are appropriate for Japanese teaching contexts. The JTEs showed relatively high levels of agreement with the statements. However, two items may need revision. The statement “I can implement the



#JALT2019 • NAGOYA 11.1-11.4

Murray, Nakamura, Matsumoto, Ito, & Tsuchihira: *Teachers' Beliefs About Classroom-Based Assessment*

5-step process" (E2) may be difficult for the respondents because they may be unfamiliar with this model of assessment development. One possible improvement would be dividing the process into five separate statements aligned with the steps of the process. Also, the statement about stakeholders (E9) needs to be revised to better reflect the stakeholders that JTEs actually interact with.

Finally, as mentioned in the feedback from the teachers section, additional statements will be required particularly about efficacy and actual practices. Separate statements are needed for summative and formative assessment. Also, teachers may work in contexts where their actual teaching practices do not align with their personal beliefs because of institutional guidelines. It is necessary to collect more information about the teaching context in the next iteration of this instrument. The quantitative analysis of responses along with the feedback gives direction for further refinement of the items. Also, none of the participants expressed any opinions (written or verbal) about the addition or deletion of items. For these reasons, it can be said that this questionnaire is appropriate for JTEs and Japanese teaching contexts.

The second objective of this study was to better understand the tendencies and characteristics of JTEs. Generally speaking, the teachers who participated showed quite strong beliefs and high levels of self-efficacy in their responses, which was one reason why their self-efficacy responses and those of actual practice were all highly correlated. As previously cautioned, the results of this study are probably not generalizable because it is highly likely that only motivated teachers responded to the online questionnaire. Also, the sample size is rather small and the respondents' backgrounds are not well balanced. For a larger-scale study, we need to find participants with a variety of backgrounds and skill levels. However, there were some interesting findings as to the areas where teachers' beliefs and self-efficacy were not directly reflected in actual practice, notably the significant differences found between E3 and P3 and E4 and P4. E3 and P3 relate to the ability to use learning targets to guide instruction, while E4 and P4 relate to the ability to use assessment results to make appropriate instructional decisions. This means that the teachers' perceptions of self-efficacy for these items were relatively low even though they thought they were actually doing those things or trying to do them in practice.

In terms of teaching experience, both less- and more-experienced teachers demonstrated high levels of agreement with the belief statements. As might be expected, less-experienced teachers (6-10 years of experiences) were more likely to feel that their practices were not as aligned with their beliefs as they could be. Although this cannot be generalized, it suggests that they were aware of the significance of these assessment principles.

This pilot study, though small in scale, gave us many insights into the instrument. We learned that the majority of the belief statements are appropriate for Japanese teaching contexts. The feedback from the teachers, particularly about efficacy and practice, will help us refine the statements and create additional ones so that the instrument better fits the present context of Japanese English teacher education. We will replicate this study with a larger number of respondents with the goal of these items being used as standards or criteria in Japanese assessment literacy education.

Thus, teachers may need further professional development through assessment literacy training opportunities, either face-to-face or online, so that they can expand and deepen their knowledge and improve their skills in these areas: establishing learning targets aligned to the standards that guide instruction, and making better use of assessment results so as to make appropriate instructional decisions for individual students and the whole class. Therefore, teacher education programs for pre- and inservice language educators may need to provide more practice with skills, rather than instruction about beliefs, in order to enable less-experienced teachers to actually do what they think they should do.

Though this research is still in the preliminary stages, we will do more fine-tuning to the questionnaire items related to these areas, and plan to provide concrete suggestions for pre- and inservice language educators based on the analysis of the next stage of the study.

Bio Data

Adam Murray has been an active member of JALT since 2003. He teaches at the University of the Ryukyus (Okinawa). His research interests are listening instruction, materials development, and assessment. <murray@gec.lab.u-ryukyu.ac.jp>

Yuji Nakamura is professor of English at Keio University, where he teaches English teaching methodology, language testing, and English as an international language. His primary research focus has been L2 reading assessment, in-house placement tests and longitudinal analyses of students' English proficiency. <nkyj@flet.keio.ac.jp>

Kahoko Matsumoto is professor at Tokai University and has been teaching English and applied linguistics in the Tokyo Metropolitan area for over 20 years. Her research interests are performance assessment and intercultural competence development. <mkahoko@tsc.u-tokai.ac.jp>



#JALT2019 • NAGOYA 11.1-11.4

Murray, Nakamura, Matsumoto, Ito, & Tsuchihira: *Teachers' Beliefs About Classroom-Based Assessment*

Yasuko Ito is professor in the Department of English at Kanda University of International Studies. She has a PhD in second language acquisition from the University of Hawaii at Manoa. Her research interests include second language acquisition, second language phonology, and language testing. <yasukoi@kanda.kuis.ac.jp>

Taiko Tsuchihira is associate professor at Seitoku University. Her main interest is language testing, especially testing listening and speaking. She also teaches English at an elementary school. <ttaiko@wa.seitoku.ac.jp>

References

- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132. <https://doi.org/10.1080/15434303.2011.642041>
- Hill, K., & McNamara, T. (2012). Developing a comprehensive research framework for classroom-based assessment – an empirical study. *Language Testing*, 29(3), 395-420. <https://doi.org/10.1177/0265532211428317>
- Klassen, R. M., Tze, V. M. C., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998-2009: Signs of progress or unfulfilled promise? *Educational Psychology Review*, 23, 21-43. <https://doi.org/10.1007/s10648-010-9141-8>
- MEXT. (2016). 教職課程コアカリキュラム [Core curriculum for teacher education courses]. Retrieved May 25, 2020, from https://www.mext.go.jp/b_menu/shingi/chousa/shotou/126/houkoku/1398442.htm
- Michigan Assessment Consortium. (2017). *Assessment literacy standards*. Mason, MI: Michigan Assessment Consortium.
- Nishino, T. (2012). Modeling teacher beliefs and practices in context: A multimethods approach. *The Modern Language Journal*, 96(3), 380-399. <https://doi.org/10.1111/j.1540-4781.2012.01364.x>
- Thompson, G., & Woodman, K. (2018). Exploring Japanese high school English teachers' foreign language teacher efficacy beliefs. *Asia-Pacific Journal of Teacher Education*. <https://doi.org/10.1080/1359866X.2018.1498062>
- Wyatt, M. (2018). Language teachers' self-efficacy beliefs: A review of the literature (2005-2016). *Australian Journal of Teacher Education*, 43(4), 92-120. <https://doi.org/10.14221/ajte.2018v43n4.6>
- Yada, A., Tolvanen, A., Malinen, O., Imai-Matsumura, K., Shimada, H., Koike, R., & Savolainen, H. (2019). Teachers' self-efficacy and the sources of efficacy: A cross-cultural investigation in Japan and Finland. *Teaching and Teacher Education*, 81, 13-24. <https://doi.org/10.1016/j.tate.2019.01.014>

Appendix

Questionnaire Items (Classroom-based Assessment Self-efficacy Scale)

BELIEFS

B1. Teachers should understand and be able to use tests.

教師はテストについて理解し、使えるべきである。

B2. Multiple assessments provide a more balanced picture of a student.

複数の評価をすることで、学生(生徒)の学力についてよりバランスのとれた把握ができる。

B3. When done correctly, the resulting data can be used to make sound educational decisions.

正しい評価をすれば、より確実な教育的決定をするためにその結果を使うことができる。

B4. Quality assessments are a critical attribute of effective teaching and learning.

質の高い評価は、効果的な指導・学習を行う上で重要な要素である。

B5. Assessment results should be used to make instructional decisions to improve student learning.

評価の結果は、学生(生徒)の学習を向上させるための指導上の決定をする際に使うべきである。

B6. Clear learning targets are necessary for learning and assessment.

学習と評価には明確な目標設定が必要である。

B7. Effective feedback is necessary for learning.

学習のためには、効果的なフィードバックが必須である。

B8. Students should learn how to use assessment results to improve their learning.

学生(生徒)は自身の学習の向上のためにどのように評価結果を使うかを学ぶべきである。

B9. Good classroom assessment and quality instruction are intricately linked to each other.

良質の教室内評価と質の高い指導は密接に結びついている。

B10. Grading is an exercise in professional judgment, not just a numerical, mechanical exercise.

評点を与えることは、単に数量的な操作ではなく、専門的判断をするということである。

B11. Language tests should resemble real-life language use.

言語テストは現実生活における言語使用に似たものであるべきである。

EFFICACY

E1. I can select and use various assessment methods appropriate to assessment purposes and learning targets.

私は、評価目的と学習指標に対して適切な様々な評価方法の選択・使用ができる。

E2. I can implement the 5-step process (plan, develop, review, field test, review, and revise) for assessment development.

私は評価方法を開発する際に、5段階のプロセス(計画、開発、チェック、実施、改良)を行うこと。

E3. I can use learning targets aligned to the standards to guide instruction.

私は指導をする際に、客観的基準に基づく学習目標に沿って行うことができる。

E4. I can use assessment results to make appropriate instructional decisions for individual students and groups of students.

私は、個々の学生(生徒)と集団の両方に対して、評価結果を使った適切な指導上の決定をすることができる。

E5. I can provide timely, descriptive, and actionable feedback to students based on assessment results.

私は、評価結果に基づいて、記述によって行動指針を示すようなフィードバックを、適切なタイミングで与えることができる。

E6. I can use grading practices that result in grades that are accurate, consistent, meaningful and supportive of learning.

私は正確かつ一貫性があり、学習に対して意義がありそれを助けるような評点の与え方ができる。

E7. I can use assessment results appropriately to modify instruction to improve student achievement.

私は、学生(生徒)の学習成果を向上させるように指導方法を修正するために、評価結果を適切に使うことができる。

E8. I can use multiple sources of data over time to identify trends in learning.

私は学習状況の傾向を認識するために、長期間にわたって様々な情報源を使用することができる。

E9. I can communicate effectively with students, parents/guardians, other teachers, administrators and community stakeholders about student learning.

私は、学生(生徒)の学習について、彼らだけでなく、親や保護者、他の教師たち、学校の管理職、そして地域の利害関係者と効果的にコミュニケーションができる。

E10. I can make tests that resemble real-life use of English.

私は現実生活における言語使用に似せたテストを作ることができる。

PRACTICE

P1. I use various assessment methods.

私は様々な評価方法を使っている。

P2. I use the 5-step process (plan, develop, review, field test, review, and revise) for assessment development.

私は評価方法を開発する際に、5段階のプロセス(計画、開発、チェック、実施、改良)を行っている。

P3. I use learning targets to guide instruction.

私は学習目標に沿った指導を行っている。

P4. I use assessment results to make appropriate instructional decisions.

私は、評価結果を使って適切な指導上の決定をしている。

P5. I provide timely, descriptive, and actionable feedback to students.

私は、記述によって行動指針を示すようなフィードバックを、適切なタイミングで与えている。

P6. I use grading practices that result in grades that are accurate, consistent, meaningful and supportive of learning.

私は正確かつ一貫性があり、学習に対して意義がありそれを助けるような評点の与え方をしている。

P7. I use assessment results to improve instruction.

私は、指導方法を改善するために評価結果を適切に使っている。

P8. I use multiple sources of data over time to identify trends in learning.

私は学習状況の傾向を認識するために、長期間にわたって様々な情報源を使用している。

P9. I communicate effectively with students, parents/guardians, other teachers, administrators and community stakeholders about student learning.

私は、学生(生徒)の学習について、彼らだけでなく、親や保護者、他の教師たち、学校の管理職、そして地域の利害関係者と効果的にコミュニケーションを行っている。

P10. I make tests that resemble real-life language use.

私は現実生活における言語使用に似せたテストを作っている。