

SD&D in the Field

An Online English Speech Contest: What, How, & Who

Vivian Bussinguer-Khavari

Kwansei Gakuin University

vivian.b-k@kwansei.ac.jp

James Carpenter

Asia University

jamesc.carpenter35@gmail.com

David Kluge

Nanzan University

klugenanzan@gmail.com

Dawn Kobayashi

Onomichi City University

kobayashi-dawn@onomichi-u.ac.jp

Abstract

The Japan Online English Speech Contest (JOESC) was created in 2014 by the Speech, Drama, and Debate Special Interest Group (SDD SIG) of the Japan Association for Language Teaching (JALT). This paper describes what the contest is, how it was created, who participated, and how it was administered. It also explains the various problems encountered, specifically regarding the creation of a judging rubric and the inter-rater reliability among all the judges. Some suggestions for improvement, including greater participation in the rubric-creating process and pre-contest training of judges, will help create a better JOESC 2015.

In 2014, the Speech, Drama, and Debate Special Interest Group (SDD SIG) of the Japan Association for Language Teaching (JALT) launched its largest project to date – the Japan Online English Speech Contest (JOESC, pronounced Joe’s C) – with the purpose of creating an inclusive and accessible speech contest in which entrants could participate from anywhere in Japan through the use of video file submissions of speeches using Internet technology. The rules were made simple (see Kobayashi, 2013 for details) and the contest was divided into three sections in order to target high school students, university students, and adults.

Preparations for JOESC 2014

One of the key logistical issues in organizing JOESC was how to collect the video file submissions from the participants. Because of practical and financial considerations, the organizing committee required a system that would be simple, efficient, and free. Although there are commercial programs that allow for heavy-file uploads with limited funding, these were not an option for this contest.

Google was chosen as the best solution. Their online submission form was used because Google automatically transfers the data to a spreadsheet which makes managing the volume of information and recording dates of submissions much simpler (Malette & Barone, 2013).

Participants were instructed to register for a Google account to which they would upload the video file of their speech to their Google drive. They were then requested to both share the file with the JOESC Gmail account and include a link to the file in the online submission form.

The system is not perfect and there are minor privacy concerns involved in asking students to create a Google account while the contest administrators maintain the security of data (Bichsel, 2013, p. 3). Another possible flaw to this system is the risk of deterring participants from submitting a speech if the submission process is considered to be overly time-consuming. Nevertheless, one positive aspect is that this system provided a backup since the contest committee effectively received the speeches twice – once when the contestant gave JOESC the right to view the video file and then when the contestant gave JOESC a link to it. This was invaluable since a couple of participants did indeed fail to give one or the other.

JOESC 2014

Following over a year of planning and creating a rough timeline (see Figure 1), JOESC was first administered in 2014, with the speech submission deadline set for the fall and announcements of winners to be made in the winter. As it was the very first experience in creating such a project, the JOESC committee could not foresee how many speech entries would be received. Committee members made efforts to publicize the contest by distributing fliers at JALT's conference sites (including smaller JALT-related workshops and meetings) and asking SDD SIG members to promote JOESC amongst their contacts and at their institutions. When the submission deadline was reached, JOESC 2014 had received a total of 49 speech submissions: two for the high school category, 31 for the university category, and 16 for the adult category. In order to better explain how these speeches were assessed, the design of the JOESC scoring rubric is described in the next section.

ACTIVITY	TIME
Discuss and decide the organization of JOESC	October 2013
Find sponsors (Cambridge University Press, Macmillan LanguageHouse)	November 2013
Decide/order awards and prizes	December 2013
Create the website and submission system	December 2013
Advertise for participants	April 2014
Select judges	March 2014
Create judging rubric	March 2014
Deadline for submissions	October 2014
Judge first stage	Mid January 2015

Judge second stage	End January 2015
Select winners	End January 2015
Notify winners	End January 2015
Send awards and prizes	February 2015
Evaluate the contest	Spring/Summer 2015

Figure 1 Timeline of JOESC

The JOESC Scoring Rubric

Because the target population for JOESC is students of English as a foreign language residing in Japan, and because it is a contest where judges evaluate the quality of the speeches, the scoring rubric for the contest may be essentially considered a tool for language assessment. Therefore, assessment principles were relevant to the construction of the JOESC scoring rubric. In other words, the JOESC rating scales have the same basic purpose as language testing because the judges were being asked to rate a particular set of language skills (in this case, the performance of an original speech) in order to determine a winner. However, the JOESC was not envisioned or designed to measure students' English proficiency – only to help rank the participants in a particular category. The testing principles were applied to aid in the creation of a fair and useful contest for participants. The important distinction here is that the results of JOESC are not meant to evaluate any participant's language ability outside of the contest. This paper assumes that the basic principles for designing a good set of rating scales are basically the same whether developing an English language assessment or a contest rubric for English language students.

With this in mind, developing an effective language assessment requires a number of considerations. First, what is meant by language assessment should be considered. For the purpose of this paper, language assessment is defined as any instrument designed to measure a particular set of language skills. The central question during the assessment design process, then, is how to identify these particular language skills, and create a metric that is understandable to students, other teachers, administrators, parents or any other stakeholders involved. In the case of

Mask & Gavel Volume 4, Issue 1, October 2015

JOESC, the major stakeholders are the participants and the judges. In order to do this effectively, it is necessary for assessment designers to consider the validity, reliability, and practicality of the proposed assessment (Stoynoff & Chapelle, 2005). Indeed, considerations of validity, reliability, and practicality played an integral part in the JOESC scoring rubric development process, and each is explored in more detail below.

Validity

Validity may be generally understood as “the adequacy and appropriateness of the interpretations and uses of assessment results” (Miller, Linn & Gronlund, 2009, p. 70). In other words, establishing validity explains why an assessment measures what it claims to measure. Performance in a speech contest using a second language constitutes a language task. Language tasks are defined by Ellis (2003) as “activities that call for meaning-focused language use” (p.3). Therefore, developing a valid assessment of speech contest performance does not involve the measurement of specific areas of language knowledge, such as the ability to use the passive voice or accurate English intonation patterns (Bachman & Palmer, 2010). Rather, assessing language ability through a speech contest requires a performance assessment, which measures how well a participant has completed the language task using a scoring rubric.

Generally, there are two types of scoring rubrics. The first is a holistic rubric. Holistic rubrics group criteria under a single score, and are generally used for placement or proficiency testing, where detailed feedback is not a priority (Brown, 2012). The second type of rubric is an analytic rubric. Analytic rubrics divide criteria into separate scores. These rubrics are used for classroom and project-based assessments, where detailed feedback is necessary (Brown, 2012). Because detailed feedback would presumably add value and incentive to participation in the JOESC, an analytic-type rubric format was chosen.

After choosing a format for the rubric, it was necessary to define the language ability JOESC wanted to measure. How language ability is defined is to a large extent related to the sorts of inferences to be drawn from the scores, and by extension the decisions that will be made on the basis of those scores. According to Bachman and Palmer (1996), “when we define an

ability this way, for the purposes of measurement, we are defining what we call...a construct” (pp. 66-67). Defining this construct is a central component of test validity, and was carefully considered during the JOESC rubric development process.

However, there is no agreed-upon definition for what constitutes a ‘good speech’. Therefore, it is difficult to create a construct definition to support the scoring rubric for any speech contest (Venema, 2013). JOESC was no exception. The rubric developer attempted to reach out to the SDD SIG community by asking the several members involved in JOESC to take part in an online survey about the construct definition for the rubric. Emails containing a link to the survey were sent out to those members two times, but ultimately only one member replied. In the end, the rubric developer based the construct definition on his own experience teaching speaking and presentation skills. In this way, three criteria were chosen for the JOESC rubric. These three criteria were (1) matter, or how well the participants organized their speech, (2) meaning, or how clearly and effectively the participants used English to communicate their ideas, and (3) manner, or how well the participants performed their speech (i.e., in terms of body language, facial expressions and gestures).

Reliability

Reliability may be understood as “a function of the consistency of scores from one set of tests and test tasks to another” (Bachman & Palmer, 1996, pp. 21-22). That is, an assessment is considered reliable if the results of that assessment are consistent among different places and different groups of participants. In other words, reliability refers to how consistently an assessment performs between different administrations, among different populations, or how an assessment is evaluated by different raters. Obviously, a major challenge for performance assessment is that different raters can interpret the scoring criteria on a rubric in different ways. In the case of contests like the JOESC, “gesture” is one criteria that can garner a variety of interpretations, and make consistency between raters difficult to achieve. Consistency between raters, or inter-rater reliability, was a significant obstacle for this first administration of the JOESC. This will be discussed in more detail in a later section.

Practicality

Lastly, practicality may be understood as “the adequacy of the available resources for the design, development, use, and evaluation of the test” (Stoynoff & Chapelle, 2005, p. 144). The available resources for the design, and development of the rubric were marginal. One person designed and developed the rubric. However, in terms of use, the rubric was easily used by all judges through the use of digital checklists and shared spreadsheets, and the rubric provided on the JOESC website gave valuable assistance to the contestants, as well as forming a good base for the judges to make comments on the speeches. In terms of evaluation, the discussion, preparation, and writing of this paper provided an excellent opportunity to evaluate the rubric. Therefore, the issues involved with practicality, that is the resources for creation and development were not optimal, but the resources for use and evaluation were adequate.

Submitting Speeches

Guidance to participants was given on the registration form found on the JOESC website Submissions page that asks participants to do the following:

1. Make a two-minute video of the speech in English on the contest theme of *bonds (kizuna)*. The recording must be a headshot of the participant giving the speech without using notes. The video must be a continuous two-minute recording and should not use any editing. The video could be in any of the following formats: WebM files (Vp8 video codec; Vorbis Audio codec), MPEG4,3GPP and MOV files - (h264 and mpeg4 video codecs; AAC audio codec), AVI (MJPEG video codec; PCM audio) MPEGPS (MPEG2 video codec; MP2 audio), WMV, FLV (Adobe - FLV1 video codec, MP3 audio). The video file needs to be titled with the participant’s full name and the name of the contest (JOESC2014).
2. Get a Gmail address.

JOESC Speech Competition Report

3. Upload the video to the contest's Google drive, and share it with the contest email address.
4. Fill in the google form, which then automatically puts all the data in a spreadsheet.

After the submission deadline, the 49 speech submissions were then sent to judges who selected the winners. Details regarding the judging stages are discussed below.

JOESC Judging

The judging procedure was left to each judge to decide, as this was the organization's first experience with an online speech contest. All judges had significant speech judging experience. In addition, most of the judges were members of the SIG and had similar views of what constituted a 'good speech'.

Background to Judging Procedure

Four judges were recruited, three from the SIG (two native speakers of English and one native-level speaker of English) and one outside the SIG (a native speaker of Japanese who is a member of an international speech-making organization). The contest judges agreed upon the judging rubric described above. A good number of participants sent audio-video files of their speeches. It was time to start judging. Two judges (both from the SIG), using the rubric, separately judged all the participants, and input their scores on one common spreadsheet file. The first stage judges found many of the same problems in language (lack of vocabulary, inappropriate word choice, and inaccurate grammar), speaking skills (problems with organization, improper development of key points, uninteresting openings), and lack of sufficient background knowledge (inadequate amount of world knowledge and life experiences) mentioned by Hsieh (2006). However, by eliminating the speeches with the greater number of these problems, they were able to narrow each section of speeches to the top five university students and free category, but for the high school section they retained both entries. They then sent the top speeches in the university and free sections, and the two high school student entries to a second group of judges

Mask & Gavel Volume 4, Issue 1, October 2015

(one judge from the SIG and one from outside the SIG), along with the same spreadsheet file used for the first round of judging. The second group of judges separately watched and listened to the videos, and using the same rubric used by the first group of judges, added their scores to the spreadsheet, each judge inputting scores on one page of the spreadsheet book. That is, during the judging process, a judge would not see the scores of the other judges, but later could look at other pages of the spreadsheet book to see how the other judges scored and to see if his or her scores were within a reasonable range.

Judging Process

As was mentioned above, there were two sets of judges: the first group made up of Judge A and Judge B, and the second group made up of Judge C and Judge D. How three of the judges assessed the speeches is described below. (Unfortunately, Judge D was not available for comment for this paper.)

In the first stage of judging, the judges assessed the speeches in two steps. In the first step, Judge A screened all the entries for any speeches that did not meet the regulations of the contest. Although there were no visible signs of dishonesty, it appeared as if some participants could have either had notes on a table in front of them or next to the camera. As there was no indisputable evidence, these entries were not excluded. However, for subsequent contests it would be better to specify that participants should have no table or other furniture in front of them and that they should stand a set distance from the camera to avoid the possibility of reading from notes. Next, all entries were evaluated using the scoring rubric along with one-point advice comments. Judge B also independently evaluated all of the entries using the scoring rubric. Happily, Judges A and B agreed on the same top 5 entries. Had there been any discrepancy, the next step would have been to discuss and re-evaluate the disputed entries until agreement could be reached. During the second step of judging, Judge A re-evaluated the top 5 entries and provided more detailed comments. Judge A took a constructive approach providing ways to improve speech-making skills in future contests.

In the second stage of judging, Judge C of the second group used a two-step system. First, Judge C watched all the speeches in each category, taking notes and ranking the speaker

holistically; that is, ranking the speakers based on overall impression. In the second step the judge used the agreed-upon rubric to assess each speaker, marking numbers for each characteristic in each category, and then writing a comment, usually based on the Sandwich Method — constructive comments sandwiched between praising comments. The judge then compared the results of the two steps, and found that they matched perfectly.

Results of Judging

In order to determine the winners, the resulting scores for each of the two judges of the second stage were examined. It was found that there was a vast difference between the scores of the SIG judge and the non-SIG judge. Therefore, the results of the two SIG judges, Judge A and B who rated the speakers in the first stage of judging, were used to determine the final ranking of speakers in each category. It was satisfying to see that the three SIG judges agreed completely on the top speakers in each category, but it was a little puzzling that the non-SIG judge, Judge D, often had the exact opposite rankings — speeches that the SIG judges rated at the bottom of the group the non-SIG Judge D ranked as the top of the group.

What could account for the discrepancy in judging between the SIG judges and the non-SIG judge? It could be that the SIG judges (Judges A, B, and C) accidentally had the same image of what a ‘good speech’ is. The non-SIG judge was certainly qualified as she was trained by her organization to judge speeches. This could mean that her rankings were correct and the three SIG judges were wrong. It could mean that she skipped the holistic rating stage and only used the rubric. As Venema (2013) points out, it could be the case of missing the forest because of counting the trees. The discrepancy could also indicate that Judge D’s English was not sufficient to understand the rubric, but even so, she would have been puzzled why following the rubric as she understood it led to results that were contrary to the ratings and rankings of the other judges. A more probable explanation is that various attributes of a ‘good speech’ are especially focused on in her international speech-making organization, and this could have skewed the ratings when a speaker, although deficient in other categories, did extremely well in the style in which Judge D was trained. Because the judges were not interviewed about their judging experience after the judging, the answer or answers to the question of the discrepancy will never be known.

What could be done to prevent this discrepancy from happening again? Some training of judges before the actual assessment process would certainly help to ensure that all judges are using the same understanding of the criteria. This is possible now that the SIG has access to the 2014 speeches that can be used as benchmarks for future training. Interviewing the judges to see what they did and how they came to their decisions could also help in improving the rubric and the judging process. These improvements are necessary if the contest grows in popularity and requires more judges.

Benefits of JOESC

Comparing JOESC to live speech contests, the following benefits are clear:

1. *An online speech contest is more convenient to judge.* Rather than devoting a large block of time on a Saturday or Sunday to judging a live speech contest, the judging of an online speech contest can be done when the judge has free time.
2. *An online speech contest allows judges to be more thorough.* Instead of listening and judging a speech once in a very limited amount of time, in an online speech contest the judge can listen to one speech as many times as desired. If there is a close call, the two speeches vying for first place can be listened to right after each other for easy comparison as opposed to having to rely on memory to compare the speeches as in a live speech contest. Instead of being rushed to finish commenting on a speech, all the time desired can be taken to make and revise comments.
3. *An online speech contest makes it easier to consult the other judges.* In a live speech contest it is possible to ask a question of the judges sitting next to you, but there is a tendency at such contests to try to be quick and this inhibits the judges from asking important questions of the other judges. With an online speech contest, judges can consult each other via e-mail or video conferencing, and in the case of JOESC, they can even look at the other judges' ratings and comments.
4. *An online speech contest takes less time to judge.* This seems to be counter to the comments made in numbers 1 to 3 above, which seem to indicate a leisurely pace to the

judging, which is indeed the case, but the two-tiered judging scheme allows the second-stage judges to only judge the top speeches — the rest of the speeches having been eliminated by the first-stage judges.

5. *An online speech contest lets judges see the contestants' best efforts.* The contestants have the ability to choose when to record the speech and which of the possibly many recordings to submit, allowing the judges to see only their best efforts. It is heartbreaking for a judge to see a contestant, after all the hard work that has been put into writing and practicing a speech, freeze during performance. An online speech, in this respect, is better for both judges and contestants.

When looking at the issue of practicality, some benefits resulted. The way the contest was designed was quite practical in that students have easy access to cameras to take the video and have access to the Internet to send the video files to the contest. Additionally, it was quite easy for the judges to receive the video files and to assess them. Practicality when applied to the creation of speeches and judging of speeches was more than adequate; in fact, it was good.

Problems with JOESC

There were some problems with the contest. When practicality is looked at in relation to the entire contest, some aspects of administration of the contest were problematic. As this was the first year for JOESC, the available resources for the design, development, use (administration), and evaluation of the contest relied entirely on the financial and human resources of the JALT SDD SIG, and largely on the resources of the authors of the present paper. One author designed the concepts and criteria for the scoring rubric. The other authors recruited judges from within their professional networks, and vetted the majority of the contest submissions before sending the finalists to the judges for evaluation. In addition, some of the authors of the present paper provided feedback to all participants about their speech performance.

Therefore, it goes without saying that the resources for this first year of JOESC were austere. The constraints described above limited the implementation of JOESC. JOESC was unable to offer rater training using the scoring rubric for the judges. This contributed directly to some inconsistent scoring among judges. This speaks directly to the limited human resources

available to JOESC this first year, which as a consequence meant that advertising for the contest was neither as robust nor as aggressive as it could have been with more human resources.

Of course, the main drawbacks to online speech contests are due to the lack of human contact — there is no sense of camaraderie among the judges, no forging of professional bonds, and no contact with the contestants. However, given the advantages of online speech contests, it seems they are highly successful and should be developed further.

Conclusion

In the case of JOESC 2014, the writing of this paper outlining the what, how, and who of the process serves as a reflection and evaluation. What was learned from JOESC 2014? One important point was that there should be more participation in the creation of the judging rubric by members of the organizing body. After the rubric is decided upon by the speech contest organizers, there needs to be training materials created from the 2014 speeches and data. Before judging the contest, judges would go through a short online training session using the training materials to orient them to the goals and criteria of the contest in order to get better inter-rater reliability. It would not be necessary, but desirable after the judging process to have the judges fill out a reflection form to gain insights into their judging process. In addition, it would be good to have more people involved in the organizing of the contest. Finally, what was learned from JOESC 2014 was that it is a very worthwhile project that, with improvements, should be repeated in 2015.

References

- Bachman, L., & Palmer, A. (1996). *Language testing in practice*. New York, NY: Oxford University Press.
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. New York, NY: Oxford University Press.
- Bichsel, J. (2013). *The state of e-learning in higher education : An eye toward growth and increased access*. Retrieved from <http://www.educause.edu/ecar>

JOESC Speech Competition Report

- Brown, J. D. (2012). *Rubrics in language assessment with case studies in Asian and Pacific languages*. Honolulu, HI: National Foreign Language Resource Center.
- Hsieh, S. (2006). Problems in preparing for the English Impromptu Speech Contest: The case of Yuanpei Institute of Science and Technology in Taiwan. *RELC*, 37(2), 216-235. Retrieved from <http://www.educause.edu/ecar>
- Kobayashi, D. (2013). Introducing JOESC: Japan online English speech contest. Brought to you by JALT Speech, Drama, and Debate SIG. *Mask and Gavel*, 2(2), 26-33. JALT Speech, Drama, and Debate SIG. Retrieved from <https://sites.google.com/site/speechdramaanddebatepublicsite/home/mask-gavel-archives/m-g-volume-2-issue-2-2013>
- Mallette, M., & Barone, D. (2013). On using Google forms. *Reading Teacher*, 66(8), 625–630. doi:10.1002/TRTR.1169
- Miller, M., Linn, R., & Gronlund, N. (2009). *Measurement and assessment in teaching*. (10th ed.). Upper Saddle River, NJ: Pearson Education, Inc.
- Shannon, M. (2013). Towards a merit-based award system for speech contests. *Mask and Gavel*, 3(1), 26-33. JALT Speech, Drama, and Debate SIG.
- Speech, Drama, and Debate SIG (2013). Submissions to JOESC. Retrieved from <https://sites.google.com/site/japanonlinespeechcontest/submissions-2>
- Stoyhoff, S., & Chappelle, C. (2005). *ESOL tests and testing*. Alexandria, VA: Teachers of English to Speakers of Other Languages, Inc.
- Venema, J. (2013). Judging speech contests: Rating scales and common sense. *Mask and Gavel*, 2(1), 25-30. JALT Speech, Drama, and Debate SIG. Retrieved from <https://sites.google.com/site/speechdramaanddebatepublicsite/home/mask-gavel-archives/m-g-volume-2-issue-1>