
TITLE

Complexity, accuracy and fluency (CAF) features of speaking performances on Aptis across different levels on the Common European Framework of Reference (CEFR)

This is a summary of a report by Xun Yan, Ha Ram Kim and Ji Young Kim as part of the ARAGs Research Online Series. For a copy of the full report, see: www.britishcouncil.org/exam/aptis/research/publications/

WHAT WE LOOKED AT:

Second language (L2) proficiency has long been recognized as a multi-componential and multi-dimensional concept, comprising three main components: linguistic complexity, accuracy, and fluency (often referred to as the CAF features). As such, CAF features have been widely used to characterize test performances and test-taker proficiency levels in both L2 speaking and writing assessments. Our aim in this study was to investigate the CAF features of speaking performances on Aptis across different levels on the Common European Framework of Reference (CEFR). Using a corpus-based approach, this study examined (1) the relationships amongst CAF features and holistic scores of speaking performance on the Aptis test; (2) CAF features that characterize and distinguish speaking performances across different CEFR levels.

Specifically, the project addresses two research questions (RQs):

RQ1: *What CAF features in Aptis speaking performances are associated with different CEFR levels of speaking ability?*

RQ2: *Do test takers across different CEFR levels display systematic differences on sub-components of CAF features on the Aptis speaking test?*

HOW WE DID IT:

To answer the research questions, we transcribed and analyzed 500 benchmark speech samples from 125 examinees on the Aptis speaking test (25 examinees from each level of A1-C, and four speech samples from each examinee). Fourteen individual CAF features were measured on all speech samples, spanning six sub-components: lexical sophistication, lexical appropriateness, grammatical complexity, grammatical accuracy, fluency and pronunciation. These linguistic features were then subjected to both univariate and multivariate statistical analyses, to identify distinguishing CAF features that can predict examinees' CEFR levels.

Complexity, accuracy and fluency (CAF) features of speaking performances on Aptis across different levels on the Common European Framework of Reference (CEFR)

WHAT WE FOUND:

The results of this project reveal distinguishing features in all three CAF components. However, Aptis speaking performances at different CEFR levels are characterized by different CAF components. Interestingly, while lower proficiency levels can be distinguished by more CAF features, no meaningful differences were observed between the B2 and C levels.

The majority of the CAF features show moderate to strong correlations with CEFR levels. Based on the patterns of statistical relationships amongst the CAF features, the features are further grouped into four CAF components: (1) lexico-grammatical knowledge, (2) automaticity in language use, (3) macro-level speech fluency, and (4) pronunciation. These components largely correspond to the scoring criteria for the Aptis speaking test and the CEFR descriptors for speaking ability. Except for pronunciation, all components showed moderately strong to strong correlations with CEFR levels. These findings suggest that overall, the rating criteria reflect the systematic differences across proficiency levels on the Aptis speaking test. They also indicate an alignment between key criteria assessed in Aptis and components of speaking ability on the CEFR.

In terms of systematic differences across CEFR levels, automaticity of lexico-grammar use distinguished four of the five CEFR levels (A1 to B2). Macro-level fluency distinguished three CEFR levels (A2 to B1). However, no meaningful differences were observed between B2 and C levels on the Aptis speaking test. There are several possible explanations for the lack of significant differences. First, it is possible that there is not a meaningful difference between B2- and C-level performances; however, such an argument cannot be simply made without exhausting all possibilities that can influence the results of the study. Second, the array of performance features examined in this study is not exhaustive. There might be other features that can characterize the difference between B2 and C level performances, such as the precision and register/style of language use (these features were included as distinguishing features in the descriptors for C1 and C2 levels on the CEFR). Third, the lack of meaningful differences between B2 and C levels might also be because the Aptis speaking tasks only target B1 and B2 levels, but not above. Therefore, the tasks might not be designed to capture the differences between B2 and C levels well.

Despite the lack of distinguishing power between B2 and C level, findings of this study provide strong support for the quality of the Aptis speaking test in two aspects: (1) the Aptis speaking test measures the intended speaking abilities, and (2) the Aptis speaking test is able to distinguish most levels of the CEFR.