

思考力を促進する英語タスクの開発

大野 秀樹 (大東文化大学経済学部准教授)

Development of EFL critical thinking tasks.

Hideki OHNO

Abstract

This study illustrates the significance of developing English as a Foreign Language (EFL) tasks dealing with critical thinking (hereafter CT). These CT tasks are intended for use at the tertiary level in EFL settings. These will be used for formative assessment of CT skills and can be a proxy for CT tests. The component of task development is to 1) select descriptors for CT skills, 2) create tasks based on the descriptors, 3) examine the tasks, 4) select and administer the tasks, and 5) validate the tasks. They are run in a cyclic way to produce optimal CT tasks.

1 Introduction

In EFL settings, there is a plethora of research into task-based language teaching (TBLT). However, little research has attempted to create EFL tasks that aim to foster CT. Some of the current EFL task research include “reasoning demands” (Skehan, 2018, p. 127), which is essential part of critical thinking. Yet the discussion in those studies is over to what extent reasoning is required in the task, not how to facilitate reasoning. The main thesis of this study is that CT tasks can foster the aspects of reasoning, rationality, or higher-order thinking.

Usually, textbooks designed to teach CT contain CT questions and activities. However, they are not in the form of tasks in the sense of TBLT (see East, 2017). This is understandable given the nature of those textbooks; they are not developed in relation to TBLT.

A framework dealing with intercultural competence (Candelier, Camilleri Grima, Castellotti, de Pietro, Lőrincz, Meißner, Molinié, Noguero, and SchröderSura, 2012) does have CT aspects (Ohno, 2015), although the authors never use that term. They contain valuable descriptors, and based on the descriptors, provide intercultural competence classroom activities (see ECML, n.d.).

Therefore, it may be useful to use them as CT tasks. Still it is hard to regard the activities as CT tasks, partly because it is not clear whether the connection between the descriptors and the activities is strong. It might seem irrelevant to point out this issue here, considering that they never use the term critical thinking, nor are they developed within the TBLT framework. As these activities are based on descriptors, however, it gives a valuable suggestion for CT task development; a valid CT task should be fully based on CT descriptors.

2 Task development

As mentioned above, task development in this study starts with choosing descriptors for CT skills. As shown in Figure 1, it is followed by “Creation of tasks,” in which tasks are created based on the descriptors. The next phase is “Examination of tasks,” in which the created tasks are examined by professionals to see if the descriptors are fully reflected in the tasks, and if the tasks have task characteristics. Then comes the administration of tasks targeted for the EFL tertiary level. The final phase is “Validation of tasks,” which examines the validity of the tasks by external tests or among the created tasks. This paper shows the first three phases of CT task development.

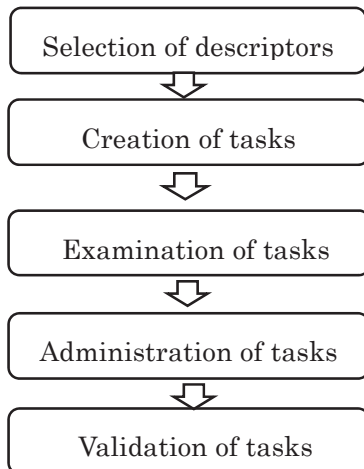


Figure 1. Process of CT Task Development

Selection of descriptors

The candidate for a descriptor list is Facione (1990). It is a CT framework created in a bid to improve CT skills and attitudes of (mainly L1 university) students. What constitutes CT was extracted via the Delphi method from forty-six researchers engaged in CT. This framework has

dimensions of skills (Interpretation, Analysis, Evaluation, Inference, Explanation, and Self-regulation) and disposition, and has descriptors, such as “to recognize a problem and define its character without prejudice to inquiry (“Interpretation” above).” Based on the descriptors, CT tests were developed and they have now widely been recognized, including translated versions. If these are used as CT “tasks,” it will be useful for EFL learners of English. However, these CT tests are not originally intended for EFL classroom tasks, nor have they been made public. Therefore, creating CT tasks based on sound framework is needed for EFL learners.

Stanovich (2011) is a powerful index for the creation of CT tasks in this study. They use the term “rationality” to refer to CT, and they created a framework of rationality based on the sound research findings of cognitive psychology. They point out that “reflective mind” is key to understanding higher-order thinking, such as rationality.

Creation of tasks

Once descriptors are selected, the next step is creation of CT tasks. Skehan (2018, p. 13) defines tasks as “an activity in which: meaning is primary; there is some sort of relationship to the real world; task completion has some priority; and the assessment of task performance is in terms of task outcome.” In this study, this definition is employed in addition to basing each task on CT descriptors.

The next issue is in what way task are presented—“task characteristics” (Skehan, 2018). In a similar vein, Liu, Frankel, and Roohr (2014) provides useful information, introducing the types of CT tests conducted in L1 settings. While they argue “the constructed-response type is more expensive to score and the multiple-choice type can measure the same constructs equally well in some cases (Liu et al. 2014, p.11),” they point out the former is based on real-life situations. Therefore, constructed-response items will be preferable for CT tasks in EFL contexts so it will fit the above task definition.

Furthermore, the aspects of “Input, Ability Tested, Instructions, Expected Response, and Marking” are taken into consideration, following ALTE (2001). These items are intended to create tests, but they are also useful to create tasks, as each item of tests can be regarded as a task.

To create CT tasks ideal for this study, appropriate prompts and directions are needed. For this, Kroll and Reid (1994) was used as a reference point. From the viewpoint of writing instructions, they argue “a well-developed prompts achieves its purpose of ranking, sorting, or placing students in a way that appears to be reflective of their “true” writing ability (Kroll and Reid, 1994, p. 248).” Along this line, this study aims to create tasks with a well-developed prompt, using framed prompts and text-based prompts.

Based on the above characteristics and definition, CT tasks are created. The following task is one such example.

CASE

Descriptor: To formulate multiple alternatives for resolving a problem

Task (20 min.)

Read the following passage and prompts, then answer the questions.

At 7:30 pm on a winter day, Mr. A was walking on one of the busiest streets in Tokyo and was hit by a bicycle. However, the bicyclist left the scene of the accident without even apologizing to Mr. A. Mr. A got a bruise and was desperate to find the bicyclist. In the next morning, he saw a man coming out of a store around the scene of the accident. Mr. A thought that man was the suspect and tried to complain to him (Mr. B). Mr. A is confident that Mr. B is the person who hit him with a bicycle the day before. However, Mr. B denies Mr. A's statement.

Suppose you are in a position to ask Mr. A about his statement.

Q1 To see if Mr. A's statement is correct, what questions would you ask him? Ask more than two questions.

Q2 Provide reasons for why you ask the questions.

When you ask questions, 1) please assume Mr. A is not lying; 2) you are not allowed to ask reasons for Mr. A's conclusion that Mr. B is the suspect.

Parameters 1) and 2) above respond to the question as to whether the task fully reflects the descriptor. They both restrict certain kinds of questions, and lead learners to probe more relevant questions, which prompts them to conjure up "alternatives." In terms of general task characteristics, this task meets all the conditions above. In this manner, descriptors were chosen and CT tasks were created.

Examination of tasks

After the task creation is the examination of the tasks. The created tasks need to be examined by professionals to see if they are valid tasks to be administered. Each set of task will undergo an examination by professionals. The following six items are asked to researchers and practitioners in the field of (applied) linguistics (1 being Strongly disagree, 2 Disagree, 3 Neither

agree nor disagree, 4 Agree, and 5 being Strongly agree).

1. Do you think this task is related to its descriptor(s)? Please choose the best answer.
2. Suppose your students* work on this task. Do you think this task is difficult for them, when they are not allowed to use a dictionary? Please choose the best answer.

*You may have various types of students in several different classes in one university or several different universities. You can pick up an average student among them and assume that student takes on this task.

3. Suppose your students* work on this task. Do you think this task is difficult for them, when they are allowed to use a dictionary? Please choose the best answers.

*same as above

4. Do you think the content* used in this task affects the difficulty level of this task? Please choose the best answer.

*content refers to the content knowledge of economics, medicine, culture, etc.

5. In order to strengthen the connection between this descriptor(s) and this task, how would you modify this task? Please use the same descriptor(s) while modifying the task. Modifying includes the time allocated, mode of language skills, addition of some other activities, or addition or deletion of some phrases, etc.

6. Using the same descriptors, how would you change the difficulty level of this task? Changing the difficulty level includes the time allocated, mode of language skills, addition of some other activities, or addition or deletion of some phrases, etc. To make this task more/less difficult, you would ()*. *Within () comes their suggestions.

Question 1 concerns the content validity. This word is used primarily in the field of testing. However, any task can be treated as a test item, hence the use of this term. Only a task that exceeds certain standards (see Stanovich, West, and Toplak, 2016) will be selected for the administration. Question 4 asks if content knowledge is needed for the task. The question 5 and 6 are used to adjust the difficult level of tasks This is about the perceived difficulty for students seen from teachers.

3 Scoring and example answers

Once tasks are administered, the next step is the scoring of tasks. As stated in the previous section, most of the tasks are expected to be “construct-response” types. It is thus necessary to choose criteria for evaluating each task, not using a set of evaluation score.

Furthermore, as tasks in this study are not assumed to be a multiple-choice test type, holding raters' conferences is important. Two raters, including the author, have had one conference and several email-discussions so far.

Not all the tasks are based on the same descriptors, and therefore depending on the types of descriptors, evaluation should vary from task to task. After evaluation criteria is chosen for each task, students' answers in each task can support the evaluation, and their answers amount to enriched collections of example answers. Practitioners will benefit from this wide range of model answers, when they give their students feedback. What is behind this principle is Ennis (1985), in which evaluation criteria and example answers are provided for nine questions required. Although the Ennis test (task) is perhaps intended for L1 English speakers, this type of test/task with model answers would be useful for EFL learners and practitioners in that feedback is readily available.

In the scoring process, what is of importance is to what extent learners use words or phrases relevant to the task. Therefore, complexity, accuracy, and fluency are of secondary importance, so long as raters understand the meaning. For example, "During Tokyo Olympics garbage problem" was acceptable by two raters. It was interpreted as "During the Tokyo Olympics, the amount of garbage will increase." It was a part of the task that requires EFL learners to generate the scenario concerning what will happen after "Japan hosts the Tokyo Olympics in 2020."

4 Conclusion

This paper has demonstrated the importance and manner of CT task development in an EFL context. Based on selected CT skill descriptors, tasks are created. Selection of descriptors and basing the created tasks on sound cognitive psychology background is important. Before administering the tasks, examination of task is necessary to ensure that each task has a strong relation with its descriptors.

Validating these CT tasks with external CT tests is preferable to demonstrate that these tasks are valid. However, there are constraints on the use of CT tests, because most of current CT tests are intended to be used in L1 settings. Therefore, administering those tests in English is not feasible in EFL settings given that the level of English in the tests prevent EFL students from exercising their CT skills. Therefore, a more realistic solution would be to find inter-task correlations.

Evaluation criteria for each task is developed so the function or intended meaning of tasks will be fully maximized. Some of learners' answers, upon permission, will be described in a task

manual so they will benefit practitioners when they give their students feedbacks. Example answers will be updated using the students' answers for each task. To build a system that keeps updating possible examples of correct answer can be a strength of this study.

The whole concept of this study is different from some TBLT practices in which grammar is intended to be fostered through authentic tasks. This study, however, aims to foster CT skills through authentic tasks, and evaluate to what extent relevant information is provided in each task. Therefore, providing linguistic input or giving students a permission of using a dictionary can be part of task completion. Once this descriptors-task-evaluation triad is established, formative evaluation is easily conducted.

Acknowledgments

This research was supported by Grant-in-Aid for Scientific Research (17K02899). The idea for this research was conceived during the sabbatical leave in 2016–2017.

References

- ALTE. (2001). *Individual component checklist for use with one task: Writing*. Retrieved from <https://www.alte.org/resources/Documents/Writing%20Checklist%20for%20one%20task.pdf>
- Candelier, M., Camilleri Grima, A., Castellotti, V., de Pietro, J-F., Lőrincz, I., Meißner, F-J., Molinié, M., Noguerol, A., & SchröderSura, A. (2012). *FREPA-A framework of reference for pluralistic approaches to languages and cultures: Competences and resources [PDF File]*. Retrieved from <https://www.ecml.at/Portals/1/documents/ECML-resources/CARAP-EN.pdf?ver=2018-03-20120658-443>
- East, M. (2017). Research into practice: Council of Europe The task-based approach to instructed second language acquisition. *Language Teaching*, 50, 412–424. doi: 10.1017/S026144481700009X
- ECML. (n. d.). Teaching and learning materials. Retrieved from <https://carap.ecml.at/Database/tabid/2313/Default.aspx>
- Ennis, R.H., & Weir, E. (1985). *The Ennis-Weir Critical Thinking Essay Test*. Retrieved from https://www.academia.edu/1847582/The_Ennis-Weir_Critical_Thinking_Essay_Test_An_Instrument_for_Teaching_and_Testing
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction [PDF File]*. Retrieved from <https://www.insightassessment.com/content/download/623/4009/file/DEX+2014+reprinting.pdf>
- Kroll, B., & Reid, J. (1994). Guidelines for designing writing prompts: Clarifications, caveats, and cautions. *Journal of Second Language Writing*, 3, 231–255. doi: 10.1016/1060-3743 (94)

90018-3

- Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. *ETS Research Report*, RR-14-10, 1-23. doi: 10.1002/ets2.12009
- Ohno, H. (2015). Examining frameworks of critical thinking for language education at the tertiary level in Japan. *Daito Bunka University Kiyō*, 53, 43-57.
- Skehan, P. (2018). *Second language task-based performance: Theory, research, assessment*. New York: Routledge.
- Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York: Oxford University Press.
- Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The Rationality quotient: Toward a test of rational thinking*. Cambridge, MA: The MIT Press.