

**Q: 自己評価はなぜ必要なのですか？ Is it good to let students evaluate their development of writing by themselves? I truly understand the importance of self-evaluation, however, isn't it essential to have a sample description of the category in an evaluation criteria?**

A: Your first question: "I just wonder if it is good to let students evaluate their development of writing by themselves."

Students' self-assessments are not expected to skillful analyses of their writing. Rather, self-assessment is part of students' learning process.

Reasons for self-assessment

- ✧ Students develop reflective self-awareness of their own learning process. That awareness causes most students to take more interest in improving their writing. When self-assessment is done regularly, students can see the improvement in their writing, which tends to increase their confidence and motivation.
- ✧ Students learn qualities of good writing by seeing them listed in the self-assessment table and thinking about them with each self-assessment.

Your second question: "Isn't it essential to have a sample description of the category in an evaluation criteria?"

I think you are referring to a rubric, which is the assessment tool used by teachers to assess students' English production (writing and speaking) for test purposes. Please see the model high school rubric below for assessing a "fun essay." As you see, it includes descriptions of the levels of accomplishment for each criterion. Teachers should also show the rubric to students early in the year or term and again as test time approaches so they understand the goals they are expected to reach.

Table 11: Modified version of the rubric for the speaking test – 20 points (December 2004)

Fluency & Content	10 points	(10) be able to maintain 3 minute-conversation fluently, with good content (7) be able to maintain a 3 minute-conversation with some silence, with adequate content (4) be able to maintain a 3 minute-conversation with some silence, with poor content (1) be hardly able to maintain a 3 minute-conversation with some long silences
Accuracy (grammar & pronunciation)	3 points	(3) be able to communicate with accuracy (2) be able to communicate with some errors (1) communicate with many errors, using mainly key words
Delivery (volume & eye contact)	3 points	(3) be able to speak with good volume and eye contact (2) occasionally speak with adequate volume and eye contact (1) be hardly able to speak with adequate volume and eye contact
Strategies (conversation strategies & follow-up questions)	4 points	(4) be able to use many conversation strategies and follow-up questions (3) be able to use some conversation strategies and follow-up questions (2) use a few conversation strategies and follow-up questions

		(1) be hardly able to use conversation strategies and follow-up questions
--	--	---

Sato & Takahashi (2008)

Nancy Mutoh (2016)

**Q:やる気を高めるための評価はどのように行いますか？ *I wonder how you assess the students. I'm sure you need to give them grades. Also, some studies show that assessment plays a great part in students' learning.***

A: Of course teachers need to assess their students and give grades, me included. By presenting in our workshop different approaches to “liking” in classroom interactions, activities, and coursework, I’m suggesting that teachers occasionally mix their approaches, and yes, sometimes flip the meaning of assessment. Flipping traditional assessment does not have to be done for every activity or assignment. What does “test” mean? In one sense, when marking assignments simply as completed or not, that in itself is kind of a test, meaning a verifiable and quantifiable challenge to the students. It asks of the students, “Did you do it or not?” and “Can you keep doing the required work consistently and in a timely fashion?” Actually, many of us teachers are already tracking similar things, such as in maintaining records of attendance, tardiness, and absence, whether for a grade or not. This form of assessment is also commonly promoted in extensive reading and involves checking whether students regularly read books, with a tally of how many books each student has read by the end of the semester. Why not assess similarly for extensive speaking or writing? Journal entries, freewriting sessions, and extemporaneous speeches would make good candidates for this type of assessment. Can you think of any more?

More importantly, by placing less focus on grammatical forms and error corrections, more time and space can be opened up for investigating the ways of meaning-making. “Liking” can promote further and better meaning-making by inviting teachers and students to share with each other in asking, “What does this mean to you?” and “How do you feel about it?” and “How might we change this or add to it to make a clearer meaning, a better idea, or an expansion upon it?”

I myself like to mix and match various teaching approaches and forms of assessment, all the while thinking about the particular course objectives and individual psychologies and group dynamics in each situation. Here are just a handful of elements that often enter into the mix for my classes (in no special order): Formative assessment; process-based writing, presenting, and conversing; portfolio assessment; negotiated syllabus; project-based learning; strategy training; and peer feedback. Also, my students and I learn a lot from each other through the process of critical participatory looping (CPL; Falout & Murphey, 2010; Murphey & Falout, 2010), which could be conceived as a kind of ongoing, iterative appreciation and assessment of the classroom that encourages student voice and student-directed learning. CPL can also be used for gathering data and providing validity in action research (Falout, Murphey, Fukuda, & Fukuda, 2016; Murphey & Falout, 2012). All of these methods bring opportunities for sharing appreciation of efforts in coursework, qualities of character, and exchanges of opinions.

The key things to remember are that there is no single best way for teaching, action research helps us make informed decisions, and that each one of us teachers can uniquely mix and match as many approaches in teaching and forms of assessment as we (and maybe our students) choose for our own particular situations. As far as “liking” goes, this is just one extra element that might be included in the mix. Even a few written exercises or speaking activities for “liking” without focusing on error correction or using the traditional forms of evaluation might be fit into a busy semester.

“Liking” frees up the pressures of being judged by others that can make students self-conscious and embarrassed when using English. Fear of being judged poorly and humiliating experiences in front of teachers and classmates seem to be notable causes, respectively, of silence in the classroom (King, 2013) and demotivation in learning (Falout & Falout, 2005; Falout & Maruyama, 2004). “Liking” provides students ideas for what they might try again or develop further with their English. In my experience, “liking” helps students gain confidence, positive affect (emotions), fluency and duration (talking and writing longer), depth of expression and meaning, and even accuracy in their speaking and writing. They work harder on their English because they themselves like their English even more when others like it too. Given training and practice for showing appreciation among themselves, “liking” also seems to help improve the quality of student interactions and the explicit teaching of other classmates.

(Joseph fallout 2016)

**Q: CLT において、生徒の評価には評価基準を設けるべきでしょうか。 Ideally, should we use the same standard for assessment?**

I think this question means: Should we use absolute (criteria-based) assessment standards for evaluating students' learning?

Answers to this important question are various. In Japan, MEXT changed the national policy on assessment in schools from relative standards (that compared the students in each class to each other - setting percentage of 5's, 4's, 3's, 2's and 1's) to absolute standards that base grades on fixed criteria of achievement (sometime phrased as “Can do” statements). Some of the workshop participants at that time told me that teachers in each subject were working hard to develop those criteria-based standards.

Criteria-based standards are fairer for each learner and support education by setting explicit learning goals at each grade. If the goals are told to students, students will understand what to aim for. Students can be partners in learning rather than competitors.

Assessment is a subject that we should all learn more about because assessment shapes learning. Telling students that the end-of-term testing includes a speaking test and process writing essay causes most students to work on those skills more diligently throughout the term. And showing students the specific criteria that their test conversations and writing will be judged by teaches them the qualities of good conversation and good writing.

Nancy Mutoh (2015)

**Q: スピーキングの評価をどのようにするのか教えてください。 I wanted to know more about how teachers evaluate student' speaking abilities.**

**A:** I ask my ALT to conduct speaking tests, having prepared all the materials needed for speaking tests (handouts and rubrics) and given enough scaffolding to my students to make them comfortable with the test formats. However, I can't give this kind of more-or-less interactive speaking tests with the ALT very often because they are very time-consuming. So, I've started to incorporate peer assessment to evaluate speaking abilities in such cases as making speeches. I have each student in groups of five make a speech in turn, and I ask the leader to evaluate if each member was able to make a speech smoothly without looking at his/her handout, based on the criteria spelled out in my rubric.

福元有希美 Yukimi Fukumoto (2014)

**A:** There is no one ideal way. These are steps, I think, teachers need to do. Teachers need to (1) clarify

the goals or objectives (CAN-DOs), (2) decide the way(ways) of testing students achievement, (3) consider the criteria, (4) make a rubric with as concrete descriptions as possible, (5) share the rubric with other teachers, (6) try out the rubric, finally (7) conduct a speaking test and evaluate their abilities.

奥田紀子 Noriko Okuda (2014)

Q: **ルーブリックとは何ですか？** Please give more information about making rubrics.

**A: Borrow rather than make!**

Making rubrics that are clear in meaning and easy to use is much more difficult than it appears. Whenever possible, use a rubric that has already proven to be useful for other teachers. You can modify a borrowed rubric to make it fit your needs better. You looked at the new series of books on assessing language performance 「英語授業を変えるパフォーマンス・テスト」佐藤一嘉編者 明治図書 (2014) at the workshop. These are full of many types of useful, classroom-tested rubrics for assessing speaking and writing. The fastest, most successful way to start performance testing is to get the book for your teaching level and use those rubrics: 中学校 1、2、3 and 高校. Doing so will get you and your students off to a better start.

#### Parts of a rubric

A rubric is usually made using a table (表). There are various formats for rubrics. One format is to list the *criteria* you use to assess the speaking or writing in the left column with the *degrees of success* in each criterion listed across the top line (often stated in points or as A, B, C. The remaining squares usually contain *descriptors*, short phrases that help assessors decide whether to give, for example, 2 or 3 points for a particular criterion. A different kind of rubric can be found at <http://www.pearsonlongman.com/ae/worldview/wvvideospeakingrubric.pdf>

This is a *proficiency bands* type of rubric. It doesn't allow the assessor to score different criteria separately, so it is less useful for feedback and ongoing learning. This kind of rubric is most suitable for standardized proficiency testing and placement decisions.

The following rubric is included in [Sato, K., & Takahashi, K. \(2008\)](#).

(<http://www3.nufs.ac.jp/~yoshi/paper/Sato%20&%20Takahashi%20%282008%29.doc> ). It has been used (with and without modification) by many teachers.

Table 10: Rubric for the speaking test – 20 points in total (July 2004)

Fluency	8 points	<ul style="list-style-type: none"> <li>• asking questions (2, 1, 0)</li> <li>• response rate (2, 1, 0)</li> <li>• answer in more than two sentences (2, 1, 0)</li> <li>• maintain two-minute conversation (2, 1, 0)</li> </ul>
Accuracy	2 points	<ul style="list-style-type: none"> <li>• grammar (1, 0)</li> <li>• pronunciation (1, 0)</li> </ul>
Delivery	4 points	<ul style="list-style-type: none"> <li>• volume (2, 1, 0)</li> <li>• spontaneity (not memorization) (2, 1, 0)</li> </ul>
Strategies	4 points	<ul style="list-style-type: none"> <li>• How ya doing? (1, 0)</li> <li>• Nice talking with you. (1, 0)</li> <li>• Shadowing (1, 0)</li> <li>• How about you? (1, 0)</li> </ul>
Impression	2 points	<ul style="list-style-type: none"> <li>• Impressive (2, 1, 0)</li> </ul>

Table 11: Modified version of the rubric for the speaking test – 20 points (December 2004)

Fluency & Content	10 points	(10) be able to maintain 3 minute-conversation fluently, with good content (7) be able to maintain a 3 minute-conversation with some silence, with adequate content (4) be able to maintain a 3 minute-conversation with some silence, with poor content (1) be hardly able to maintain a 3 minute-conversation with some long silences
Accuracy (grammar & pronunciation)	3 points	(3) be able to communicate with accuracy (2) be able to communicate with some errors (1) communicate with many errors, using mainly key words
Delivery (volume & eye contact)	3 points	(3) be able to speak with good volume and eye contact (2) occasionally speak with adequate volume and eye contact (1) be hardly able to speak with adequate volume and eye contact
Strategies (conversation strategies & follow-up questions)	4 points	(4) be able to use many conversation strategies and follow-up questions (3) be able to use some conversation strategies and follow-up questions (2) use a few conversation strategies and follow-up questions (1) be hardly able to use conversation strategies and follow-up questions

Nancy Mutoh (2014)

**Q: コミュニケーションの評価はどのようなものでしょうか。 We talked about communicative assessment but it was largely summative assessment. I would like to hear about assessment methods other than tests and how to incorporate them into classes on a daily basis. I am also interested in formative assessment.**

A: Formative, or ongoing, assessment can be grouped into three types, I find.

*One type* is visible: the explicit “short test” (quiz) type of assessment that is common, the teacher walking around looking at students’ work as they do it, teacher feedback on student writing or worksheets, etc.

*The second type* is assessment that is woven into learning activities.

- (a) One example of this second type of assessment is in the teaching of grammar. With “focus on form instruction” (also called “structured input” of grammar), a student’s success in doing the “intake task” shows to what extent the student has *understood* the meaning of the new target grammar point; while success in the output task shows to how well the student has understood the *form and use* of the grammar point.

For information about this way of teaching grammar, please see the books containing grammar lessons written by teachers of JHS 1, 2, and 3 and SHS, フォーカス・オン・フォームでできる！ 英文法指導アイデアワーク edited by Sato, K. and published by 明治図書, 2012. Also click on the two links below to see published journal articles about this approach:

[Sato, K., Fukumoto, Y., Ishitobi, N., & Morioka, T. \(2012\).](#) Focus-on-form instruction and student learning in Japanese junior high schools. In A. Stewart & N. Sonda (Eds.), *JALT2011 Conference Proceedings* (pp.283-303). Tokyo: JALT.

And, about high school grammar teaching: (Please remove the space between these two lines.)

[Sato, K., Iwai, R., Kato, M., & Kushiro, M. \(2009\).](#) Focus-on form instruction (FFI) and its effect on student learning. In A. M. Stoke (Ed.), *JALT2008 Conference Proceedings* (pp. 521-548). Tokyo: JALT.

- (b) Another example is assessment that is part of an integrated-skills, information-exchange chain activity. Partners have a conversation asking and answering questions about, say, sports they like (sports they play, how often they play, sports they like to watch, how often they watch, and their favorite team). As they chat, they write very brief notes (one- or two-word answers) in an information table that has space for the student's own answers and those of each partner spoken with (often 3 partners). Students can then be asked to verbally report the information they collected to a small group of different students or write the information in a paragraph. By successfully doing this chain activity, students generate visible evidence of their ability to use the target language in a real conversation. Target grammar and previously-learned grammar, vocabulary, pronunciation, spelling, etc. can be assessed at once. The purpose isn't to collect gradable data, but to learn what follow-up review or teaching is needed for which students. Teachers have always done this, but integrated skills activities provide natural, fruitful opportunities at the same time that students are getting good communication practice.

Here is a typical information table format.

Collect the answers of three classmates. Then share your answers with other classmates.

	You	Name	Name	Name
1. What sport do you play/do?				
2. How often do you play/do it?				
3. What sport do you like to watch?				
4. How often do you watch it?				
5. What team is your favorite?				

*The third type* is assessment that is conducted by students. Students learn important things from these types of assessment.

- (a) Peer-assessment is often used in writing. Each student gives their partner their feelings as a *reader* of the text (not as an editor). Peer assessors write a star beside sentences that have especially interesting content or are especially well-written, a question mark beside places that

are difficult to understand and the word “more” beside places where the peer-reader would like more detail or an example. Then the peer assessor writes one positive comment about the text. Many teachers have students exchange papers with three classmates. This amount of feedback is good stimulus for writers as they prepare to write their second draft, and is especially helpful for expanding the content of their text. It also causes readers to read more thoughtfully and to think about what good writing is.

- (b) Self-evaluation of a student’s own conversation or writing. These are very simple, requiring only 2 or 3 minutes. By doing these regular self-evaluations, students learn the qualities of good speaking and writing because the qualities to be self-assessed are listed by the teacher. They also track their own development and set goals for improving specific aspects of their speaking or writing. Here is a sample self-assessment format.

評価項目		自己評価			今回の会話の感想・気が付いたこと・分かったこと:
		A	B	C	
Eye contact	アイ・コンタクト	A	B	C	
Voice	適当な声の大きさ	A	B	C	
Used CSs	C S の利用	A	B	C	
Smoothness	スラスラ話せた	A	B	C	

Nancy Mutoh (2014)

**Q: リーディング能力を測るテストの作成はどのようにしたらいいのでしょうか。It is not easy for me to make a test which can assess students’ reading ability. If you could give me any suggestions, I would be very grateful. (I am teaching at a high school now.)**

A: I agree with you that writing good tests is not easy!

“Reading” to most people means to get the meaning from a written text. When you choose a text for a reading test, therefore, it needs to be one that the students haven’t seen before and that is close enough to their current level of vocabulary and English structure that they have a fair chance of understanding it with reasonable reading skills. (Using the passage from the textbook is a test of memory, not a reading test.) The test passage needs to be short enough that students have enough time to read it needs enough content to base questions on. Some points to keep in mind:

- (a) Avoid questions based on one sentence in the text because they can often be correctly answered without understanding the meaning of the sentence. They can often be answered by using general knowledge of English grammar. This is especially so when both the question and answers are also in English. I have found that many textbook comprehension questions have this problem.

Please take this reading test on a short excerpt from *Jabberwocky* (in *Through the Looking-Glass and What Alice Found There* by Lewis Carroll, 1872).

Instructions: Read the four-line text below and answer the three questions.

‘Twas brillig, and the slithy toves  
 Did gyre and gimble in the wabe:  
 All mimsy were the borogoves,

And the mome raths outgrabe.

Questions

1. What did the slithy troves do?
2. How were the borogoves?
3. What did the mome raths do?

The correct answers to the test are at the end of this reply to 7).

(b) Better questions are based on understanding two or more places in the text. Questions like:

~Why did Toby want to go the city?

~What did he learn from his week there?

~How did Toby feel after that experience?

~What is the main idea of this passage?

(These draft questions and the intended correct answers need checking by colleagues because the first draft of such a questions often has two possible answers if multiple choice.)

(c) The choice of using Japanese or English or a mix in the questions and answers needs thought.

Things to consider: Straight translation questions about particular words, phrases or a sentence don't show whether a student has understood the text as a whole. Multiple choice questions in English let students scan the text for words in the question. Open-ended questions that ask students to answer in Japanese allow them to more clearly communicate their understanding to the teacher. A reading test score shouldn't depend on English writing ability. For the same reason, if answers to reading test questions are written in English, points should not be deducted for English language mistakes in the answers.

(Answers to the reading test: 1. They gyred and gimbled. 2. They were mimsy. 3. They outgrabe.

The words are nonsense words, so we can't get any meaning from this text. However, all of the teachers and students who have taken this text scored 100%.)

Nancy Mutoh (2014)

**Q: 評価方法を変えるために、もっとも大切なことは何でしょうか。What's the most important thing to change the situation in testing from now on? Maybe talking with our colleagues about this...**

A: Yes, that's it! Within one school, colleagues can significantly improve students' learning by

- 1) being clear about, and agreeing on, the goals of their overall English program and each course in the three years, and
- 2) planning types of term-end assessment (performance tests and sections of the paper test) that match the goals, then
- 3) planning the curriculum for each course so that students are lead, step by step, to succeed in the ongoing (formative) assessment tasks and the final assessment tasks, and finally
- 4) assess your own program by examining results of final assessment, student surveys, reflections of teachers, etc. to identify the strengths and weaknesses in the program and make a plan to improve it.

Nancy Mutoh (2014)

**Q: 生徒の批判的思考(critical thinking)をどう評価したらいいですか? Are there any good ways to assess students' Critical Thinking skills?**

**A:** One of the CT skills, 'differentiating facts and opinions', can be assessed rather easily, which I will explain in the answer to Question 6. Also, students' CT attitudes or awareness can be assessed through questionnaires. You could ask them:

1. Did you understand the issue clearly?
2. Did you have chances to think about it deeply?
3. Were you able to form your own opinions?
4. Did you enjoy exchanging opinions with your classmates?
5. How did you feel when your classmate had a different opinion from yours?
6. Did you have any new findings or awareness about this issue?
7. Did this activity (or lesson) change your way of thinking? Why or why not?

You could ask such questions either in Japanese or English (depending on their language levels) and get feedback from them. This is one way of assessing how much your lesson has helped them enhance their CT attitudes. I would not use the questionnaire results, however, for evaluation or grading purposes.

For your reference, I suggest reading the following paper. It is for L1 situations and not really for EFL classes, but you can get good hints. I found it very interesting. (It is written in Japanese)

南学 Manabu MINAMI

「クリティカルシンキングをうながすゲーミング教材の開発と評価」

“The development and estimate of a game material to promote critical thinking”

三重大大学教育学部研究紀要第 64 巻教育科学 (2013) 337- 348 頁

<http://miuse.mie-u.ac.jp/bitstream/10076/12347/1/20C16238.pdf>

山中純子 Junko Yamanaka (2014)