Evaluation of a Japanese Language Trial Lesson Using FishWatchr

Mari Nishitani, Takuya Iwasaki

Hitotsubashi University, Tokyo, Japan

FishWatchr (FW) was developed as a discussion and observation support tool, incorporating functions of easy annotation and ability to visually access results. In this study, an experimental lesson was conducted by using FW for Japanese language trial lesson evaluation. Results showed that FW evaluation was perceived to be more beneficial than paper evaluation and evaluators took a learner perspective in paper evaluations, but took a teacher's perspective in FW evaluations when evaluating the lesson. It was also clear that comments received from others on learners' own failure were considered beneficial. It seems that combining FW evaluation and traditional paper-based evaluation was extremely effective in evaluation activity in trial lessons. Through the experience of using both evaluation methods, it is possible to foster different evaluative perspectives.

Keywords: FishWatchr (FW), trial lesson, evaluation, failure, Japanese language education

Introduction

The author (Nishitani) has been implementing the trial elementary level Japanese language lessons in a graduate course for the past 10 years. During this time, students playing the student role in trial lessons have been asked to fill out paper evaluation sheets that include the good points and points for improvement after the end of the trial lesson, and these have been given to the person leading the lesson together with the author's evaluation. However, those doing the evaluations were taking student roles and completed the evaluation sheet after the end of the lesson, there is a tendency toward overall impressions after the lesson, and there has not been much time to discuss specific points for improvement.

Many previous studies have pointed out that the importance of lesson evaluation in Japanese language teaching practice and in trial lessons, alongside the essential development of self-reflective capability on the part of the person conducting the lesson (Indoh, 2007; Yanagida, 2015). In school education contexts, research has been progressing on teaching practice by using video-recording and guidance on trial lessons. In trial lessons in the school subject area, Hongo (2009) reported on students doing peer assessment based on videos of recorded trial lessons, finding that,

> Information compared to the traditional method, which tended to promote an atmosphere of vague evaluation, a tendency to evaluate from an analytical viewpoint is evident, and there has been a greater sense of class participation through the activity of closely observing each other's lessons in detail.

Yamazaki, Kato, and Yamazaki (2011) stated that,

Mari Nishitani, Ph.D., professor, Department of Center for Global Education, Hitotsubashi University.
Takuya Iwasaki, M.A., graduate students, Graduate School of Language and Society, Hitotsubashi University.

In order to improve the quality of lessons, it is important to look back at the lesson based on others' evaluations and reflect, but the problem is that it can be difficult to sufficiently understand the intention of the observer through the method of listening to opinions at post-lesson reflection meetings, and so on. The method of video-recording in the class and analyzing the recording as it is shown at later reflection meetings also exists, but it is extremely difficult to examine everything in the limited time available.

In order to resolve these problems, a system was developed whereby video-recorded images and synchronized comments from observers could be browsed together.

The Aim of the Study

Based on knowledge obtained from the previous studies, this study provided an opportunity to conduct evaluation simultaneous with implementation of a lesson by using the information and communication technology (ICT) tool and FishWatchr (FW). FW is a free tool that enables learners to easily annotate something occurring before their eyes, incorporating a function to access results visually (Yamaguchi, Ootsuda, & Kitamura, 2016). It was developed as a tool to support observation of discussion and contribute to discussion education, but it was used in this study for the purpose of evaluating trial Japanese language lessons. Figure 1 shows an actual FW screen. By using the "merge" function provided in FW to collate individual student files into one file, it becomes possible to see the distribution of evaluations, the way in which evaluations of individual evaluators are spread and it points in time at which comments are concentrated.
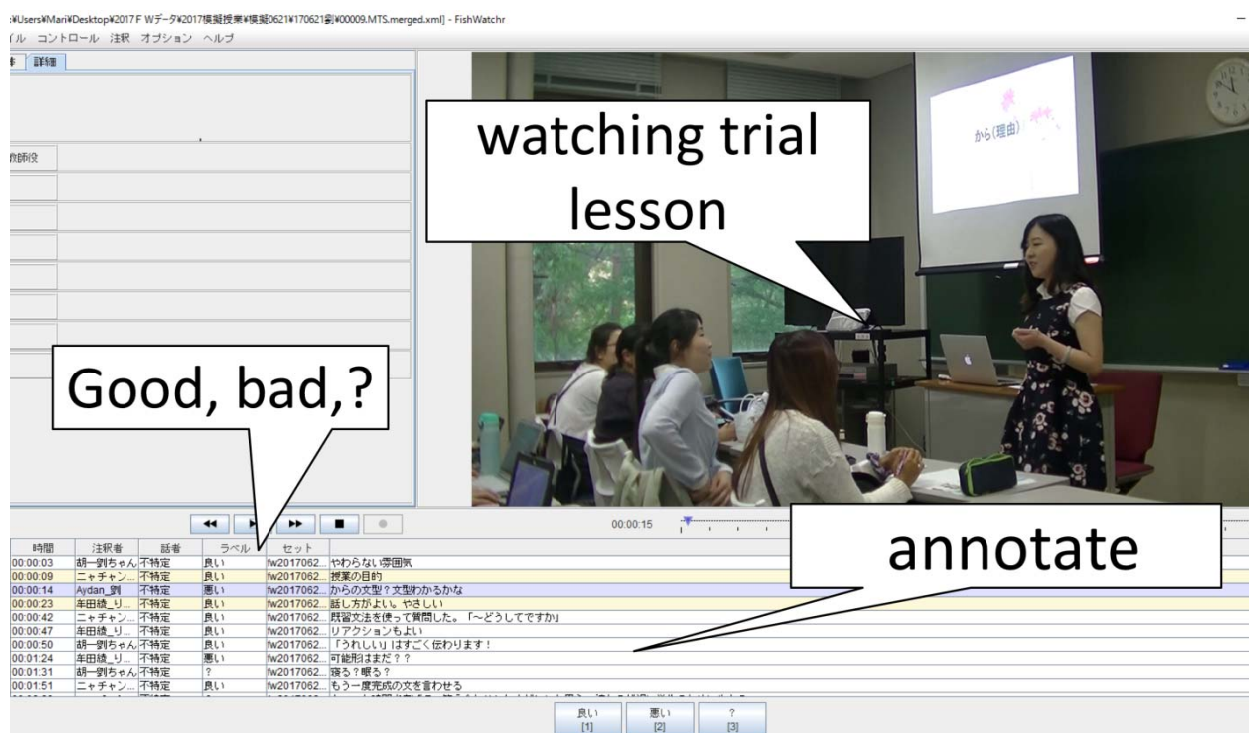

Figure 1. Example of a FW screen.

The three aims of this study were to:

1. Compare traditional post-trial lesson paper evaluation with simultaneous evaluation during the lesson using an FW and to clarify differences in evaluation viewpoints;

2. Clarify differences between self-evaluation and evaluation by others when using FW;

3. Clarify points of evaluation difference among those with experience in implementing Japanese language education, those without experience, and international students.

Through addressing these issues, measures for effective evaluation of trial lessons are then examined. At this point, the perspective of how to interpret one's own and others' failure is introduced.

Overview of Experimental Lesson

The lessons in this study were conducted as part of a graduate level course. It holds an extremely important place in the program as preparation for the smooth implementation of practice teaching that follows this course. Participants in this experimental course were 12 first-year Master's program students, of whom five were Japanese with no experience in Japanese language education, three were Japanese with experience in Japanese language education, and four were international students with no experience in Japanese language education (three Chinese and one Taiwanese). The period of the experimental course was from April to July 2016, comprising a total of 24 trial lessons with each student conducting two lessons of 15 to 20 minutes each.

During the lesson, FW evaluators and those playing the student role sat in different areas, with time allocated after the lesson for those in the student role to complete a paper evaluation sheet (see Figure 2). As two trial lessons were conducted in the 90-minute class, those who took the student role in the first trial lesson switched to be FW evaluators in the next trial lesson. That is, within each experimental class, students experienced both FW evaluation and paper evaluation.

As FW evaluators focus on evaluation, there is the disadvantage that they do not participate in the student role, but they can experience the student role within the same class through the way the course is designed. For FW annotation labels (buttons), "?" was added as an item to the existing "good" and "bad" buttons. The reason for this was that it can be difficult for students to simply evaluate something as "bad" if they do not understand the intention of the person conducting the lesson or cannot understand why something happens within the process of the lesson, and it also seemed to be a useful option for those in the teacher role as they reflected on their own lesson.
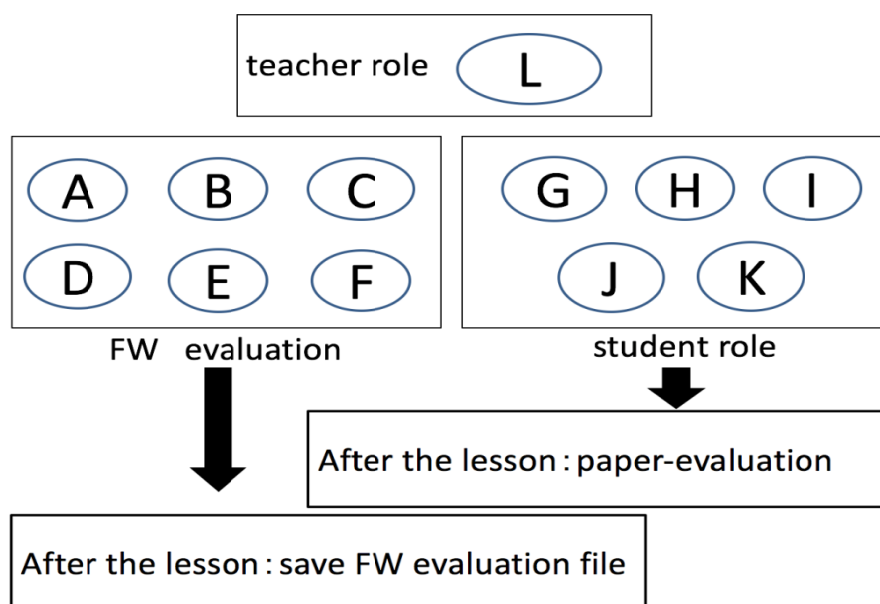

Figure 2. Class configuration.

①Send lesson plan to lecturer by e-mail Interact two or three times

②Lecturer sends comments by e-mail

③Conduct trial lesson, take recorded video home

④Do FW evaluation at home while watching recorded video

⑤Send FW evaluation and reflection sheet to lecturer

⑥Lecturer merges FW evaluation with everyone else's and sends to the lesson teacher

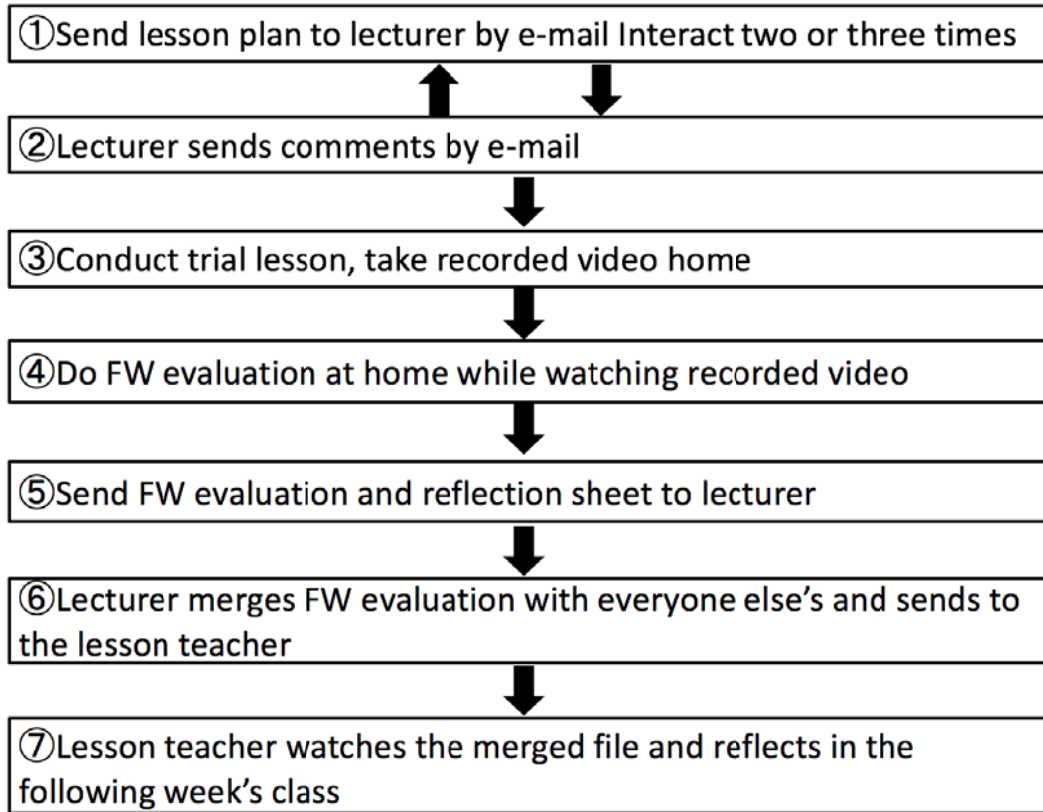⑦Lesson teacher watches the merged file and reflects in the following week's class

Figure 3. Flowchart for conducting trial lessons.

Before the lesson, the person in the role of lesson teacher sent the lesson plan and PowerPoint file to be used by e-mail to the author (Nishitani), who provided feedback to improve the teaching plan. For each trial lesson, the average number of pre-lesson interactions was two or three times, but it sometimes reached four or five times. After the lesson, the video-recorded was given to the student who had conducted the trial lesson, who then sent the FW evaluation done after the trial lesson and the reflection sheet (reflecting on things that went well and failure from planning the class through its implementation) to Nishitani. That is to say, the person conducting the trial lesson performed self-evaluation without seeing the evaluations from others in the student role. After receiving the FW file from the person conducting the class, the author merged this FW file with the FW evaluations of those in the student role, and sent it to the student in the trial lesson teacher role. The person who conducted the trial lesson then looked at the merged file of his/her own FW evaluation and the evaluations of those in the student role, in preparation for a reflective discussion in the following class. The tasks of the person conducting the trial lesson are shown in Figure 3.

Results of Interim Survey

An interim survey was conducted at the point in the course when all students had completed one trial lesson. Students were asked about the extent to which the following aspects were beneficial: FW evaluation, paper evaluation, reflection sheet, reflection session, comments from Nishitani regarding the lesson plan, and feedback sent by Nishitani after the lesson. As well as rating items on a 5-point scale from completely (1 = "Useless" and 5 = "Very useful"), students were asked to write the three most beneficial things in order from first to third. Table 1 shows the mean scores of interim survey results divided into categories of those with no

experience in Japanese language education, those with experience and international students. Results of the survey showed that evaluation of the lecturer's comments was highest, followed by watching one's own lesson on video, then receiving FW evaluation. The point selected as most beneficial was "watching one's own video" (four people), followed by "comments from the lecturer on the lesson plan" (three people), "being evaluated by FW" (three people), and "lecturer feedback after the lesson" (one person).

Students were asked to evaluate the extent to which they paid attention to evaluation on a 5-point scale (1 = "Do not pay attention at all" and 5 = "Pay great attention"). The highest mean was obtained on FW comments (4.58), followed by paper comments (4.41). On evaluation of the FW buttons, "Good," "Bad," and "?," the mean scores of international students (4.50) and those without experience in Japanese language education (4.00) were high, while those with experience in Japanese language education scored low (2.33). The question of whether FW evaluation would prove to be a burden to students was an issue of concern to the author, and so was evaluated on a 4-point scale (1 = "Very difficult" and 4 = "Very easy"). The mean response was 3.01, suggesting that participants in the experimental course found FW evaluation to be simple overall.

Table 1

Points Considered to be Beneficial in the Interim Survey

	Evaluate		Receive evaluation		Video	Write reflection sheet	Reflection session	Evaluation from lecture	
	FW	Paper	FW	Paper				Lesson plan	Lesson
Inexperienced	4.20	3.80	4.60	4.20	4.80	4.60	3.40	4.60	5.00
Experienced	3.00	3.00	3.67	3.67	4.67	4.00	4.00	5.00	4.67
International	4.25	3.50	4.75	4.00	4.75	4.50	4.50	5.00	5.00
Average	3.90	3.48	4.40	3.99	4.75	4.41	3.93	4.84	4.91

Results of Final Survey

A final survey was conducted on the final day of the course. The question—Please explain what was beneficial in improving your own lesson (writing the second lesson plan and teaching the second lesson), was asked together with questions about differences between paper evaluation and FW evaluation (see Table 2). Statistically significant differences were not evident due to the small number of respondents, but it could be seen that students learned a lot from their own failure and comments pointing out the weak points of their lesson. The overall mean score for learning from successes and positive comments was 4.47, whereas the overall mean score for learning from failure and negative comments was 4.82.

All students found the video of their own lesson to be extremely beneficial. It can be assumed that both failure and successes were useful in this case. All students also thought that they could learn from the weak points of their classmates' trial lessons to improve their own subsequent lessons, and wanted to do so.

Table 2

Points Considered to be Beneficial in the Final Survey

	Positive comment			Successes		Average	Negative comments			Failure		Average
	FW	Paper	Lecture	Video	Classmate		FW	Paper	Lecture	Video	Classmate	
Inexperienced	4.00	4.00	3.80	5.00	4.80	4.32	4.40	4.80	4.80	5.00	5.00	4.80
Experienced	4.33	4.33	4.33	5.00	4.67	4.53	5.00	4.67	4.67	5.00	5.00	4.87
International	4.75	4.50	4.25	5.00	4.50	4.60	4.25	5.00	4.75	5.00	5.00	4.80
Average	4.34	4.26	4.09	5.00	4.66	4.47	4.51	4.83	4.75	5.00	5.00	4.82

Next, there were many comments to the effect that it was beneficial to experience both FW evaluations while watching the trial lesson and paper evaluation after participating in a student role. As far as differences in evaluation viewpoints were concerned, two of the 12 respondents replied that both were basically the same, while the other 10 respondents said they were different. It was felt that taking the student role led to evaluating as a learner, whereas FW evaluation led to evaluating from the viewpoint of a teacher. The point was made that,

> In the student role, we are looking back on the lesson and evaluating, so it is not possible to reflect on everything we felt in the evaluation, and it is mainly the points that left an impression that get evaluated.

For evaluation in the student role, many students commented that it is easy to empathize with the feelings of students. Specifically, judgments about whether the teacher was a good teacher from their perspective were made, such as "good teacher," "enjoyable lesson," "the practice exercise was monotonous," "the board was difficult to see," and "the teacher's facial expressions were good." In contrast, when doing FW evaluation, it is worth noting that the lesson content comes to students from the same viewpoint as the teacher. Comments included,

> I could see more objectively as we were focusing on the evaluation, because we were not playing the role of students. It is possible to evaluate in more detail and more objectively, because comments about what was good and what was bad in the moment of each particular situation remain.

Difference Between Self-Evaluation and Evaluation of Others

Table 3 and Table 4 show the results of Pearson's Chi-Squared Test, which was conducted to identify any differences between evaluation scores on self-evaluation and evaluation of others' trial lessons according to the three categories of those who had no experience in Japanese language education, those with experience, and international students. Statistically, significant differences were found in both cases. For evaluation of others' trial lessons, all categories of students used the "Good" button frequently. However, compared to other categories of students, those with experience in Japanese language education used the "Bad" button infrequently and the "Good" button more frequently ($\chi^2 = 40.84$, $p < 0.01$).

This was a major discrepancy from the initial expectation that those with experience in Japanese language education would evaluate more strictly. In contrast to the evaluation of others, the "Bad" button was substantially more frequently used in self-evaluation. In particular, while the tendency for those with experience in Japanese language education to press the "Good" button was stronger than for other categories for evaluation of others, they showed the opposite tendency in self-evaluation with a lower percentage of "Good" button selections than other categories ($\chi^2 = 12.79$, $p < 0.01$). It could be assumed that this shows that those with experience evaluate themselves from a harsher perspective than they evaluate others.

Table 3

Evaluation of Others

	Good	Bad	?	Total
Inexperienced	387 (62%)	105 (17%)	131 (21%)	623
Experienced	273 (71%)	29 (8%)	81 (21%)	383
International	307 (77%)	49 (12%)	43 (11%)	399
Total	967 (69%)	183 (13%)	255 (18%)	1,405

Table 4

Self-Evaluation

	Good	Bad	?	Total
Inexperienced	33 (20%)	123 (72%)	13 (8%)	169
Experienced	8 (11%)	54 (75%)	10 (14%)	72
International	20 (15%)	83 (64%)	27 (21%)	130
Total	61 (16%)	260 (70%)	50 (13%)	371

Summary and Future Issues

Results of implementation of both methods of FW evaluation while watching a trial lesson and paper evaluation after a trial lesson clearly showed that both forms of evaluation have their own distinctive features. From the perspective of being on the receiving end of evaluation, FW evaluation is perceived to be more beneficial than paper evaluation. Students also see comments on their own failure and bad points as more beneficial than comments about their successes and good points. Comparing evaluations of those with experience in Japanese language education, those without experience, and international students, results showed that those with experience in Japanese language education focus on good points when evaluating others, but focus on bad points and failure when evaluating themselves.

From the perspective of providing evaluation, conducting FW evaluation from the viewpoint of the teacher and experiencing both forms of evaluation were seen to be beneficial. It was clear that combining FW evaluation conducted during the lesson as it progressed by using ICT with traditional paper-based evaluation conducted after assuming student roles in the lesson was extremely effective in the evaluation activity in trial lessons. Through the experience of using both evaluation methods, it is possible to foster different evaluative perspectives.

In this study, after discussion with students, it was decided to use the method of asking the person conducting the lesson to do FW evaluation before showing them the merged file of evaluators and the paper evaluations, and only after this show them the merged file of others' evaluations. Another method would be to send the merged file of the other evaluators first, before the person doing the lesson conducted self-evaluation, so that the person in the teacher role could see the evaluation of others before doing self-evaluation. It may be that conducting self-evaluation before seeing others' evaluations provokes a tendency to focus on bad points, whereas seeing positive evaluations from others before doing self-evaluation would restore self-esteem and help those in the teacher role to see their own good points. Given the evident attitude in the survey among students of learning from failure, another area to examine further would be explicit presentation in FW evaluation from the perspective of failure and learning.

References

Hongo, T. (2009). Educational practical training using class interactive evaluation system. Presentation paper of *Japan Society of Educational Information, 25*, 128-131. (In Japanese)

Indoh, M. (2007). Designing teacher development activities for promoting decision-making abilities in student teachers. *CIEE Journal, the University of Kitakyushu, 5*, 23-39. (In Japanese)

Yamaguchi, M, Ootsuda, Y, & Kitamura, M. (2011). Design of the system that supports observation of collaborative learning process using annotation by learner. In *Let's go to the city with evaluation* (Chapter 13). Tokyo, Japan: Kuroshio Publishers. (In Japanese)

Yamazaki, G., Kato, N., & Yamazaki, K. (2011). A development of the reflection support system to synchronize the evaluators memos with a lecture video. Presentation paper of *National Convention for Information Processing Society of Japan, 4*, 499-500. (In Japanese)

Yanagida, N. (2015). Development of reflection program for student teachers in Japanese teaching practice: Tentative program of using recording materials. *The Hitotsubashi Journal for Japanese Language Education, 3*, 13-23. (In Japanese)