

Student-Staff Partnership in a Chinese Language Class in Japan: A Case Study of Speaking Test Practice

Received: 28 March 2024; Revised: 29 May 2024; Published: 16 June 2024

Ming Qu

Muroran Institute of Technology, Hokkaido, Japan
Email: quming2014@gmail.com
ORCID: <https://orcid.org/0000-0002-4213-693X>

Abstract:

In recent years, there has been a growing emphasis on fostering collaboration between staff and students to co-design learning and teaching methods in higher education. However, the inclusion of speaking tests within student-staff partnerships remains relatively uncommon. This paper addresses this gap by presenting two main parts. First, it outlines the stages of engaging students to develop and administer the speaking test. It then analyses the validity of student assessments and their perceptions of this educational practice. Student-staff partnerships create an environment that fosters empowerment, collaboration, voice, and responsibility. Through these partnerships, students are empowered to take control of their learning journey and develop into self-directed learners.

Keywords: Chinese language teaching, assessment literacy, democratic practices, speaking test

1. Introduction

Student-staff partnership is an approach in which students and staff work collaboratively in the design of teaching & learning. According to Mercer Mapstone & Marie (2019), in conventional educational approaches, modifications to course designs typically occur post-course, incorporating feedback from students. In contrast, within student-staff partnerships, the responsibility for change is shared, which involves engaging in collaboration earlier in the design process, before finalizing any alterations.

A student-staff partnership transcends formal mechanisms of student representation and feedback by positioning students at the forefront of their educational experience. Actively engaging students in shaping the curriculum, and collaborating with staff to enhance the design, delivery, and evaluation of teaching, learning, and assessment, not only increases their learning motivation but also fosters a sense of belonging to an engaged learning community (Pascale, Steve & Mark, 2019). Furthermore, involving students in decision-making empowers them to take an active role in shaping their educational experience. For the staff, working in partnership with students can stimulate creativity, share responsibility with students, thus enhancing the quality of teaching and learning. It is one of the most effective ways to engage students, create a learning community, and ensure a better educational experience (Deeley & Bovill, 2017).

While reports indicate the implementation of some student-staff partnership practices, it is essential to recognize that there is no universally applicable approach to expanding such partnerships in education. The effectiveness of student-staff partnerships can significantly vary based on the context, institutional culture, and the specific goals and needs of the educational institution. This study focuses



on a case study examining the development and implementation of a student-staff partnership in a Chinese language class in Japan. Specifically, our attention centers on the collaboration between students and staff in designing, delivering, and evaluating a term-end speaking test in the Chinese language classroom.

This study consists of two parts, first, the stages involved in engaging students to develop and administer the speaking test were presented, then, the validity of student assessments and their perceptions of this educational practice were analyzed.

While recent years have seen an increased focus on research and practice related to staff and students collaborating to co-design learning and teaching in higher education, the inclusion of speaking tests within student-staff partnerships remains relatively uncommon. Furthermore, previous studies have focused on the effects of this educational practice in terms of its impact on cognitive domains, such as comprehension of learning content and metacognitive ability, and affective domains, such as motivation to learn and autonomous motivation when students are involved in the assessment process. This study deviates slightly from this trend by examining the validity of student evaluations and analyzing Japanese students' perceptions of student-staff partnerships. This research aims to lead readers through the diverse phases of implementing our student-staff partnership practice. Simultaneously, the objective is to address those research gaps and provide a substantial contribution to the field of Chinese language education.

2. Student-Staff Partnership in Assessment and Feedback

When considering assessments, our thoughts often gravitate toward high-stakes tests such as university entrance examinations, term-end exams, or examinations that hold significant importance, such as the TOEFL test for studying abroad. The primary purpose of these high-stakes tests is to evaluate student performance and differentiate the achievement levels. High-stakes tests can be a source of stress for some students. However, in the case of informal classroom-based tests, they not only serve as proof that learning has occurred but also offer evidence of learning, with the tests themselves becoming integral to and contributing to the learning process.

The influence of a test on student learning is referred to as the washback effect. For instance, if a reading test is administered, students are expected to engage in extensive reading practice; if a vocabulary test is given, students are anticipated to focus on studying vocabulary; and if a summary test is conducted, students are encouraged to practice summarizing texts. Additionally, offering specific instructions on how to prepare for the test can be effective, and this approach enables students to understand what and how to study, facilitating better preparation for the test. In essence, beyond being a means of judgment, the mode of the test we select not only determines what students learn but also influences how they learn. The issue revolves around how to raise students' awareness of the expected performance from their teacher and how to furnish them with specific instructions for test preparation. In this study, we aim to present an educational approach wherein teachers and students engage in collaborative test creation, namely, a student-staff partnership approach to assessment. Through this collaboration, students gain insights not only into the test format and expected performance but also into effective test preparation strategies. As a result, this process cultivates self-awareness of areas for



improvement, encourages students to think about improving their performance, and promotes independent and autonomous learning.

Empowering and engaging students through assessment design and providing opportunities for dialogic feedback are crucial elements of the learning experience. In recent years, many studies have examined the student-staff partnership approach to assessment. (e.g., Rust et al., 2003; Boud and Molloy, 2013; Handley and Williams, 2011; Orsmond et al., 2002; Mulder et al., 2014; Deeley and Bovill, 2017). The findings from these studies suggest that the partnership approach to assessment can yield advantages for staff, students, and higher education institutions. This approach, with its focus on democratizing the assessment process, not only grants students greater agency in their learning but also contributes to their assessment literacy.

As indicated in the review conducted by Deeley and Bovill (2017), previous studies suggest that students can enhance their understanding of the language of assessment through active engagement in various methods such as observation, imitation, dialogue, and practice (Rust et al., 2003). Additionally, students demonstrated enhanced assessment literacy through collaborative exploration of marking criteria with staff (Handley and Williams, 2011) and active involvement in assessment design (Orsmond et al., 2002). Attributes like the ability to make complex judgments (Sadler, 2010), self-monitoring (Boud and Molloy, 2013), and proficiency in giving and receiving criticism (Mulder et al., 2014) are highly regarded. The partnership approach is viewed as a method to enhance these skills. Boud and Falchikov (2006) additionally assert that engaging in a partnership for assessment can effectively prepare students for a lifetime of self-assessment in their learning journey.

To date, previous research on the effectiveness of educational practices that involve students in assessment has predominantly concentrated on cognitive domains, such as comprehension of learning content and metacognitive ability, as well as affective domains, including motivation to learn and autonomous motivation. This study, however, shifts the focus slightly by analyzing the validity of student evaluation itself. Additionally, it also examines Japanese students' perceptions of student-staff partnerships, a dimension not yet explored in the field of Chinese language education.

The degree of agreement between peer and teacher evaluation results has been extensively studied (e.g., Ashenafi et al., 2017; Campbell et al., 2001; Lin et al., 2001; Patri, 2002; Rudy et al., 2001; Naganuma et al., 2019; Qu, 2018). Empirical evidence suggests that the outcomes of peer evaluations cannot be uniformly considered as reasonable assessment results. This is often attributed to students not understanding the rubric or unclear of the judgment criteria. To address this issue, developing assessment tasks and scoring methods that make rating criteria more accessible to learners may lead to peer assessment results aligning more closely with teacher assessments, thereby enhancing validity. Notably, there have been no known attempts to incorporate such an approach. Therefore, this paper demonstrates the multiple stages of implementing our student-staff partnership practice in a speaking test, subsequently, we analyze and discuss the correlation between peer ratings and teacher ratings, along with Japanese students' attitudes toward student-staff partnership assessment.



3. Student-Staff Partnership Case Study

This student-staff partnership practice engaged students as co-designers throughout the assessment and feedback processes. The study was conducted at a Japanese university with second-year undergraduate students. Twenty-one students studying the Chinese language participated in the study. Assessment for this course included a term-end speaking test, accounting for 25% of the overall course grade, a one-hour paper test weighted at 60%, and the remaining 15% attributed to homework. This study focuses on how students and staff cooperate in designing, developing, and delivering the term-end speaking test. The reason we conduct the speaking test is that the score attained in a paper test may provide insights into a student's knowledge, however, it falls short in assessing their practical language usage. Especially in Japan, it's commonly acknowledged that many students possess substantial vocabulary and grammar knowledge yet struggle to apply it in real-life communication (Sato, K., & Takahashi, K., 2003). In response to this challenge, learners need to be encouraged to think about activities in which they can talk using the vocabulary and grammar they have learned.

During the course, the following student-staff partnership activities were undertaken:

- Student-staff co-creation of speaking test topics.
- Student-staff co-creation of speaking test rating category.
- Student-staff co-creation of speaking test marking criteria.
- Rater training activity.
- Creation of checklist.
- Students' peer assessment of their speaking test.

3.1 Methods of data collection and analyzation

Student-staff partnership activities began in week 12 and continued, albeit a little each week, until week 14. Group discussion was utilized to create the speaking test topic, rating categories, and marking criteria. Semi-structured group interviews were conducted to explore Japanese students' perceptions of their attitudes toward this educational practice. Microsoft Excel (2000) was used for data analysis. Descriptive statistics were calculated initially, followed by Spearman's correlation analysis to explore the correlation between peer ratings and teacher ratings.

3.2 Participants and the speaking test

The setting: This study was conducted at a university in Hokkaido, Japan. Students are mandated to study a second foreign language, choosing from Chinese, German, or Russian. Each class comprises around 25 students, with 12 classes for Chinese and German, and 2 for Russian. All foreign language classes adhere to the CEFR A1 level.

Participants: Twenty-one Japanese students participated in the study, majoring in information technology, engineering, and science. All students start from scratch. The class convened once a week for 90 minutes over a 15-week duration.



Instructor: The instructor, a female with approximately 20 years of experience in teaching Chinese as a foreign language, conducted the study.

Term-end speaking test: Students were tasked with delivering a presentation introducing family members, friends, or themselves using PowerPoint. Both the teacher and peers assessed the presentations simultaneously in class. Students used a score sheet with four rating categories, circling the appropriate number for each category.

Study content during the semester: The study content and the learning objectives were as follows, roughly equivalent to the A1 level of the CEFR.

- Introduce yourself simply and use basic greetings.
- Tell where someone is from and give a basic, simple description of the city.
- Talk simply about family members, and describe their ages, hobbies, and occupations.
- Talk about your favorite foods and make simple orders in a restaurant.
- Talk about everyday life at your university.
- Describe places and give simple directions.
- Talk about hobbies and interests and make plans.
- Discuss common products and make basic purchases.

4. Findings

4.1 Student-staff co-creation of speaking test topics

To create the speaking test topics, the following activity was conducted in week 12. Students were divided into groups of four or five and tasked with discussing, “What are you able to do from the content you have studied this term?” They were asked to discuss all the topics they could talk about, using the grammar and vocabulary they had learned in this term. As a result, the following seven topics were listed.

(1) self-introduction, (2) introducing a friend, (3) introducing one family member, (4) introducing one's pet, (5) introducing a place where someone lives, and (6) sharing one's university life. Given that introductions of pets and places of residence involved unfamiliar vocabulary, we opted to exclude (4) and (5) from the options.

To streamline the process, students were instructed to select one of the subjects (1), (2), (3), or (6) for their presentation. During the preparation phase, students were asked to determine the number of sentences they planned to use and the anticipated duration of their presentation. Most students expressed their intention to use between 12 and 16 sentences, with a presentation duration ranging from 40 seconds to 1 minute 30 seconds. Consequently, a presentation time of one minute was agreed upon.



4.2 Student-staff co-creation of speaking test rating category

In week 13, students collaborated with the teacher to develop the rating scale. They were tasked with imagining themselves as teachers and determining what type of analytic rating scale they would use to evaluate the speaking ability of their peers. Before the discussion, the teacher provided a brief introduction to the rating categories and scales of high-stakes Chinese speaking tests, such as the HSK (Hanyu Shuiping Kaoshi) speaking test and Chuken speaking test. Drawing on their understanding of these rating scales, the students proposed more than ten rating categories, encompassing facial expression, voice quality, content, fluency, design of the PowerPoint file, organization of the content, pronunciation, accuracy of grammar, and vocabulary, natural expression, and posture.

However, recognizing the impracticality of ten categories for a speaking test, after much discussion, the students decided to streamline it to five at this stage: pronunciation, grammar and vocabulary, fluency, presentation content, and body language.

At this point, many questions were asked about the evaluation of pronunciation, so the students were reminded of the pronunciation knowledge they had learned in their classes and reviewed the points they should pay attention to when pronouncing the words. Examples of good Chinese pronunciation for students at this stage included, for example, the single vowels "e" and "ü", the retroflex sound "zh", "ch", "sh" and "r", the bilabial consonant sound "b" and "p", and the nasal vowels "ang", "eng", "ing" and "ong". The pronunciation of syllables with compound vowels, such as "jiā", and the strong pronunciation of the medial "i" in such syllables were also stressed.

4.3 Co-created speaking test marking criteria

Students were tasked with discussing specific points to be rated within each category. For instance, in defining the rating points for pronunciation, they delved into what constitutes good and bad pronunciation in Chinese, particularly for Japanese students at the beginner level. They also considered the practicality of these points during assessments, ensuring the evaluation could be completed in less than one minute.

The final rating points for each category are presented in Table 2, comprising five categories with a full score of 25. Weightings were determined through collaboration with the students. Although content and body language are not inherently part of speaking skills, many students insisted on their inclusion. Broadly speaking, these skills are deemed integral to presentation skills and were therefore incorporated into the assessment items for this session.

Table 1. Marking Criteria

Pronunciation	Vocabulary & Grammar	Fluency	Body Language	Contents
1 Easy to understand 2 common mistakes for beginner Japanese students	1 vocabulary accuracy 2 Grammar accuracy	1 Comfortable pause (3 seconds) 2 Natural speed	1 Eye contact 2 Good posture, gesture	1 Interesting 2 Good organization



4.4 Rater training

To help students better understand the rating categories, a mini-rater training was conducted. We asked one student to do a speaking example in front of the other students (with that student's permission, we also made a recording) and asked the other students to do the rating together. The actual assessment revealed that too many rating categories would be burdensome for the raters, it is not easy to evaluate five categories in one minute, so they decided to delete the "content", and reduce the number of evaluation categories to four. At last, four categories remain: pronunciation, vocabulary & grammar, fluency, and body language.

Table 2. The Last Version of the Marking Criteria

Pronunciation (5 points)	Vocabulary & Grammar (10 points)	Fluency (5 points)	Body language (5 points)
1 Easy understanding 2 common mistakes for beginner Japanese students	1 vocabulary accuracy 2 Grammar accuracy	1 Comfortable pause (within 3 seconds) 2 Nature speed	1 Eye contact 2 Good posture, gesture

We generated and uploaded three spoken samples, illustrating upper, middle, and lower proficiency levels, onto the Moodle course page for student assessment. As the students found vocabulary and grammar to be the most challenging and were unsure about how to score on a scale from 10 to 2, a rule was established for the category of vocabulary and grammar: 'In cases where there are almost no mistakes, give 10 points; for minor errors, give 8 points; when there are mistakes however they do not hinder comprehension, assign 6 points; if there are some errors hindering comprehension, give 4 points; and in situations where there are many errors hindering understanding of the content, give 2 points.' The students then practiced scoring based on this rule.

4.5 Creation of checklist for preparing for the speaking test

Making a test preparation checklist was week 13's homework. Students were asked to create a checklist to prepare for the speaking test. Each student's list is different, though, as the topics and points to note are different, Table 4 is a checklist that was made by one of the students who did a "self-introduction." This student wrote in particular detail about grammar.



Table 3. Checklist

(1) Pronounce the four tones correctly and pay attention to the pronunciation of “日本人 (Rìběnrén) ”.
(2) Ensure that 'r' does not become 'l'.
(3) Stand up straight (posture). Give the presentation as if speaking, not memorizing.
(4) Keep silence within 3 seconds.
(5) Omit "是" before numbers. E.g. “我 19 岁”.
(6) When saying a full name, use "叫" instead of "姓".
(7) Omit "的" in "我 的妈妈".
(8) Time adverbs should be used before verbs. “我 明年去北京”.
(9) Pay attention to the position of "两次", I ate the Beijing roast twice. 我吃过 两次北京烤鸭.

4.6 Peer evaluation

In the 15th week of the term, a presentation-style monologue speaking test was conducted and assessed by the teacher and peers at the same time in the class. Students were asked to fill out the score sheet which included four rating categories, scored from 1 to 5 or 1 to 10, varying according to the weighting, with 1 being poor and 5 or 10 being excellent, by circling the appropriate number for each category. A sample of the score sheet is shown in Table 5.

Table 4. Score Sheet

Student number	Name	Pronunciation (5 points)	Vocabulary & Grammar (10 points)	Fluency (5 points)	Body language (5 points)
14999411	Sato	1 2 3 4 5	2 4 6 8 10	1 2 3 4 5	1 2 3 4 5
14999412	Shimizu	1 2 3 4 5	2 4 6 8 10	1 2 3 4 5	1 2 3 4 5
14999413	Suzuki	1 2 3 4 5	2 4 6 8 10	1 2 3 4 5	1 2 3 4 5

4.7 Students' ratings and the teacher's ratings

Table 6 presents the descriptive statistics of both peer and teacher assessments. Except for body language, the mean scores for each category in peer ratings were slightly lower than those given by the teacher. This suggests that, in comparison to the teacher's ratings, students tended to be more stringent in their evaluations. Interestingly, for body language, students' ratings tended to be more lenient. Notably, the standard deviation (SD) for all four categories was higher in the student assessments than that of the teacher. This indicates that the teacher tended to rate presentations within a narrower range, while students exhibited a wider range in their evaluations of their peers.



Table 5. Descriptive Statistics of Peer and Teacher Assessments

	Student		Teacher	
	M	SD	M	SD
Pronunciation	3.31	0.84	3.6	0.55
Vocabulary & Grammar	7.14	0.73	7.55	0.51
Fluency	3.74	0.52	4.01	0.46
Body language	3.76	0.48	3.25	0.41

To investigate the extent to which peer assessment correlated with the teacher's assessment, Spearman's correlation analysis was conducted between peer and teacher assessments. The results are presented in Table 7, indicating medium correlation coefficients for all four categories.

Table 6. Correlations Between Peer and Teacher's Assessment

Categories	r
Pronunciation	.41**
Vocabulary+ Grammar	.38**
Fluency	.32*
Body Language	.37**

Note: ** $P < .01$, * $P < .05$

4.8 Japanese students' attitudes toward student-staff partnership in assessment

To investigate Japanese students' attitudes toward student-staff partnership in the speaking test practice, semi-structured group interviews were conducted. The students were asked a single question: "What do you think of this term's student-staff partnership in the speaking test?" They were divided into small groups, with each group consisting of 4 to 5 students. Their responses were categorized into two groups: positive responses and negative responses. The comments were originally given in Japanese and were translated into English by the author. The content of the responses is outlined below.

4.8.1 Positive responses

Student-staff partnership in the speaking test is helpful and useful.

- Rater training, we were shown three speaking examples that presented upper, middle, and lower levels, which was really very helpful. I quickly understood what performance was being expected, and what performance was not expected.
- Working together with classmates and teacher on the speaking test was helpful. Before we graded each other's speaking, I didn't think about what made someone good at speaking. But now, I know clearly how I am evaluated. I've learned a lot about it.



- When we all collaborated on creating the grading materials and underwent rater training, it significantly eased the process of understanding how to prepare for the speaking test.
- Evaluate my classmates' speaking performance and giving them a grade wasn't just about grading – it helped me see what I could do better too. When we all looked at each other's presentations, I could see where I needed to improve.
- By looking at what others did well and where they struggled, I got some good ideas to improve my performance. Now, I know what to do to be better next time.
- Both the rater training and the creation of the checklist taught me a lot. Rater training allowed me to see other people's performance, and creating the checklist allowed me to see myself objectively. I was almost confident during the exam because I knew I was well prepared.
- I started to understand how to prepare for the speaking test and what I needed to do to get a high score.

Student-staff partnership in assessment is motivating.

- When I assess someone else, like if they have a good posture, maintain good eye contact, or create a well-designed PPT file, it motivates me to learn from them and try to do the same thing as well.
- By creating a checklist during practice, I now have a clear idea of the next steps to enhance my Chinese speaking skills, I want to get a better score next time.
- Knowing what to do and how to prepare has made learning Chinese more enjoyable for me. I want to learn more.
- The evaluation criteria were straightforward, which enhanced my motivation to get a good score, so I practiced my presentation multiple times before the test. This gave me the confidence that my peers would give me good scores.

Student-staff partnership in assessment is interesting.

- This was my first time doing this kind of student-staff partnership in assessment, and it's pretty interesting. I felt like I was playing the role of a teacher and had a strong sense of being involved.
- This assessment experience has given me insights into what the teacher thinks about us, and it's been a really interesting experience. It's made me think about teacher' feelings, even outside of the classroom, I started to think the other people's feelings.
- Even though I'm a student, I got to do things that usually only teachers do. It was a nice experience.



4.8.2 Negative responses

Student-staff partnership in assessment is difficult.

- Peer assessment is difficult. Sometimes, I really don't know how to assess the other students, especially when assessing vocabulary and grammar, as I couldn't catch everything said by the presenters.
- It is difficult to assess the other students, so I think the students' ratings are not reliable, I also don't like an arbitrary grading.
- Frequently, I gave higher scores to those presentations featuring compelling content and captivating photographs. Is this appropriate?
- I don't know how to choose from 1 to 5 points, especially 2, 3, and 4 points, and I often end up choosing 3 points.
- Scoring within one minute is difficult, there is not enough time to pay attention to grammar and vocabulary.

Student-staff partnership in assessment is boring.

- Personally, I think creating the checklist and marking criteria was very useful, however, peer assessment is troublesome and boring, I would prefer dedicating my time to learning from textbooks rather than evaluating other students.

5. Discussion and Implications

This paper highlights the active involvement of students in designing the topic, rating categories, marking criteria, and peer assessment process for the term-end speaking test in a Chinese language class in Japan. Previous studies have explored student-staff partnerships in assessment, however, there is a notable lack of examples of staff and students jointly developing speaking tests - this study addresses this gap. Furthermore, while some research has explored cognitive aspects of student-staff partnerships' effectiveness in assessment, this study stands out for its unique emphasis on the validity of student assessment, an area that has received comparatively less attention. The following discussion explores these notable findings.

5.1 Correlation between the student and teacher ratings

In this study, the rating scale comprised four categories, and correlation coefficients were computed for each category between student and teacher ratings. The findings indicated that all four rating categories (pronunciation, vocabulary & grammar, fluency, and body language) exhibited medium correlation coefficients. In contrast to Qu (2018), where three rating categories (vocabulary, grammar, and fluency) showed low correlation coefficients, the introduction of two specific tasks in this study is believed to account for this change. By involving students in co-create rating categories and marking criteria together, the students had a clear understanding of the marking points for each rating category. For example, there were three rating points for body language- eye contact, posture, and gestures, so a



decision was easily made. Pronunciation is a rating category that focuses on linguistic elements of the presentation, it was considered very difficult at the time of creating the rating categories, and there were a great many questions from students about it. There are two criteria for assessing pronunciation: first, the pronunciation should not hinder semantic understanding, and second, it should not be mispronounced in a way easily mistaken by beginner Japanese speakers. Although students were given the responsibility to determine the ease and clarity of understanding, conducting a pre-review session with students on pronunciations prone to confusion could have contributed to achieving better results. There were two primary considerations for fluency: maintaining a pause of no more than three seconds and achieving a natural speaking speed. Both the restriction on pause duration and the assessment of naturalness in silences were regarded as easily understandable.

Qu (2018) pointed out that students struggled to identify vocabulary and grammar errors due to a lack of necessary language knowledge. To address this, a preparation checklist was created before the speaking test, allowing the students to review common vocabulary and grammar issues. This preparation facilitated effective peer assessments. Additionally, at this stage, students had only covered a limited range of vocabulary and constructions, and the vocabulary and grammar were relatively uncomplicated and easy to grasp, potentially contributing to the positive outcomes observed.

In this study, the task was intentionally designed to be straightforward, with students employing words and expressions they had learned and the ease of the task allowed for effective peer evaluations. However, if the task were more challenging or included more unfamiliar vocabulary and complex grammar, the results might differ. Therefore, it's important to note that, at this stage, the study remains a case study, and generalizing the results could pose challenges.

5.2 The students' perception of this student-staff cooperation

The results indicated a positive reception among students toward the practice of student-staff cooperation. Favorable opinions included sentiments such as finding the practice helpful, motivating, useful, and interesting. Participation in student-staff collaboration during the speaking test enabled students to comprehend the rating criteria and understand the expected performance standards set by the teacher. Moreover, engaging in creating and reviewing a checklist for test preparation prompted students to contemplate their test readiness and consider areas for future study to enhance their skills.

Through their independent thinking and efforts, students expressed that the experience was "interesting," found student-staff cooperation to be "helpful and useful," felt confident in their preparation, and expressed a desire to further engage in Chinese language learning. This positive feedback reflects a cyclical pattern: interest leads to utility, confidence increases, and a desire for further engagement emerges. This cycle is expected to enhance self-efficacy, fostering a high sense of self-efficacy that, in turn, promotes independent and autonomous learning.

However, it's important to note that negative comments primarily revolved around the challenges associated with peer assessment, particularly in the context of evaluating peers. Upon closer scrutiny, positive remarks were predominantly linked to the preparation stages of the test, including creating speaking test topics, developing rating criteria, and undergoing rater training. In contrast, negative comments were more focused on the process of peer assessment itself.



Concerning the negative feedback on peer assessment, it's possible that students perceived it as challenging due to a lack of sufficient language knowledge for making judgments using the rating scale. Despite concerted efforts by the students, it's acknowledged that some may still struggle to keep up linguistically. The perception of boredom and hassle might be attributed to the specific procedure used in this study, where students were tasked with assessing around 20 peers in a single class. The sheer volume of assessments could have left them feeling unengaged and overwhelmed. Reducing the scope and focusing on a more manageable number of assessments per student might contribute to a more positive and engaged perception of the peer assessment process.

6. Conclusion

This study centered on a case study that actively sought to meaningfully involve students in their learning through assessment, employing a collaborative approach grounded in democratic classroom practices. The methodology included students in co-creating speaking test topics, determining rating categories and marking criteria, and participating in peer evaluation using these established criteria. This inclusive approach enhanced both the comprehension of the expected performance and the validity of the students' assessment. When students have clarity about the expectations set by the teacher and are familiar with the learning methods that support those expectations, they are better equipped to take control of their learning journey. Those educational activities significantly increased the students' test literacy and facilitated the promotion of independent learning.

Collaborative partnerships between students and staff present an opportunity to challenge traditional pedagogical norms and assumptions concerning the establishment of learning objectives and the evaluation of learning outcomes. This approach extends from existing literature advocating for educators to share authority with students, enabling them to actively engage in the learning process. Such a paradigm shift not only serves as a means of assessing accumulated knowledge but also functions as a potent mechanism for cultivating deeper learning experiences. We believe that adopting a collaborative approach between students and staff in assessment aligns with the principles of learning-focused assessment. It encourages greater involvement in both assessment and learning and offers chances to improve students' ability to evaluate their own progress.

Declarations and Acknowledgment:

The author declares there is no conflict of interest.



References

- Ashenafi, M. M. (2017). Peer-assessment in higher education-twenty-first century practices, challenges and the way forward. *Assessment & Evaluation in Higher Education*, 42(2), 226–251.
- Boud, D., and N. Falchikov. (2006). “Aligning Assessment with Long-term Learning.” *Assessment & Evaluation in Higher Education* 31 (4): 399–413.
- Boud, D., and E. Molloy. (2013). “Rethinking Models of Feedback for Learning: The Challenge of Design.” *Assessment & Evaluation in Higher Education* 38 (6): 698–712.
- Campbell, K. S., Mothersbaugh, D. L., Brammer, C., & Taylor, T. (2001). Peer versus self-assessment of oral business presentation performance. *Business Communication Quarterly*, 64(3), 23–40.
- Handley, K., and L. Williams. (2011). “From Copying to Learning: Using Exemplars to Engage with Assessment Criteria and Feedback.” *Assessment & Evaluation in Higher Education* 36 (1): 95–108.
- Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2001). Web based peer assessment: Feedback for students with various thinking-styles. *Journal of Computer Assisted Learning*, 17(4), 420–432.
- Mulder, R. A., J. M. Pearce, and C. Baik. (2014). “Peer Review in Higher Education: Student Perceptions before and after Participation.” *Active Learning in Higher Education* 15 (2): 157–171.
- Ming Qu & Margit Krause-Ono. (2016). Peer Assessment for Testing Classroom Chinese Speaking in a Japanese University: Correlations and Attitudes. *The Japan Association for Language Teaching Gunma chapter Annual journal*. 24 ~ 35.
- Mercer-Mapstone, L., & Marie, J. (2019). Scaling up student-staff partnership in higher education.
<https://www.advance-he.ac.uk/news-and-views/scaling-up-student-staff-partnership-higher-education>
- Naganuma, S., Sugiyama, Y., Shibukawa, S., Asakawa, Y., & Matsushita, K. (2019). Can Students’ Self-Assessment or Peer-Assessment in Performance Assessment be Close to Valid Assessment? A Case from the Field of Civic Online Reasoning. *Kyoto University Research in Higher Education* 25 13-24.
- Orsmond, P., S. Merry, and K. Reiling. (2002). “The Use of Exemplars and Formative Feedback when using Student Derived Marking Criteria in Peer and Self-assessment.” *Assessment & Evaluation in Higher Education* 27 (4): 309–323.
- Patri, M. (2002). The influence of peer feedback on self and peer-assessment of oral skills. *Language Testing*, 19(2), 109–131.
- Pascale Lorber, Steve Rooney & Mark Van Der Enden (2019). Making assessment accessible: a student–staff partnership perspective, *Higher Education Pedagogies*, 4:1, 488-502, DOI: 10.1080/23752696.2019.1695524
- Rudy, D. W., Fejfar, M. C., Griffith, C. H., & Wilson, J. F. (2001). Self- and peer assessment in a first-year communication and interviewing course. *Evaluation & the Health Professions*, 24(4), 436–445.
- Rust, C., M. Price, and B. O’Donovan. (2003). “Improving Students’ Learning by Developing Their Understanding of Assessment Criteria and Processes.” *Assessment & Evaluation in Higher Education* 28 (2): 147–164.
- Sato, K., & Takahashi, K. (2003). Teacher and student learning in the workplace: The impact of performance tests. *JALT 2002 Proceedings*, 325-336.
- Susan J. Deeley & Catherine Bovill (2017) Staff student partnership in assessment: enhancing assessment literacy through democratic practices, *Assessment & Evaluation in Higher Education*, 42:3, 463-477, DOI: 10.1080/02602938.2015.1126551
- Sadler, D. R. (2010). “Beyond Feedback: Developing Student Capability in Complex Appraisal.” *Assessment & Evaluation in Higher Education* 35 (5): 535–550.



About the Author:

Qu Ming is a professor in the liberal arts group at Muroran Institute of Technology in Japan. Her interests are in pedagogy, language testing, and intercultural communication. She has published over 20 articles, with the most recent ones focusing on teaching methodologies, speaking tests, and intercultural communication.



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)