広島大学学術情報リポジトリ
Hiroshima University Institutional Repository

Title	How to implement interactive speaking tests in junior and senior high school English classes: Insights from first-hand experiences
Author(s)	Sengiku, Motoji; Jennings, Nick
Citation	中等教育研究紀要 ／広島大学附属福山中・高等学校 , 62 : 174 - 185
Issue Date	2022-03-31
DOI	
Self DOI	10.15027/53484
URL	https://ir.lib.hiroshima-u.ac.jp/00053484
Right	
Relation	

How to implement interactive speaking tests in junior and senior high school English classes: Insights from first-hand experiences

Motoji Sengiku
Nick Jennings

Although many English classes in Japan have taken on the communicative qualities, speaking tests have yet to become an integral part of the instructional routine at a junior and senior high school level. This paper provides a quick guide of how to create speaking tests to enable teachers to evaluate interactive performance. The key to the production of well-balanced tests for classroom assessment will be discussed with special emphasis on the treatment to lessen the vocabulary constraints. The discussion as to how to ensure validity, reliability, and practicality of the test follows with original speaking test sets.

1．Introduction

The new *Course of Study for High School* will be put into operation in April, 2022. Under the new policy, the authentic evaluation process including performance tests will strongly be encouraged. Therefore, there will surely be a growing concern among high school English teachers regarding how to create speaking test concepts for their students, particularly those for evaluating their speaking performance through interactive activities. The new government syllabus has drawn English teachers' attention from receptive skill development to that of productive skills, by turning the previous speaking into the newly introduced categories, spoken production and spoken interaction, in the description of goals or instructional contents of each course. A government research conducted in 2020, however, shows that more than 50% of the English teachers who reacted had not provided any form of speaking performance tests in their courses. Considering the washback effects of the tests on the outcome of learning, this shouldn't be left unsolved.

Many English teachers are going to face problems in their forthcoming classes under the new curriculum when they assess their student performance. How can they develop speaking performance tests suitable to judge the achievement of the unit? How can they reduce subjectivity in the rater or among the raters and create a consistent rating system? How can they carry out the speaking performance tests for so many students in a limited period? These are the questions that English teachers may have before they actually implement their speaking tests.

Ideally, training programs should be provided so that English teachers will acquire skills necessary to implement speaking tests. In reality, English teachers will have to solve the problems only with the use of an official manual, made by National Institute for Educational Policy Research (2021), which has several sample lesson plans regarding performance tests. Is developing speaking tests supposed to be easy? An examiner's subjective judgement is also a concern among researchers, though it plays an essential role in classroom assessment. It is pointed out that each English teacher, though they are not aware of them, seems to have developed internal evaluation criteria (Nekoda, 2019), and they may assess the student performance intuitively even when they are instructed to use scoring criteria (Sato, 2011). Rating inconsistency can be averaged, in theory, with the participation of multiple scorers. Is it realistic in the classroom assessment? Teaching resources are limited in most cases, and teachers should find other ways to avoid rating inconsistency.

This paper is a concise guide for teachers of junior and senior high schools. It covers issues regarding what it takes to design a speaking test that will assess interactional performance of junior and senior high school students, used as an achievement test in class. Thus, problems that English teachers are likely to face while testing are stated, along with some suggestions for improvements, based on experience. They are based on test performance data and reflections, which were obtained through the tests created and used by the writers of this paper in 2021.

The writers have always made a test with a simple scoring matrix, usually having only a brief talk about testing procedures. They have then started the assessment, with a discussion about scoring of one or two performances at the beginning, under the leadership of the first author of this article. Unfortunately, some parts of the tests were unsuccessful and were unable to be used for grading. However, by adding the descriptions of their revising process, with the attempts to

overcome what was thought as failure at that time, the readers of this paper can gain insights on what it means to offer authentic English interview tests, which will help them to create their own speaking tests to better fit their instructional goals.

２．How to enhance the quality of a speaking test for interactional skill assessment: An overview

To create speaking tests, the teachers should follow the following steps (Underhill, 1987). First, they should define the aim of their test (a proficiency, placement, diagnosis, or achievement test), being aware of the resources they can use for the test (who can be included and how much time can be used in the testing process), which refers to specific equipment and facilities they should obtain. These define, or more accurately limit, the type of the test, or how to elicit student performance (whether or not the teacher is involved as an interlocutor in the conversation to elicit the student performance). Popular elicitation techniques for interactive tests are a group discussion, a role-play, or an interview.

A test of any kind can be evaluated in terms of its validity, reliability, and practicality. Validity questions if the test is going to do what it is supposed to do. A good achievement test should be a good representation of the material that needs to be tested. Two major reliability types are the inter-rater and the intra-rater reliabilities and they question the consistency of scoring. Practicality questions whether a test is practical in terms of financial limitations, time constraints, ease of imple-mentation, such as scoring procedures, and the interpretation of the test results. If these three criteria are met, a test will then be administrable.

Then the teachers should specify which part of the student performance they should see during the test. Skills the students should demonstrate in the interactive speaking tests are twofold: interactional management and discourse organization (Lyn et al, 2020). The former includes (1) topic management (to begin, maintain, or end the conversation), (2) turn management (to gain, maintain, or offer turns), (3) clarifying meanings in their interlocutors' speech, and (4) breakdown repairs (to get into a new topic or not). The latter include (1) students' use of discourse markers to lessen the burdens of their interlocutors in comprehending their speech, (2) establishment of coherence and cohesion over several turns, and (3) selection of the frame (the type of information structure) appropriate to the purpose of the interaction they are engaging in.

Then the teachers can think of fair administration: how they can surely gain stable performance from each student over the class period they use for testing. They should also be aware of how they can maintain the test conditions equal to each of the examinees throughout their speaking test process.

３．Steps to be followed in developing an interview test for classroom assessments

The first step for English teachers to follow is to know the purpose of the test. This paper is written for the teachers working at junior or senior high schools, so its target test type is the achievement test to be used at the end of a teaching unit (or a lesson). In the stages of instruction, considering the speaking skill development, the teachers provide speaking activities so that the students will be equipped with specific target skills and notions. They are supposed to have become functional in the settings with the skills and necessary expressions which they are supposed to use in the performance test provided in the end of the teaching unit.

The common features of the test, such as the settings where the interaction takes place, the purpose of the interaction, or most, or at least, some of the expressions to be used in the test, should reflect on the classroom activities in the unit. In an achievement test, the teachers assess their student performance, focusing on whether the purpose of the language use is fulfilled, whether the necessary message exchange is made, and whether or the target expressions are actually used. These will explain the validity of the test.

The second step is to know the examiner factors. How many teachers will take part in the test in the same class period? This influences the length of the test period, and what should be done before actual scoring starts. Many school settings has one ALT (Assistant Language Teacher), so if more than one examiner is necessary, other teachers should be added to the examiners. A JTE (Japanese Teacher of English) monitors the students waiting for their turns in the classroom if they hold the speaking test as a part of their normal class schedule. If the ALT takes the job of interlocutor in the interview test for all the students of the class, one minute can be assigned to each interview session, while two minutes can be assigned to each student if two teachers are available. If more than one teacher is available, it is better to use more than one room to ensure the equal testing conditions. The teachers may have to give up providing two-minute interview tests due to the limit of either

the teachers available or the rooms available at that time.

When the teachers develop the testing item, they should first set the conversational context of the speaking test. It includes the purposes and setting of the conversation, language function and items to be used to achieve the purposes, and some conditions the examinees should meet in the manners of speech. The examinees should be provided chances to get familiar with them while they take part in the speaking activities in class. They should then make the evaluation form which includes a checklist of the skills to be used, of the verbal and non-verbal behaviors to be expressed during the interview test.

The teachers can now start thinking about the marking system. The purpose of holding a test is for teachers to reflect their instruction, and for students to reflect their own learning. This means that the teachers should provide their students with a simple tool to interpret their learning attainment with. Accordingly, the teachers can help their students become more responsible for their own learning. To achieve this, the use of a rubric is very useful. A rubric is a scoring guide to assess the quality of student performance, which has evaluation criteria and the definitions of quality of performance required for particular grades. With rubrics, the teachers can quickly evaluate their student performance with great consistency in marking even while engaging in the test as examinees. Rubrics also enable learners to become able to exercise self-control over their own learning with the feedback provided through rubrics.

However, the actual process that the teachers should follow is often the reverse. They should think of practicality issues first. It will no doubt be an impossible situation to involve two teachers in the actual testing period, mainly due to individual teaching schedules, or perhaps due to the unavailability of separate testing venues. Accordingly, the other factors like validity and reliability are likely to deteriorate. If this is possible, students may then be allocated two minutes for each individual testing and evaluation. This extra time per student may well relate to providing the student more opportunity to converse and possibly sufficient time for on-site oral feedback directly from their examiners instead of written comments from them after one-minute examinations.

In the following sections come the writers attempt to make their speaking tests appropriate to evaluate their students' interactive skills. Typical problems are sorted out, with descriptions of test tasks and of the examiner-examinee conversation excerpts.

4. Discussion: Complications of speaking tests and their implementation and suggestions for improvements

4-1 Validity-related problems

4-1-1 How can the teachers focus on what is significant on during the test?

The students should be able to show interactional skills in a performance test in the unit-end achievement test. However, in order to get their message across, the students need the contents of their speech, through which they show the target behavior including target skills. Since the performance test shouldn't be the chance for the students to show their achieved vocabulary, we should think of effective ways to free the examinees from their vocabulary constraints. According to Adams (1980), the vocabulary accounts for much of the oral performance of novice learners. Vocabulary is, indeed, the fuel to speak. This should be taken into account in the test design.

One way is to make the setting of the language use similar to the scenes from the students' everyday lives. This will work particularly well in ESL settings, but EFL textbooks also have model conversations which take place in some familiar settings like their schools or their homes. The examinees are already familiar with the settings in their life. Thus, they form schema in their memory, which will help them in L2 communication. However, we shouldn't forget that some examinees lack the language-related experiences to which they can refer as schema when they create verbal messages in L2.

For example, in an advanced test named HOPE, High School Oral Proficiency Exam (Imai & Yoshida, 2007), the examinee, as a customer of a shop should ask a shop clerk, an examiner, to get a discount or to make complaints about the purchased items to get a refund or a new item. This is a test to assess the proficiency of using the skills and languages to negotiate. However, many of the high school students still live with their parents and aren't responsible for such issues. Hence the real-life negotiation skills are of little importance. Creating a realistic context to elicit the target speech and behaviors in the test can thus be difficult.

To avoid having the students in imaginative contexts, many people may think it a good idea to ask the examinees to talk about their own experience. However, this creates problems. First, some examinees are familiar with the topic of the test, while others aren't. Also, it is simply impossible to judge whether the examinees are producing their speech based on their own experience.

As a solution, the writers often use test tasks of the role-play

type, in which they are asked to play a role of a major character in a textbook story with which they are already familiar through the classes prior to the test. Usually, they have spent at least four or five classes using a textbook passage, with several activities to enhance the students' comprehension of the passage itself. Some sample tests of this kind come in the next section.

The writers have used textbook stories to write speaking performance test tasks where their students should take roles of the main characters of the stories. This treatment brings three major advantages. First of all, the explanation of the setting of the test can be very simple. You are "(a character)" tells the examinees a lot. Another is that this will surely promote the integration of reading and speaking skills in the activities used in the teaching unit, and this will motivate the examinees to read the textbook stories carefully. The other is that the role-play type speaking tests free the examinees from the vocabulary constraints. If the examinees take a role of a character in the textbook stories, whose contents are already familiar to them, the examinees' active vocabulary levels will naturally be standardized, due to the various tasks in their classes prior to the test.

Speaking is, by nature, context-bound since the contexts define the vocabulary the examinees are supposed to have obtained. Therefore, if the teachers can successfully create speaking tasks where the vocabulary acquisition will be achieved through the instruction prior to the test, the examiners don't have to worry too much about the vocabulary factors in the process of testing. Thus both the scoring procedures and the interpretation of the test score will be simple and easy, and this will enable the examiners to focus on the target skill use by the examinees during the test.

4-1-2 How can the test be made as realistic as possible?

In a role-play type speaking test, the examinee is asked to take on a particular role and to imagine themselves in that role in a particular situation (Underhill, 1987). They are required to use language items and skills necessary to achieve a communicative goal described in the role play scenario. The easiest version is to repeat the memorized dialogue and the most difficult one is to act wholly on improvisation according to the given situation, with, somewhere in the middle of the two, a realistic version which requires the students to select the relevant items and skills in the same or similar situation to the one found in the textbook.

If the students are not familiar with the role to play or the situation where the interaction takes place, they cannot fully demonstrate their knowledge or skills in the test. So, the writers often use a textbook story in order to make up a role-play scenario, and the scenario is usually written in Japanese. Some students may show personal reluctance to participate because "role-playing by definition implies pretending (Underhill, 1987, p.52). The teachers must be aware that some students can do this more easily than others. The teachers, therefore, should make the role-play setting as fair and real as possible by carefully choosing the story whose characters are about the same age and have "problems" that the students feel sympathy with or at least can relate to.

Also, the students can find a wise way to get through the speaking test protocol without showing that they aren't really understanding what is going on in the dialogue. The following is the excerpt of the conversation between the examiner and the examinee derived from a speaking test (See Appendix 2). In the conversation below, the student are managing to keep up with the flow of the dialogue, though he acknowledged in the self-reflection worksheet after the test that he hadn't really understand the answer given to his question by the examiner (Statement 04).

01 Student C: Is it a non-fiction?
02 ALT: Yes, it is.
03 Student C: Oh, non-fiction. What can we learn from the story?
04 ALT: I guess many people are working hard to make our world a better place.
05 Student C: Oh, I see. Can you tell me about the main character?
06 ALT: Yes, it's about a man who built a landmine museum in Cambodia.
07 Student C: A man did what?
08 ALT: He built a landmine museum.
09 Student C: Oh, I see. Is it "Landmines and AkiRa?"

This can be considered a very successful performance in light of the goals of the task. The student, however, made a remark in the a written report that he wrote in the self-reflective activity immediately after the test that he was successful based on the fact that he remembered all the questions and repeated them in order, though he actually didn't understand some of the statements given by the examiner. His desperate memorization strategy worked perfectly in this case. Following that remark, he also acknowledged that he should have used the question for clarification request then (at Statement 05) instead of saying

"Oh, I see" then. How can the language teachers judge whether the examinees are using some language items because they are understanding their true value? It is not always possible for a teacher to create the natural setting where the examinee can show their true understanding of their language items.

The students who took part in this test had actually read all the books as a part of their English classes during the previous year, and as a result, the purpose of their participation in the test was changed into the demonstration of their knowledge of the relevant expressions. That setting, however, was made because of the time constraints. If two-minutes had been assigned to each student, the students could have managed the conversation as to an unfamiliar book, though they had to be more patient in the process of verbalization of their thoughts. In this case, a trade-off between practicality and validity undermined the communicative aspects of the task.

4-1-3 How can the teachers help students feel less stress during the test?

Most students leave remarks about the stress they experienced during the interview tests in the self-reflection worksheet. Their teachers, on the other hand, try hard to relax them as much as possible. However, some attempts have result in vain, or even left negative influence on the student performance. The following is a typical example. The testing materials are in Appendix 1.

01 Student B: Hello. (Handing the evaluation form)
02 ALT: How are you?
03 Student B: Pretty good.
04 ALT: Oh, you wear new glasses.
05 Student B: Yes.
06 ALT: I thought so. OK. (Started the timer.) Which grade are you in? (Noticed that Student B wanted to listen to the question again by her gesture.) Which grade are you in?
07 Student B: I'm in Hiroshima.
08 ALT: No, no, no. You are an elementary school student. What grade.
09 Student B: I'm sixth grade.
10 ALT: Sixth grade. Oh, what do you have in your hands?
11 Student B: I have a textbook.
12 ALT: (Laughter) No, you're Sadako.
13 Student B: Oh… I have a paper crane.
14 ALT: But you have many paper cranes. Why? Why do you have so many cranes?
15 Student B: Cranes are a symbol of long life.

The examiner, one of the writers, thought that the examinee was nervous, and tried to ease her mind by commenting about her new glasses. The examiner saw her smiling and thought his attempt went well, but immediately after that, it turned out that his remarks had confused her. Though she showed up as Sadako, a character the examinee was supposed to play, in front of him, but he talked about the glasses of the examinee and she replied to his comments, which resulted in the loss of concentration.

To avoid the confusion of this kind, the examiners should say, before they actually start the test, that they are going to have a test, and that the examinee is supposed to answer by taking the role of the character of the story during the test. This may sound a waste of time, but the above example shows it is necessary. Also, it may be a good idea to tell the examinees to bring in a symbolic item that the character has or uses in the story. If the writers have another chance to provide this test, they will have each examinee to come into the test venue with a paper crane. This will not always work because the teachers cannot always think of suitable items for all the test settings.

4-1-4 Other factors: A limit to the instructional design or difference in cognitive demands

It is also difficult to design speaking tests to evaluate some aspects of interaction because they cannot always be seen in the natural conversation. For example, in theory, the conversation for negotiation typically has three stages. And according to Sengiku (2019), high school students, who engaged in experimental classes including role-play tasks for negotiation for two months, used more expressions instructed to use in negotiation in the post-test interview. However, while various expressions recommended to use in the first stage were observed in all the cases, ones for the latter stages were observed in less cases, with less varieties. Even after the two-month engagement, many students were still unable to advance to the latter stages. The time constraints in the experimental classes may have prevented many of them from doing so, and many of the target expressions that were taught to use in the latter stages of negotiations weren't actually used. Further research should be done to investigate whether this was caused by the instructional design or difference in cognitive demand that the use of each expression impose on the students.

For example, it is highly probable that among seven aspects of interactive competence (See Section 2), few breakdown repair strategies will be naturally taken in the activities, since the students may not get into any breakdown until they achieved the goal of the task, or they will be instructed to finish

their practice when "the clock expires" during the role play, without using the proper strategies necessary to get out of communication breakdown, in both of which case they will never actually use the target expressions.

If the students take speaking tests in student-to-student pairs, their inability to take proper strategies to maintain interaction may cause a deadlock in the conversation flow during the test. A simple solution to this problem is to design a test of a teacher-student pair although, by taking this measure, the number of tests that one teacher can handle will be reduced by half, so this may cause another problem. But if the target skill of the test is the use of breakdown repair strategies by the students with low-level proficiency, the advantage will override all other concerns. The test specification should tell the examiner to include wrong information deliberately in at least one of the statements they give. If the examinees notice it, they will react to it properly. This will ensure the validity of the test and meet the test objectives.

4-2 Reliability-related problems

Reliability concerns the inconsistency of the grading by the examiners. Among the present writers, the JTE designs the test and the ALT takes the role of the examiner. The followings are the measures they take to ensure the reliability of a test.

The writers usually have discussion on test objectives briefly, and have rehearsal of the test, the JTE acting as an examinee and the ALT using its scoring rubric. Since it is usually very difficult for the ALT to have discussion about the test specification, their brief talk before the test is mainly or exclusively about the testing procedures. However, as is stated in Introduction, teachers often assess student performance with their own scoring criteria developed in their mind through their own learning or teaching experiences, which will bring the inter-rater inconsistency to the scoring outcomes.

For the convenience of the rater, the JTE makes a scoring matrix which the rater can use as the test dialogue proceeds, and the scoring rubric that the students can analyze their performance with the scoring matrix. The writers see the use of the scoring matrix helps the examiner can proceed the dialogue step by step, and evaluate the performance analytically. The detailed information given on the matrix helps the students look back on their performance more positively.

The design of the scoring tools needs careful consideration since it will influence inter-rater reliability. A word used in one of the scoring criteria once caused serious discrepancy in the scoring results of the writers. The ALT gave full-mark to the reactions of almost all the students, while the JTE thought he could give it to those of only few in Speaking Test 2 (see Appendix 2). The JTE realized that the label read "reaction" and the ALT evaluated the student performance on that category in terms of both verbal and non-verbal reactions. The ALT taught the students only once every week, while the JTE taught other classes too. The ALT was not as aware of the instructional objectives as the JTE, so the ALT saw this category as that of a proficiency test. Also, the ALT tends to naturally praise the student if he or she try to communicate with him.

Due to the JTE's inappropriate labeling, the writers failed to share the common ground for the fair and reasonable testing from the beginning of the test. Fortunately, this did not create a disaster. All the test conversations were videotaped, and the JTE could evaluate the performance with the intended scoring criteria, though it took him an extra hour.

5. Conclusion

The key to the production of well-balanced test for classroom assessment is the development of role-play tasks based on the passages of the textbooks, thereby helping the students to meet the vocabulary-related demands of the test. In this way, the teachers can focus on the interactive skills to be evaluated to ensure validity of the test. The quality of the dialogue embedded in the interactional test really matters. Some students are clever enough to find a way to lessen the cognitive demands of the test, and to achieve the goal of the task without going through the critical stage of the test where the students are supposed to use the target skills. Teachers should tell the students the outcome of the tests with rubrics as soon as the tests are over. Since the rubrics indicate the extent to which the students have achieved the target skill acquisition, they can look back on their language learning by then or at least the performance during the test. With rubrics, the teachers can visually share the expected outcome of the student performance, which ensure the consistency in rating across the examiners. However, subjective judgement can deviate the scoring results from the original objectivities set in the test specification. It can happen at any stage of the evaluation process. The collaboration among teachers doesn't simply mean working together to sharing the examinees among the raters. The collaboration of the teachers based on the forthright talk as well as the professional knowledge is the key to success.

References

Adams, M. (1980). Five co-occurring factors in speaking proficiency. In J. R. Firth (Ed.) *Measuring spoken language proficiency*. Georgetown University Press.

Brown, G., & Yule, G. (1983). *Teaching the spoken language: An approach based on the analysis of conversational English*. Cambridge University Press.

Goh, C. (2012). *Teaching speaking: A holistic approach*. Cambridge University Press.

Council of Europe. (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge University Press.

Koizumi, R and Watanabe, A. (2021). Rater reliability in classroom speaking assessment in a Japanese senior high school. *Annual Review of English Language Education in Japan, 32*, pp.129-144.

Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.

May, L., Nakatsuhara, F., Lam, D. and Glaczi, E. (2020). Developing tools for learning oriented assessment of interactional competence: Bridging theory and practice. *Language Testing, 37*(2), pp.165-189.

Nation, I. S. P. and Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. Routledge.

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge University Press.

Sato, T. (2011). Japanese and native English-speaking teacher's perspectives on learner's oral English performance. *Annual Review of English Language Education in Japan, 22*, pp.17-32.

Underhill, N. (1987). *Testing spoken language: A handbook of oral testing techniques*. Cambridge University Press.

Imai, H. & Yoshida, T. (2007) 『HOPE：中高生のための英語スピーキングテスト (HOPE: English speaking test for junior and high school students)』教育出版

Ministry of Education, Culture, Sports, Science and Technology (2020)『令和元年度「英語教育実施状況調査」概要』 https://www.mext.go.jp/content/20200715-mxt_kyoiku01-000008761_2.pdf

National Institute for Educational Policy Research (2021)『「指導と評価の一体化」のための学習評価に関する参考資料』

Nekoda, H. (2019)「スピーキング能力の評価傾向を自己診断するためのオンラインシステムの開発(Online system to have users reflect upon their rating tendency on speaking skills)」『中国地区英語教育学会研究紀要 (*CASELE Journal*)』, 第 49 号, 21-31.

Sengiku, M. (2019)「交渉を行う言語活動を用いたスピーキング指導と高校生の英語発話の質の向上 (Speaking instruction including negotiation tasks and its effect on high school students' speech performance)」『日本教科教育学会誌 (*The Bulletin of Japanese Curriculum and Research and Development*)』第 42 巻, 第 1 号, 13-26.

Appendix 1: Speaking Test #1

1 Test Specification

A. Test type / Examinees: Achievement Test / The 9th grade students

B. Objectives:

(1) Students should be able to maintain the short conversation with a stranger by answering the given questions in a sentence.

(2) Students should be able to use typical expressions to buy time before they actually speak.

C. Setting/Roles: *The appearance of an English man and her English ability are "made up" for this test.

Examiner = An English man visiting one of his friends in the hospital, happening to see Sadako, who is folding a paper crane then.

Examinee = Sadako, an A-bomb victim, who has just started her stay in the hospital, diagnosed with leukemia ("a kind of cancer" is the term in the textbook). She is still fine enough to fold paper cranes and have a conversation, able to speak in English.

D. Time: 50 seconds each (Please DO NOT shorten or extend this).

E. Testing Venue: Outdoors (near the entrance to the classroom building)

F: Procedures:

・Evaluate one examinee at a time. Enjoy interaction with them.

・You have three different cards which has questions to elicit response from the examinees, and use the cards in order. Two other examinees are waiting behind the examinee in test, so this is important.

・Grading the performance on the three-point scale using the scoring matrix (See Evaluation Form). *Preferably this process is done during the test.

G: Notes:
・You are a stranger to the girl. But the girl is young, so be friendly. She will be, too. This is a conversation on a daily topic.
・If the examinee cannot answer one of your questions, please skip it.
・You can repeat the question if requested.

2 Textbook Passage: from Lesson 3 The Story of Sadako, *New Crown English Series 3*, pp.40-41

It began with a flash. On August 6, 1945, an atomic bomb was dropped over Hiroshima. Sadako was two years old. At least 130,000 people died by the end of the year, but she survived.

When Sadako was in elementary school, she especially liked her P.E. class and was good at sports. She wanted to be a P.E. teacher when she grew up. Sadako was a fast runner. In the sixth grade, she was selected as a member of the relay team for the school's sports day.

About a month after the sports day, Sadako suddenly became sick. At first she thought that she just had a cold. However, her sickness got worse, so she went to the hospital with her family. A doctor told her parents, "She has a kind of cancer caused by the bomb. I doubt she'll survive for more than one year."

In the hospital, Sadako received some paper cranes. In Asia, cranes are a symbol of long life. Sadako began to fold paper cranes and wished for good health. She wanted to go back to school. She never lost hope. Sadako made over 1,000 cranes. However, she never left the hospital. Her life ended when she was only twelve.

3 Scenarios: *SET 1 and SET 3 cannot be included due to the limited space.
SET 2
(Q1) Which grade are you in?   - I'm in the sixth grade.
(Q2) Oh, you are an elementary school student.   – Yes, I am.
(Q3) Oh, what do you have in your hand?   - I have origami / a paper crane.
(Q4) But you have many cranes.   Why do you have so many cranes?   - If I make one thousand cranes, I will be fine then.
(Q5) When you get better, what do you want to do first?   - I want to see my classmates. / I want to run.   etc
(Q6) Nice talking to you.   – Nice talking to you too.
*To the examiner: The examiner and the examinee are supposed to have a conversation, so do not say the numbers aloud.

4 Handout for the examinees

Speaking Test #1 「Lesson 3 の Sadako　と　Nick が話したら」要項
A. 日時と場所：　　☆当日連絡します。　☆「評価用紙」は当日配布します。
B. 問われる力：　登場人物の立場で英語で質問に回答できる。
※わからない時，答えられないときには，英語でそう伝えましょう。
C. ロールプレイの詳細：
(1)状況　まだ入院してすぐのころ。ベッドで折り鶴を作り，普通に会話はできる。
(2)場面　隣のベッドの患者のお見舞いに来た男の人（Nick）に話しかけられる。
D. 準備の内容：　Lesson 3 を暗唱するぐらい読んでおく。書いていないことを聞かれたら（例えば，彼女の「気持ち」）、文脈から想像し，答えましょう。
E. 会話の手順：　※50 秒経ったら，会話の途中でも終了します。
(1) 廊下で待つ間は教科書を参照して良い。
(2) 順番が来たら，部屋に入り，Nick に「評価用紙」を渡す。
(3) 制限時間内に，できるだけたくさん質問に答えましょう。
(4) テストが終わったら「評価用紙」をもらって帰りましょう。
F. 結果と成績：　結果と提出物への取り組みは得点化し，1 学期の成績に含めます。
G. 備考 1) 会話の様子は録画します（再判定の必要が出たときに参照するため）。
2) 待っている間は，Speaking Test と直接の関係が無い課題があります。

5 Evaluation form

```
Speaking Test #1                                              2021.6.
Evaluation Form           3年　　組　　番 名前
テストの手順 Testing Procedures：以下のように動きましょう。
① 自分の番が来たら，試験会場へ移動する。この用紙とテキストを忘れない。
② 試験の前に，採点用紙（この用紙）を Jennings 先生に提出する。
③ 質問に答える。※教科書を見て答えたら試験中止です。
④ 試験が終わったら，採点用紙を Jennings 先生から受け取って待機場所に戻る。
⑤ 感想を記入して提出する。※提出が授業後になる場合は千菊まで（必ず当日中に！）。
テスト得点 Scoring Matrix：　※以下は先生が記入するところです。【SET　　　】
Criteria 1. Questions answered:
Question #1  2・1・0
Question #2  2・1・0
Question #3  2・1・0
Question #4  2・1・0
Question #5  2・1・0
Question #6  2・1・0


Criteria 2. Interaction:
verbal reaction / sometimes non-verbal reaction only / Mostly non-verbal


Comments (if any)
☆得点への質問や不満があれば直接千菊まで（ビデオ判定します）。
テストの感想 Self-reflection：　※日本語でどうぞ。（記入欄省略）
```

6 Self-reflection worksheet (The original version in Japanese)

A. Rubric ※Scores are based on "Scoring Matrix" in your Evaluation Form

	Maintain the conversation	Interact in a natural manner
3	Answered more than 4	Showed verbal reactions suitably
2	Answered 3 or 4	Managed interactions sometimes non-verbally
1	Answered 2 or less	Stopped in silence at times

B. Analysis of the performance ※ Circle the Number of the questions you successfully answered.

1 Can answer corresponding to the textbook story   Q3 ・ Q4 ・ Q5

2 Can shows verbal reactions to maintain the conversation   Q1 ・ Q2 ・ Q6

C. The factors which influenced today's performance

※ Check the boxes you must be aware of when studying English

1 ☐ Poor understanding of the story ☐ Incomplete answer ☐ Unable to understand the questions

2 ☐ Lack of necessary reaction ☐ bad timing for reaction ☐ Unable to understand the questions

D. What should you be aware of while studying English?

```
Appendix 2: Speaking Test # 2
```

<u>1</u> Test Specification

A. Test type / Examinees : Achievement Test / The 9th grade students

B. Objectives:

   (1) Students should be able to lead the interaction with the examiner by asking questions to get information about a book.

   (2) Students should be able to make a question for clarification request properly.

   (3) Students should be able to make contribution to the extension of the interaction by giving valid information on the topic.

C. Setting/Roles:

   Examiner = A friend of the examinee who has finished reading a book.

   Examinee = A junior high school student who shows interest in the book the examinee has read.

D. Time: 50 seconds each (Please DO NOT shorten or extend this).

E. Testing Venue: Counseling Room (a small room on the third floor)

F: Procedures:

・Evaluate one examinee at a time. Enjoy interaction with them.

・You have three different cards which has questions to elicit questions from the examinees, and use the cards in order. Two other examinees are waiting behind the examinee in the test, so this is important.

・Grading the performance on the three-point scale using the scoring matrix (See Evaluation Form).

G: Notes:

・You are a friend of the examinee. This is a casual conversation on a daily topic.

・Even if the examinee cannot initiate the conversation at the right timing, please be patient.

・If the examinee skips the clarification request, just let the conversation go.

・You can repeat the question if requested.

<u>2</u> Textbook passage: from Lesson 6 Imagine to act, *New Crown English Series 3*, p.88

Kate: What are you reading?

Riku: A novel about a scientist who invents a time machine and travels to the future.

Kate: Cool. What would you do if you had one?

Riku: If I had a time machine, I would go to the past. I want to see the dinosaurs.

<u>3</u> Scenarios:

The Frame of Dialogue that the examinee can use as a reference:

Students should ask the following questions and continue the conversation:
(Ask Questions) #1 The type of the story:
#2 The message of the story:
#3 Characters:
(Answer)   #4 Title：
(Add information) #5 Contribution:

Choices for the content:

   The students were already told to read the following stories to get necessary information prior to the test day.

・Landmines and Ak iRa (Let's read 2 in *New Crown English Series 2*)

・A pot of poison   (Let's read 1 in *New Crown English Series 2*)

・Zorba's Promise   (Further Reading 1 in *New Crown English Series 2*)

A proto-frame for the exam conversation

ALT: Hi. Please begin.

Student: Is it (a type of book   )？   *Closed-questions only!

ALT: Yes, it is. / No, it isn't.

  *If the Student asks twice but still cannot get the right one, please tell her/him the answer.

Student: Oh, <   reaction by echoing   >. What can we learn from the book?

ALT: (Message of the book   )

Student: <Reaction by set phrase>. Can you tell me about the main character?

ALT: It's about a (character of the book   ) who / which (did something   ).

Student: Sorry, <Question with a Wh-word>

ALT: A man who (did something   )

Student: Thank you. Oh, I remember. It's (title of the book   ), isn't it?

ALT: Yes.

Student: <   Add other relevant information or their impression of the book   >

Teacher's resources   *SET 2 and SET 3 cannot be included due to the limited space.

SET 1   Landmines and AkiRa   ☆(S) = student   (A) = ALT

#1 (S) Is it a non-fiction?   (A) Yes, it is. It's a non-fiction.

#2 (S) <echoing>   What can we learn from the book? (A) I guess many people are working hard to make our world a better place.

#3 (S) <set phrase>   Can you tell me about the main character?

  (A) It's about a man who built the Landmine Museum in Cambodia.

  (S) Sorry, a man built what?   (Or A man built the museum where?)

  (A) He built the Landmine Museum.   (Or He built it in Cambodia.)

#4 (S) Oh, I remember. It's "Landmines and AkiRa," isn't it?

  (A) Yes, it is.   (If the answer is wrong, give them the correct answer.)

#5 (S) I've read it before. I learned that …. (A) <Comments to the statement>

---

4 Handout for the examinees

Speaking Test #2 「Tell me about the book you're reading!」要項
A. 日時と場所：　11 月第 4 週 または 12 月第 1 週の TT ／ 授業開始時に A3 教室に集合
B. 問われる力：　①本について質問して対話をリードできる。
      ②本の内容について話を深めたり拡げたりして，対話の発展に貢献できる。
      ③対話の中で，わからない語句があった時に質問して解決できる。
C. ロールプレイの詳細：
(1)状況　友人が本（以下のいずれか）を読んでいるところを訪れる。声を出して話してよい場所とする。
・Landmines and Ak iRa　・A pot of poison　・Zorba's Promise
・Peter Rabbit　・The story of the Wright Brothers　・Moby Dick
(2)場面　あなたが興味を示し，質問する。
D. 準備の内容：　指定された 3 つの話について，Lesson 6 USE Read の内容程度に英語で紹介できるよう
     に準備しておく。対話のフレームは，授業で練習したものを同じです。
E. 会話の手順：　※50 秒経ったら，会話の途中でも終了します。
(1) 廊下で待つ間は教科書を参照して良い。
(2) 順番が来たら，部屋に入り，Nick に「評価用紙」を渡す。
(3) 制限時間内に，できるだけ会話を先に進めましょう。
(4) テストが終わったら「評価用紙」をもらって帰りましょう。
F. 結果と成績：　結果と提出物への取り組みは得点化し，2 学期の成績に含めます。
G. 備考 1) 会話の様子は録画します（再判定の必要が出たときに参照するため）。
   2) 待っている間は，Speaking Test と直接の関係が無い課題があります。

5 Evaluation form

Speaking Test #4	2021.11.

Evaluation Form　　　　　　　3 年　　組　　番 名前＿＿＿＿＿＿＿＿＿＿＿＿＿＿

テストの手順 Testing Procedures ：
① 自分の番が来たら，試験会場へ移動する。この用紙を忘れない。
　※待っている間にテキストを参照してもよい。
② 受験前に採点用紙（この用紙）を Jennings 先生に提出する。
③ 会話を進める。※教科書などを見て答えたら試験中止です。
④ 試験が終わったら，採点用紙を Jennings 先生から受け取って待機場所に戻る。
⑤ 感想を記入して提出する。※提出が授業後になる場合は千菊まで（必ず当日中に！）

テスト得点 Scoring Matrix ：　↓These will be marked by Mr. Jennings.
Relevant Questions　　　　2(Good)・1(OK)・0(No good)
Clarification Request　　　2(Good)・1(OK)・0(No good / Not made)
Contribution to Extension　2(Good)・1(OK)・0(No good / Not made)
Comments (if any)
　　　　　　☆得点への質問や不満があれば直接千菊まで（ビデオ判定します）。
テストの反省 Self-reflection ：
英語でのやり取りについて，テストを通じて見えた「課題と展望」を書こう＜日本語＞。
（記入欄省略）

6 The self-evaluation worksheet　(The original version in Japanese)

A. Rubric ※Scores are based on "Scoring Matrix" in your Evaluation Form

	Asking Questions	Clarification Request	Providing relevant info about story
2	Two or more	Made in a sentence	Sufficient
1	One	Made but incomplete	Provided, but not insufficient
0	None	Not made	Not provided

B. Analysis of the performance　※ Check ☑ and show how far you have got.
□ #1 Question about the type of book:
□ #2 Question about the message of the story:
□ #3 Question about the main character of the story:
□ #4 Answer (the title) :
□ #5 Contribution by adding relevant information about the book

C. The factors which influenced today's performance
　※ Check the boxes you must be aware of when studying English
1 □ Poor understanding of the story □ Incomplete answer　　□ Unable to understand the questions
2 □ Poor use of expressions □ bad timing for reaction □　Unable to remember the expressions

D. What should you be aware while practicing interaction in English classes?