

英語要約採点における認知的負荷が採点者の意思決定に及ぼす影響
— 眼球運動測定装置を用いた心理測定的エビデンスの検証に向けて —
Influence of Cognitive Load on Raters' Decision Making in Scoring English
Summaries: Toward an Examining Psychometric Evidence Using Eye Tracker

加藤万紀子 (東北大学)
Makiko KATO (Tohoku University)

In educational settings, interest is growing regarding the assessment of integrated language skills, such as summary writing. Summarization is considered one of the most difficult academic skills to acquire for learners of English as a Foreign Language (EFL) because it involves multiple abilities, such as reading and writing. When assessing EFL learners' summaries, an analytic rating scale is pedagogically appropriate because teachers can provide feedback. However, the addition of various abilities and skills complicates the assessment; thus, issues remain regarding the reliability and validity of analytical assessments and their practicality in educational settings. In this talk, I will introduce my outline of a four-year project, made possible through Grant-in-Aid for Scientific Research (KAKENHI), to develop an analytic rubric for EFL learners' summaries. I will present progress thus far and future issues to be addressed through research.

As a first step, I used an analytic rating scale to investigate what trained raters with diverse backgrounds found difficult and their strategies for evaluating summaries written by English learners with various English language abilities. These investigations were conducted using open- and closed-ended questionnaires after the raters completed the scoring summaries. Seven trained raters evaluated 36 English summaries written by Japanese university students using an analytic rubric containing the following items: (1) integration (INT), which includes two phases: (a) whether the statements in the summary are written logically and (b) whether the statements have a global interpretation; (2) language use (LU), which includes appropriate grammar and vocabulary and the use of sophisticated syntactic structures; and (3) source use (SU), which includes (a) whether the summary is written in the writer's own words and (b) whether the information in the summary is included correctly. The results showed that the raters did not find it difficult to score LU, but they did find it difficult to score INT and SU. These findings can be explained as follows. First, raters were required to repeatedly review the rubric, source text, and written summaries while evaluating the summaries. Second, the evaluation items included elements from two phases. Additionally, the scores given by the raters for INT and SU seemed subjective, especially because of the difficulty in judging the extent of plagiarism. Furthermore, recognition of ratings and score decisions did not differ by raters' native languages; rather, they tended to differ depending on the raters' experience. However, there were no significant differences between raters in strategies used during the evaluation.

As the next step, I designed a tentative rubric for English summary writing with reference to the raters' comments and the results of inter-rater and intra-rater reliability tests performed in the first step. After revising the descriptors of each of the evaluation items in the rubric used in the first step, I wished to identify items that were difficult to assess in the revised descriptors and evaluation items, as well as the causes of such difficulties. Therefore, I determined that eye movements of the raters could be examined using an eye-tracking device to identify where and to what extent cognitive load was applied by the raters when they evaluated each item.

This research was supported by JSPS KAKENHI Grant Number JP 22K13174.