Teacher Cognition of EFL Assessment: A Case Study of Professional Development on Performance-based Language Assessment in Japan

BORDIN CHINDA*
Faculty of Humanities, Chiang Mai University, Thailand
DON HINKELMAN
Faculty of Humanities, Sapporo Gakuin University, Japan
Corresponding author email: bordin.chinda@cmu.ac.th

Article information	Abstract
	This qualitative study, conducted in Hokkaido, Japan, concerns the investigation of the teacher cognition and practice of English as a Foreign Language (EFL) assessment and the impact of a professional development (PD) program on the participants. The PD program was carried out as a series of seven in-service workshops with five native speakers of English teaching in Japanese schools and universities. The workshops aimed to provide these teachers with a basic theoretical and practical understanding of performance-based language assessment. Another major purpose of the PD was for the teachers to conduct research and publish a paper concerning language assessment. In the study, the recording of teacher reflections was used as the data collection method. From the analysis of the data, the findings show that the PD program has positively affected the teachers. They have learned about alternative ways to assess the students, gained greater self-reliance, and became confident as teacher-assessors.

INTRODUCTION

In applied linguistics, it has been noted that performance-based assessment done in a classroom is a complex operation filled with uncertainty that offers greater trustworthiness and ethical potential than standard testing of knowledge (Lynch, 2001) since performance-based assessment investigates developmental sequences in student learning and assesses candidates' ability to perform specific tasks, such as second language performance, to measure productive language skills in real-world contexts. Performance tests differ from traditional tests by having a performance by the candidate, being judged using agreed criteria, and ensuring authenticity of assessment tasks (Chinda et al., 2022). When it is implemented, teachers have to both teach and assess students, especially in ESL/EFL contexts. In these contexts, many formats and tasks are used to collect information about the abilities of the students apart from traditional closed-item tests. However, for language teachers, testing and assessment have been considered the somewhat obscure domain of the "expert", with marginal relevance to regular classroom issues (Brindley, 2001). Teachers have always seen this standardized form of testing and assessment

as their "enemy" or something to be taken care of by the testing experts (Hamp-Lyons, 2003).

A further significant consideration is the availability of teacher professional development (PD). Crandall (2000) points out teaching should be viewed as a profession where the role of teachers is to develop theory and manage their own PD through sustained in-service programs. In addition, Crandall and Finn (2014) affirm that attending local, state, national, or international workshops or seminars, some offered online, is one of the best practices to stay current. These PD programs offer a platform for discussing issues and possible solutions, as well as strategies for enhancing one's teaching style or locating potential topics for in-class research. Malone (2008) agrees that language teachers should participate in regular in-service training to supplement the pre-service teacher training programs because ongoing PD can keep teachers informed of recent advancements in language testing and provide them the opportunities to integrate new techniques into their instruction.

Therefore, the present study initially aimed at providing a series of in-service PD in language assessment for EFL teachers teaching in Hokkaido, Japan, who may not have a background in language assessment, but were responsible for assessment in the classroom and institutional levels. One of the major objectives of the PD was to create a positive impact on the teachers who participated in the program. In order to understand the impact of the PD, this qualitative study investigated "teacher cognition" (Borg, 2006), that is how each teacher viewed language assessment, conducted the assessment in the classrooms, and thought about the impact of the PD.

LITERATURE REVIEW

Professional development and language assessment

As far as a professional development (PD) program is concerned, Guskey (2000) points out that PD includes the processes and activities which are designed to enhance the professional knowledge, skills, and attitudes of educators so that they might, in turn, improve the learning of students. Furthermore, Guskey explains the following characteristics of PD:

- PD is an *intentional process* designed to bring about positive change and improvement, and guided by a clear vision of purposes and planned goals;
- PD is an *ongoing process*, a job-embedded process, in which educators at all levels must continuously learn throughout the entire span of their professional careers; and
- PD is a *systemic process* that considers change over an extended period of time and takes into account all levels of the organization.

Borg (2018) proposes that when investigating the impact of a PD on teachers, the researchers can examine the following topics: satisfaction, language proficiency, teacher knowledge, instructional skills, attitudes and beliefs, classroom practice, and reflective competence. In the study, "impact" (also known as "washback") refers to individual change in these skills, attitudes, and practice as noted in teacher reflections.

In developing PD programs in language assessment, Brindley (2001) emphasizes that a PD

program in language assessment should involve the whole system, capitalize on existing practices, recognize and deal with the reality and constraints influencing teachers' assessment practices, encourage a research orientation to PD, and plan for change. He also points out that the implementation of the PD program could be done in a modular fashion, in the form of a short course, series of seminars/workshops, or individual seminars/workshops. Similarly, Malone (2008) proposes that the first step in implementing a PD program is to determine what teachers need to know about language assessment in order to perform their jobs, and secondly, to determine how to provide such training. She also stresses that it is very crucial to identify gaps: "what do instructors know about assessment, what do they need to know and how can this information best and most effectively be shared?" (p. 237). Malone adds that the major goal of training in language assessment is to empower language teachers. In addition, the training will "improve the language assessment being conducted and promote positive washback to teaching and learning" (p. 237).

With available resources, especially textbooks, implementing training in language testing and assessment should be more practical for language educators. When teacher-assessors receive adequate preparation, the teachers will have opportunities to "critique their position in the education society, identify points of opportunity and mechanisms to influence education planning, including assessment, and to find ways to contribute to positive change" (Hamp-Lyons, 2007, p. 492). When teachers get together, arguments, understandings, clarifications, and interpretations are constructed through discussion with other teachers (Mann, 2005). From his study about the impact of a PD on EFL teachers' cognition in Thailand, Chinda (2009) found that a PD program could help teachers to become more self-consistent when rating their students' performances (teachers' practices) and to be more critical to the assessment being implement (teacher's beliefs, knowledge and attitudes).

Teacher cognition

One crucial domain in language education which has been fairly well established in the field of applied linguistics is language teacher cognition. Borg (2006) reviews more than 180 studies in teacher cognition in the areas of first, second and foreign language contexts published between 1976 and 2006. He has found that research in teacher education can benefit greatly from focusing on the content, structure, and development process in language teachers' cognition. He points out that the studies of teachers' cognition should include what teachers at all career stages think, know, or believe in respect to any part of their profession, which may also involve the study of real classroom practices and the connections between cognitions and these behaviors. In other words, teacher cognition includes teachers' beliefs, knowledge, attitudes and practices.

In this paper, the definitions of the terms "belief" and "attitude" proposed by Dörnyei (2005) are adopted. Dörnyei (p. 214) defines attitudes to have "a stronger factual support [than beliefs]" whereas beliefs "are more deeply embedded in our mind and can be rooted back in our past or in the influence of the modeling example of some significant person around us". In addition, Pajares (1992) has made a clear distinction between beliefs and knowledge. He states that beliefs are "based on evaluation and judgment" and knowledge is "based on

objective fact" (p. 313). He also adds that beliefs are "an individual's judgment of the truth or falsify of a proposition" (p. 316).

Based on numerous studies on teacher beliefs, Kagan (1992) points out that greater attention to the social and institutional contexts of classrooms is required in studies of what language teachers do. She also proposed that further research into the process of transformation of language teachers' cognitions and practices as they accumulate experience is required, in addition to the study of cognitions and their patterns amongst groups of teachers working in a similar context. Moreover, she concludes teachers' beliefs are complex and tend to be linked to a consistent teaching style that is frequently visible across classes and grade levels, and they appear to be rather stable and resistant to change. In addition, a lot of what instructors understand or know about their profession is implicit. Further, beliefs are not just attitudes or knowledge but are a form of literacy that is practice-based. Language assessment literacy can be defined as "the knowledge, skills and principles" that teachers must master in order to carry out assessment tasks (Inbar-Lourie, 2017, p. 257).

Furthermore, in an in-service context, Woods (1996) explains that the constructs of beliefs, assumptions, and knowledge (BAK) are interwoven and integrated. In other words, they are points on a spectrum of meaning rather than distinct concepts (Borg, 2006). Woods (1996) stresses that these constructs affect the decisions a teacher makes in interpreting events related to teaching. From his qualitative data collected from eight teachers in Canada, he concludes that BAK evolves in the face of conflicts and inconsistencies, and gains depth and breadth as varied events are interpreted and reflected upon (p. 212).

In terms of beliefs in assessment, Büyükkarci (2014) investigated the formative assessment of sixty-nine primary school English teachers in Turkey. The data from the questionnaire survey revealed that the teachers had positive attitudes about formative assessment since they believed that formative rather than summative assessment should be implemented in their classrooms. However, it was found that these teachers could not sufficiently implement formative assessment because the large class size and the time-consuming nature of employing formative assessment. In contrast, Chinda et al. (2022) examined washback of performance-based language assessment in Japan. From their qualitative data, they found that teachers in the study implemented formative assessment because they believed that it could enhance students' communicative skills. In addition, the teachers thought that rubrics were crucial in the assessment process. Likewise, Hinkelman and Cotter (2018) used rubric formats, and addressed the significant issue of time-consuming performance assessment by designing a video-based e-learning plugin for peer, self and teacher assessment, in Japan.

RESEARCH METHODOLOGY

Research contexts

According to Borg (2006, 2015), in order to deeply comprehend the relationships between PD and teachers' beliefs, qualitative research studies are preferred over large-scale quantitative studies. The purpose of the present qualitative research was twofold: (1) provide the participants with a series of seven PD workshops offering them fundamental concepts of performance-based language assessment and washback theory, and (2) to investigate the impact of the PD on the participants. The study was conducted in Hokkaido, Japan. Since the present study was conducted during the Covid19 pandemic (December 2020 to March 2021), the researchers were restricted to seven online workshops, 90 to 120 minutes each, conducted via the Zoom application by the main researcher. The topics of the workshops included basic considerations in language testing and assessment, performance-based language assessment, assessment for learning, washback and impact in language testing, and qualitative research in ELT. More importantly, these workshops aimed to provide the participants with opportunities to cooperate in conducting a research project, on washback, for publication (see Chinda et al., 2022).

For the impact of the PD on the participants, the research questions of the study included:

1. How do teachers view language assessment?
2. What are teachers' practices in assessment?
3. How have the workshops affected the teachers' views on language assessment and their roles as assessors?

Participants

There were five teachers, native speakers of English who have been teaching in Japan, contacted to participate in the study. They learned about the current project from a presentation given by the main author. In the paper, the names and genders of the teachers were anonymized for confidentiality purposes; that is, Teacher 1, Teacher 2, Teacher 3, Teacher 4, and Teacher 5, and the pronoun "he" is used for all teachers. Teacher 1 has an undergraduate degree in music and Italian, a Master of Music, and a TESOL Certificate. He began teaching English in Japan twenty years ago. Teacher 2 holds a Graduate Diploma in Teaching (Primary) and is presently completing his Master's in Education. He is currently teaching at a junior high school in Sapporo. Teacher 3 has a BA in Psychology (major) English (minor) and at a teachers' college and was certified as an elementary and middle school teacher in Ontario, Canada. He has been teaching English students from elementary to high school as well as adults at universities for 25 years. Teacher 4 holds undergraduate B.A and B.Sc. degrees, a Diploma of Primary Teaching, and an M.Ed. in TESOL. He has been teaching in Japan from young learners through to tertiary level for the last 20 years in Hokkaido. Finally, Teacher 5 has been teaching university EFL in Japan for 35 years. His studies included an undergraduate degree in cultural anthropology, a Master's in Education, and a doctorate in applied linguistics.

Data collection and analysis

Borg (2015) points to five major data collection methods for investigating teacher cognition: oral accounts, self-report instruments, observation, written accounts, and visual accounts. To investigate the impact of a professional development program, Borg (2019) also offered similar methods to collect the data. Based on these recommendations by Borg (2006, 2015, 2019), the present study employed a self-report method instead of a semi-structured interview to lower the Hawthorne effect and the social desirability bias of the participants which could hinder the trustworthiness of the study. The participants were requested to make self-reports on their thoughts about their views on language assessment, practices in performance-based assessment, and opinions about the PD workshops. They made the audio-recorded reflections three times following the guided questions in the guidelines (see the Appendix for the guidelines of each reflection): the first one prior to the first workshop, the second after the fifth workshop, and the last one when the first draft of the research paper was completed (see Chinda et al., 2022). Each recording lasts from five to fifteen minutes. The recordings of the reflections were transcribed and analyzed qualitatively by coding.

For the analysis of the teachers' reflections, the first author did the first coding, and the coding was verified by the second researcher. To ensure trustworthiness, the analysis of each teacher was emailed to the teacher for verification and changes. This analysis followed the grounded theory guidelines for doing qualitative content analysis, using open coding and axial coding (Corbin &Strauss, 2008).

FINDINGS

In order to have a well-rounded and comprehensive understanding of each teacher's cognition, the perspectives held by each teacher, based on the data analysis, are reported under three thematic categories. The subsequent charts report the examination of the perspectives of the teachers on assessment via the lens of three overarching themes, in alignment with the research questions of the study: 1) Views on assessment, 2) Teacher's practices in assessment, and 3) views on and impact of PD. Additionally, the charts incorporate summaries and quotations derived from the reflections of each teacher.

Views on assessment

In order to discern the five teachers' assessment literacy, the following "views" on assessment, or "teacher cognition", were reported and summarized by two researchers in Table 1.

Table 1

Teacher views on assessment

Beliefs, statements, and comments by teacher-participants

Teacher 1

- Feedback is important as it encourages students' learning.
- "Students need to know how they are doing and what kind of progress they are making to encourage them to keep them going."
- The teacher's responsibility is to provide students with feedback as it encourages students' learning
- The formal levels of large-scale testing are useful and the external reputation of these tests do encourage students to study for the tests
- "I find students really enjoy working together to perform something."
- Telling students only their good points is "one of my weaknesses" and attempts to "give them my honest ideas, but I always have to cushion it between compliments."
- Rubrics are an important assessment tool to guide students' learning.
- He believes in doing performance-based assessment because of the encouraging results.
- Some students have trouble with performance-based tasks and he may have to reduce the number of such assessments.

Teacher 2

- He sees two types of assessments: 1) knowledge itself and then 2) usage of the knowledge (or performance assessment).
- He thinks assessment should be authentic, performance-based language assessment and feels that Japan pays too much attention to traditional testing.
- He views himself as being strict about knowledge-based tests. However, he thinks he is changing as he wants to help students with the performance tasks.
- Large-scale testing (including performance-based assessment in the Cambridge tests) does not help student learning because students do not receive sufficiently targeted feedback from the test; they only know the scores, which they might not understand.
- He realizes that assessment has to be authentic (real-life situations) to be meaningful for students.

Teacher 3

- Assessment is an important diagnostic tool for students.
- He believes that a rubric helps guide students through a performance-based assessment.
- He is aware of alternative ways to assess students, but Japan focuses on traditional testing.
- Teacher's autonomy is very important in designing assessment. He enjoyed the course he was teaching because he had freedom with the materials and assessment.
- He wants to incorporate the student into as much of the assessment process as possible so that both the teacher and the student were providing feedback.
- He tries to get students actively involved as an agent in talking about the learning. He liked when students do self-assessments.
- He likes to incorporate student goals and try to put those into course design.
- He prefers the student to be able to negotiate their marks.

Teacher 4

- His views on assessment have changed as he has been given more responsibility and autonomy as a teacher to both design and administer assessments.

- He "loved being involved" in 'assessment for learning' classroom research projects, and through that experience, he believes that self and peer assessments are beneficial for students in a classroom.
- He thinks that Japan focuses too much on large-scale traditional testing, which is largely based on grammar-translation approaches as well as a reading-listening emphasis.
- As a paid professional test examiner in his spare time, he has become experienced with a variety of testing formats. In his evaluation, one particular commercial test was "way higher [quality] than other tests" but faces issues with implementation and logistics in Japan.
- Fairness in testing means that students "know how they'll be assessed before they take part". This is what he defines as a very important quality of assessment.
- Feedback, especially positive feedback, is an important part of the assessment for learning and encourages learning.
- In Japan, teachers focus too much on negative feedback.
- He dislikes traditional testing and prefers performance-based assessment.
- Both teaching, assessment tasks, and rubrics have changed because of the online situation. Points should be given for preparation work, not just the final performance work that a student does.

Teacher 5

- Teacher 5 viewed his role as a motivating teacher who used assessment to motivate students' learning.
- He believes in assessment for learning—formative assessment over summative assessment.
- Assessments motivate students' learning when students get involved in the assessment process.
- Japan, as a test-oriented country, focuses mainly on testing receptive skills in large-scale settings.
- He believes that performance-based assessment should be implemented, and teachers' professional development could help change the current practice.
- The dominant learning approach should move to task-based learning and project-based learning as they could lead to authentic assessment in real-world situations.
- Colleagues are passively against performance-based assessment because it is "messy" and "unreliable", and only trust commercial traditional testing.
- The use of rubrics as self-assessment is a powerful part of the learning process.
- Teachers need to teach students how to use the rubrics and teachers need to learn how to write rubrics for students (avoiding holistic scales and emphasizing analytic scales with specific can-do scales).
- He realizes that his self-assessment criteria can be classified as a "checklist", not a "rubric".
- A checklist was more suitable for self- and peer assessment.

Teacher's practices in assessment

The second theme of assessment literacy, "Teacher's practices as an assessor", was investigated by asking questions to the five teacher-participants with self-recorded interviews. The teachers' replies on roles of assessment were reported and summarized by two researchers in Table 2.

Table 2
Teacher practices in assessment

Beliefs, statements, comments by teacher-participants
Teacher 1
• Despite viewing himself as a capable teacher, he feels that giving students grades with performance-based assessment is difficult.

- He did not follow the rubrics strictly or uniformly as he also wanted to grade students based on their improvement. He thinks flexible rubrics would be helpful.
- He provides immediate feedback because he believes that it could encourage students to perform better.
- His role is to encourage students to use L2 while doing performance-based tasks.
- He does all speaking assessments instead of focusing on listening or reading.
- He could not monitor students well in a Zoom online classroom or in a break-out room.
- Monitoring students in breakout rooms on Zoom is difficult, and without teacher's presence, many students go back to using L1.
- He regrets he could not do his paper-based pop quizzes because of online learning, he did not have time to convert the quizzes online.
- His job is to design the final assessment task and prepare them for it.
- He said he enjoys teaching but does not want to assess students too severely.
- He uses assessment as a diagnostic tool.
- When giving feedback, his role is to give immediate, positive comments.

Teacher 2

- Teacher 2 is only familiar with traditional testing. So he has not implemented any performance-based language assessments.
- Designing assessments will be a new role for him. that he is willing to try. In the past, he has focused on preparing students for the tests.
- A barrier to designing assessments is that the textbook he has been using does not facilitate performance-based assessment because the context is unsuitable for students.
- He would love to assess students more than one time at a task.
- He recalled a bad experience when he gave a whole task assessment in a writing exercise at his high school. Unfortunately, multiple students could not complete the writing assessment and therefore got zeros. Consequently, he returned to using multiple choice questions, with which students were more familiar.

Teacher 3

- Teacher 3 tried to walk the line between meeting the institution's needs and the students' needs when it came to assessing their work.
- He modelled performance-based assessment, as a form of scaffolding with practice activities related to making movies, staging skits and role plays, and interactive presentations.
- He stressed that the audience must be involved from the get-go. But he was unsure whether he was giving too much scaffolding.
- He was positive and confident with how he assessed the students because he could guide the students to achieve the tasks. He was satisfied "as long as students can specify and develop their learning goals and understand that in my class they're going to build a learner's resume of things that they'll be able to do and demonstrate upon completion of our class."
- He helped students set short- and long-term English learning goals for the class's assessment and future careers.
- He used a rubric to guide students through the assessment. However, but the students did not use the rubrics. He believed that an assessor needed to be fair and found himself being a fair assessor when he got students involved in the assessment process and got them to do more peer assessment using the co-developed rubric.

Teacher 4

- Teacher 4 realized that the teacher's roles as an assessor is being recreated through performance-based assessment, which is completely different than knowledge-based testing.

- He realizes that giving feedback is time-consuming, but crucial in the assessment process, stating "I just don't have time to give proper feedback … That kind of comes down to workload and that's the thing. but I like to devote time to giving good feedback to students. If I can."
- He has not had problems with implementing performance-based assessment himself, but problems collaborating with colleagues who refused to adopt it.
- He also has his students do online self- and peer-assessment based on the online rubric. He uses rubrics and guidelines to provide feedback (by teacher and students).
- His role changed from when teaching was face to face. Before he could do a game, going over vocabulary … where they'd "get into groups and have races against each other, answering the questions. But this year we couldn't really do that on zoom, so we pretty much got straight into a PowerPoint."
- He usually focuses on assessment for learning through group discussion and asking questions, but he has found that it is very difficult to do in the online situation.
- He uses rubrics to assess performance-based assessment and provide immediate feedback. The amount of feedback depends on the class size.
- He points out that to motivate students to do performance-based tasks online he shows them the rubric with extra percentage points for students who do the tasks. He believes that the extra credits would motivate students to do the task.
- His job is to "devote more time to feedback. I think if I could, I would like to find ways in making that easier because it is a big-time … part of the teachers' workload to be giving good quality feedback."
- His reluctance to give good feedback depends on whether the students will act on the feedback immediately in re-doing the task.
- He is planning to continue employing task-based and performance-based assessment as well as doing research on it. He also plans to do more group tasks because they are authentic (real world) both in school and in future career situations.

Teacher 5

- Because he has research funds, he develops open source LMS plugins for self- and peer-assessment. He uses these video assessment plugins in class to support blended assessment (online commenting and face-to-face feedback combined).
- He trains other teachers in performance assessment at his school, his teachers' association, and nationally in conferences.
- He does a warmup role to encourage students to perform in class by asking questions to their partners (asking questions in casual small talk) as a lead-in activity.
- He designs major assignments which rely heavily on the rubrics for assessment. He teaches the rubrics to the students using L2 while demonstrating the assignment samples.
- Because of the pandemic, he found that he could not monitor students nor provide sufficient feedback. Teacher 5, therefore, tried to do more self- and peer- feedback instead. However, he noted that with online teaching, some students ignored self- and peer-assessment whereas all students would have done them with pre-training in a face-to-face setting. He also focused on the actual performance rather than grammar and vocabulary when employing performance-based assessment.
- He wishes he could do a standardization workshop with his colleagues focusing on assessment aims and the materials development for the course.

Views on and impact of PD

The third theme of assessment literacy, "Views on and impact of PD.", was also investigated by asking questions to the five teacher-participants within the same self-recorded interviews. The teachers' replies on the impact of professional development of assessment were reported and summarized by two researchers in Table 3 based on the criteria of satisfaction, language proficiency, teacher knowledge, instructional skills, teaching beliefs, classroom practice, and reflective competence (Borg, 2018).

Table 3
Views on and impact of PD

Beliefs, statements, comments by teacher-participants
Teacher 1
• Teacher 1 realizes that he needs to do research and publish. He views this classroom research workshop as important PD especially since the pandemic has reduced available PD conferences.
• He expected to publish a research paper, and was able to achieve that.
• He also would love to do more research. He especially enjoyed working with other teachers but found the online format of the workshop a disappointment.
• As for the impact of the professional development, he has learned about washback. Learning about washback also affected how he views assessment and how he would conduct classes. "I have started to view language assessment differently. I'm thinking more these days about how my assessments affect how I teach and affect my students, the way they prepare and what they do in the class."
• He has also learned from the PD about rubrics which will help solve his problem and he plans to study more about washback research.
Teacher 2
• Teacher 2 did not have any expectations and participated in the PD because he was invited.
• He found the PD beneficial because he has learned much about assessment for learning (formative assessment).
• Reflecting on the impact of this professional development, he compares himself with other members and feels that he is the least experienced teacher and therefore needs to improve.
• After learning about washback and assessment for learning, he plans to get students involved in the assessment process to help them succeed.
Teacher 3
• Teacher 3 was interested in professional development, working with other teachers, and learning from each other: "I was able to work and share and learn from other teachers and to encourage and feel like I can keep in touch with these teachers and continue to share."
• He did not like the fact that the PD was conducted online (via Zoom).
• He found that doing research (as part of the PD) was beneficial as he has learned about other teachers, students, and washback.
• He felt that the publication of the research project could have an impact on students and other Japanese.
• He became more interested in performance-based assessment as he believed that this type of assessment was meaningful for students.

Teacher 4

- He explained that he took part in this professional development research group because of his personal connection with the main investigator.
- He was motivated to join because the topic of the PD is related to his interest in performance assessment and his ongoing conference presentations on peer and self-assessment used in video assessment.
- He thought that the participants could have learned more if it had been done face-to-face.
- At first, he "couldn't really see … what we were trying to research and what we were doing, but then… and I think part of that was because of being online …, but then, especially culminating with the presentation, it was like, oh, this is what we're doing."
- He was fascinated by the student interview data as he has learned that the assessment tasks motivated students to learn.
- Though he has other areas of interest, he would like to explore more in this area of research.

Teacher 5

- Teacher 5 believed that he could learn more about language assessment from this professional development workshop than by research article reading.
- He believed that a PD approach was ideal for conducting a research study with his colleagues. He also participated in this PD because he wanted to work with the main researcher who specializes in language assessment.
- He thought that the teachers who participated in the program are a productive team of both novice and experienced researchers.
- This PD program had an impact on a local teacher's organization which forced accountability and focus to the research. The same PD teacher's organization also provided a platform to host two online presentations concerning the present study which attracted over twenty local teachers to attend.
- Teacher 5 disliked the decontextualized online communication as it hindered participant bonding and community-building that would happen through participants meeting in person.
- He learned from the PD that rubric and checklist assessments are different. A checklist is more suitable for his students because a general rubric would be "too abstract" for the students to understand.
- He pointed out that the PD has confirmed his views about the benefits of formative assessment and assessment for learning.
- He would like the team to continue PD by doing more research, attending conferences, and publishing research papers.

DISCUSSION

In terms of teacher assessment cognition, the above data illustrates that teachers' beliefs about assessment for performance-based learning gradually changed, especially in relation to their views on 1) assessment for learning, 2) task design, and 3) standardized testing. Firstly, from attending the PD workshops, the participants realized the benefits of assessment for learning or formative assessment, though some of them did not use these terms in the first reflections. According to Black and Wiliam (1998, 2009), the assessment includes all activities carried out by teachers and students in assessing themselves that provide information for teachers to use as feedback to modify the teaching and learning activities in the classrooms. The assessment becomes formative when the evidence is used to adapt the classroom practices to meet the needs of the students.

All teachers in the study stood firmly in the formative purpose of assessment, that assessment is for learning, rather than for educational administration. One indication of that is that terms of feedback and assessment were constantly mixed. All of the teachers did not separate feedback from assessment as they were aware of the significant roles of feedback plays in students' learning. For instance, a teacher would say that feedback is essential to students and wonder how to give the best feedback, most timely feedback, and the most frequent feedback/ assessments. Similar to Büyükkarci (2014)'s findings, one teacher stated that because of the time-consuming nature of giving feedback, he could not provide sufficient feedback because of his class size and workload, though he believes that immediate feedback is beneficial for students' learning.

It was also found that even prior to the present study, the teachers had already employed formative assessment principles when they designed the assessment tasks. These concepts are not unlike the ten principles for formative assessment (Broadfoot et al., 2002) set by the Assessment Reform Group. For example, they assert that formative assessment should be part of the effective planning of teaching and learning, focus on how students learn and their motivation, and enable students to receive constructive feedback on how to improve. From the reflections of the participants, it can be concluded that task design resulting in student enjoyment was a base of motivation for assessment. The teachers described their own motivation rising when students enjoyed the task, particularly when it was a collaborative task. While the process of enjoyment is sufficient for a teacher to continue and repeat a task, the success of the task was more important. For the students to believe in their own success, the teachers pointed out that giving students specific verbal feedback and written assessment scores/ comments gave students greater confidence. The teachers believed that this could make students feel that their own development was real and measurable, and not just a temporary good feeling.

Furthermore, as teachers play an indirect role as consumers of testing programs by large-scale testing bodies (for example, Eiken, TOEIC, Cambridge, IELTS which are generally viewed as high-stake testing in Japan), these participants felt overwhelmed and humbled by the power of traditional testing on their students. They denigrated their own intuitions on assessment, by assuming large-scale testing standards are dominant and irrefutable. This agrees with Chinda (2014)'s earlier study, which was carried out in Thailand, in which the participants were not satisfied with the classroom assessment formats, which were mandated to the teachers by the department. Here, the teachers in the present study, through sharing of practice and group reflection, gradually put their faith in alternative assessment, especially formative classroom-based assessment as they designed their own assessment instruments in contrast to traditional testing. Traditional testing emphasizes "the rank ordering of students, privileges quantifiable data for isolated, individual test performances"; whereas alternative assessments are often based on "an investigation of developmental sequences in student learning, a sampling of genuine performances that reveal the underlying thinking processes, and the provision of an opportunity for further learning" (Lynch, 2001, pp. 228-229).

Similarly, in an extensive review of testing literature, Abu-Alhija (2007) concluded that teachers are aware of both the positive and negative consequences of large-scale testing. The examples

of positive consequences of large-scale testing are promoting a more accurate assessment of each student's needs, assisting in identifying the curriculum's strengths and shortcomings, allowing students and instructors to better understand their own strengths and limitations, and encouraging teachers to connect instruction with standards. In terms of negative consequences, large-scale testing could encourage teachers to spend more time teaching to the tests than for the curriculum requirements, influence them to prepare for tests improperly, diminish their sense of professional worth, and persuade them to give advantage to some students when administering the exams, increase their stress, and decrease their morale. In this study, similar negative influences of standardized testing were acknowledged by the teachers, and self-designed performance tests were strongly preferred over those "professional" tests. Therefore, it can be stated that because the teachers distrusted the role of the large-scale traditional testing culture and its negative consequences, they relied heavily on alternative ways of assessing their students. For instance, they used feedback and assessment that they conducted spontaneously as they worked through classroom tasks because they believed that these could help motivate and improve students' learning.

Furthermore, in this study, research and PD are mixed and while the process of research is assumed to be the base, there was significant professional development happening naturally as a result. For example, Teacher 4 was fascinated by the student interview reports, and from student enthusiasm, learned about what forms of tasks and assessment students valued. Teacher 5 felt the impact of this PD was foremost related to self- and peer-assessment, and students would learn most if they understood the rubrics and could describe the specific points to watch for and reward. In other words, the level and quality of teacher assessment literacy are related to the reflective processes used in both the classroom research inquiry and the professional development process. According to Burns and Khalifa (2022), it is possible that the reflection required in research approaches like action research may cause important long-lasting, practice-based development in teachers involved in the research.

In a Japanese university context, teachers are expected to design all classroom assessments and assign a grade to each student's class performance. Therefore, these teachers were adept at making their own assessment criteria, feedback, assessments or exams, and employing self- and peer-assessment, but they were also searching for alternative ways of assessing, that co-teachers in the study could provide. In addition to developing assessment formats, rubrics, and rating systems, Teachers 3 and 5 developed online plugins to manage the recorded videos of students, which means the teacher role is evolving to include technology development in making the plugins as teacher-programmer teams who design and use open-source plugins. Thus, as Taylor (2013) predicted, language assessment literacy is evolving in scope, requiring new dimensions of knowledge and skill.

CONCLUSION AND IMPLICATIONS

This qualitative study examined the views of five EFL teachers on language assessment, their roles as assessors, and the impact of the PD in performance-based language assessment on teachers in Hokkaido, Japan. The teachers' self-reports revealed that the five teachers held

consistent beliefs in the superiority of performance-based assessment for student motivation and a more trustworthy evaluation of student progress. In addition, teachers reported positive washback in individual teacher motivation. However, teachers still experienced doubts about their own ability to assess, and a deference to experts in the field of standardized testing.

Concerning the implications, the findings in this study suggest current assessment literacy research and practice require institutional and policy changes to produce positive impact effects within language learning programs.

1. For greater teacher motivation and more trustworthy assessment formats, learning institutions can emphasize teacher-designed assessment and de-emphasize global or nationally-designed testing instruments for assessment.

2. Assessment activities can prioritize task performance (the use of skills) rather than the explicit testing of retained knowledge among students. In EFL learning contexts, knowledge of grammar and vocabulary can be tested "in-use" within performances of authentic and real-world tasks and projects.

3. Self and peer assessment force a different form of checklists and rubrics than teacher-only criteria. Specific, analytic points are easy to understand and can increase learner cognition and learner motivation as students learn the qualities of high-level performance.

4. Teacher cognition research requires qualitative methods which encourage teacher reflection, especially collaborative reflection. Teachers involved in this study learned from each other during the process of research through discussion. In interviews, the reflective questions were open-ended

5. Performance-based assessment is eclectic, requiring personal topic choices and individualized performances instead of "answers". With this assessment design, expensive technology to prevent cheating on knowledge-based testing may become less necessary. Reliance on collections of learner performances and assessments via e-learning platforms may increase.

ACKNOWLEDGMENTS

THE AUTHORS

Bordin Chinda, PhD, is an Assistant Professor at Chiang Mai University and a visiting professor at Sapporo Gakuin University (while conducting this study). His research interests include performance-based language assessment, washback studies, professional development, and language teacher education.
bordin.chinda@cmu.ac.th

Don Hinkelman, PhD, is a Professor teaching intercultural communication skills at Sapporo Gakuin University. He has researched performance assessment, co-developed the video assessment module for Moodle, and written on assessment theory in "Blending Technologies in Second Language Classrooms".
hinkelman@mac.com

REFERENCES

Abu-Alhija, F. (2007). Large-scale testing: Benefits and pitfalls. *Studies in Educational Evaluation, 33*(1), 50-68. https://doi.org/10.1016/j.stueduc.2007.01.005

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educ Asse Eval Acc, 21*(5). https://doi.org/10.1007/s11092-008-9068-5

Borg, S. (2006). *Teacher cognition and language education*. Continuum.

Borg, S. (2015). Researching language teacher education. In B. Paltridge & A. Phakiti (Eds.), *Research methods in applied linguistics: A practical resource* (2nd ed., pp. 541-560). Bloomsbury.

Borg, S. (2018). Evaluating the impact of professional development. *RELC Journal, 49*(2), 195-216. https://doi.org/10.1177/0033688218784371

Borg, S. (2019). The impact of language teacher professional development. In G. Barkhuizen (Ed.), *Qualitative research topics in language teacher education* (pp. 44-49). Routledge.

Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (Vol. 11, pp. 137-143). Cambridge University Press.

Broadfoot, P., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles*. University of Cambridge School of Education.

Burns, A. & Khalifi, H. (2022). Engaging action research to explore use of assessment for improving language ability. In A. Burns & H. Khalifi (Eds), *Second language assessment and action research* (Vol. 48, pp. 1-16). Cambridge University Press.

Büyükkarci, K. (2014). Assessment beliefs and practices of language teachers in primary education. *International Journal of Instruction, 7*(1), 107–120. https://eric.ed.gov/?id=EJ1085246

Chinda, B. (2009). *Professional development in language testing and assessment: A case study of supporting change in assessment practice in in-service EFL teachers in Thailand* [Doctoral dissertation, University of Nottingham]. Nottingham eTheses. https://eprints.nottingham.ac.uk/10963/

Chinda, B. (2014). Teachers' reactions towards performance-based language assessment. *PASAA: Journal of Language Teaching and Learning in Thailand, 48*, 57-88. https://eric.ed.gov/?id=EJ1077896

Chinda, B., Cotter, M., Ebrey, M., Hinkelman, D., Lambert, P., & Miller, A. (2022). Reactions of teachers and students towards the implementation of performance-based language assessment: A washback study in Hokkaido, Japan. *LEARN Journal: Language Education and Acquisition Research Network, 15*(1), 524-547. https://eric.ed.gov/?id=EJ1336154

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Sage.

Crandall, J. (2000). Language teacher education. *Annual Review of Applied Linguistics, 20*(1), 34-55. https://doi.org/10.1017/S0267190500200032

Crandall, J. A., & Finn M, S. (2014). Effective professional development for English as a second/foreign language teachers. In M. Celce-Murcia, D. M. Brinton & M. A. Snow (Eds.), *Teaching English as a second or foreign language* (4th ed., pp. 630-648). National Geographic Learning/Cengage Learning.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition.* Lawrence Erlbaum.

Guskey, T. R. (2000). *Evaluating professional development*. Corwin Press.

Hamp-Lyons, L. (2003). Writing teachers as assessors. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 162-189). Cambridge University Press. https://doi.org/10.1017/CBO9781139524810.012

Hamp-Lyons, L. (2007). The impact of testing practices on teaching: Ideologies and alternatives. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (Vol. Part I, pp. 487-504). Springer. https://doi.org/10.1007/978-0-387-46301-8_35

Hinkelman, D., & Cotter, M. (2018). Balancing real-time vs. post-performance feedback for EFL presentation classes. In P. Clements, A. Krause & P. Bennett (Eds.), *Language teaching in a global age: Shaping the classroom, shaping the world* (pp. 369-380). JALT. https://jalt-publications.org/articles/24396-balancing-real-time-vs-postperformance-feedback-efl-presentation-classes

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. Or & S. May (Eds.), *Language testing and assessment, Encyclopedia of language and education* (3rd ed., pp. 257–270). Springer. https://doi.org/10.1007/978-3-319-02261-1_19

Kagan, D. M. (1992). Implications of research on teacher belief. *Educational Psychologist, 27*(1), 65-90. https://doi.org/10.1207/s15326985ep2701_6

Lynch, B. K. (2001). The ethical potential of alternative language assessment. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Language testing in honour of Alan Davies* (Vol. 11, pp. 228-239). Cambridge University Press.

Malone, M. E. (2008). Training in language assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 7, pp. 225-239). Springer Science+Business Media.

Mann, S. (2005). The language teacher's development: State-of-the-art article. *Language Teaching, 38*, 103-118. https://doi.org/10.1017/S0261444805002867

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307-332. https://doi.org/10.3102/00346543062003307

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing, 30*(3), 403–412. https://doi.org/10.1177/0265532213480338

Woods, D. (1996). *Teacher cognition in language teaching: Beliefs, decision-making and classroom practice*. Cambridge University Press.

Appendix

Guidelines for teachers' reflection 1

Please reflect on your thoughts about language assessment by answering the following questions. Please use your mobile phone to record your reflections and email me the file (bordin.chinda@cmu.ac.th). Thank you for your cooperation. Please DO NOT study the questions before making a recording.

1. How would you describe your views toward language testing and assessment?
2. How would you describe language testing and assessment in Japan?
3. What do you think are the most important ways to assess students?
4. How would you describe yourself as a teacher and assessor?
5. Have you adopted performance-based assessment in your classes? What problems have you encountered? What are the sources of these problems? How did you learned to handle this kind of assessment?
6. How have your thoughts and feelings about assessment changed since you first taught English in Japan?
7. What advice would you give to someone who is new to performance-based assessment?

Guidelines for teachers' reflection 2

Please reflect on your thoughts about language assessment by answering the following questions. Please use your mobile phone to record your reflections and email me the file (bordin.chinda@cmu.ac.th). Thank you for your cooperation. Please DO NOT study the questions before making a recording.

1. The course under investigation:
    a. title
    b. main objectives
    c. number of students in class
    d. the status of the course (e.g. compulsory? how popular it is among students,
    e. what do other teachers think about it?, etc.)
    f. your impression about the course (e.g. do you like it? why/ why not?)
2. How you generally conducted the class (i.e. what happened in the class which may or may not be related to the assessment)
3. How you would have conducted the class differently (now that the class is over)
4. All performance-based language assessment you used in your class this semester, especially the final project
5. How you will do the assessment for the course again (next semester or next year)
6. How you prepared the students for the final project (processes both inside and outside the classroom)
7. How you would have prepared the students differently (now that the class is over)

Guidelines for teachers' reflection 3

Please reflect on your thoughts about language assessment by answering the following questions. Please use your mobile phone to record your reflections and email me the file (bordin.chinda@cmu.ac.th). Thank you for your cooperation. Please DO NOT study the questions before making a recording.

1. As an English teacher, what are your strengths and weaknesses in general, assessment, and research?
2. What made you decide to participate in the program (workshops and research). What did you expect from participating?
3. Did the program (workshops and research) meet your expectations?
4. What did you like or dislike about the program (workshops and research)?
5. Did you learn anything new from the program (workshops and research)? What are they?
6. Have you had problems assessing your students? How did you deal with them?
7. Have you started to view language assessment differently? How?
8. Have you considered doing assessment in your class differently? How?
9. Now that we have finished the program, do you have any plan for your professional development (teaching and/or research)?