

February 2021



Linking CAEL CE Scores to the Common European Framework of Languages

Introduction

In July of 2019, a study was conducted to link test taker performances on the Canadian Academic English Language Assessment, Computer Edition (CAEL CE) to the Common European Framework of Languages (CEFR). The study was focused on the CEFR levels relevant to academic studies – B1, B2, and C1. An internal group of assessment experts led 12 experienced panellists through the required steps of the standard-setting procedure. The methodological steps and the results of the study are presented below.

Methodology

To achieve a sufficiently large and representative group of experts (AERA/APA/NCME, 2014, p. 101), members of the following target professional groups were recruited: English for Academic Purposes (EAP) teachers, language evaluators, test content writers, language policy experts, and language assessment specialists. Table 1 below presents the composition of the panel.

Gender	Male	6
	Female	6
Role	Language instructor	5
	language evaluator	3
	Content writer	2
	Language policy expert	1
	Language assessment specialist	1

Before the panels, each participant performed three preparatory steps: (1) familiarizing themselves with the CAEL CE and CEFR design, (2) taking the CAEL CE test under authentic test conditions, and (3) completing CEFR training and calibration exercises.

A test form was selected from the operational item bank to align panellist judgments as closely as possible to the CAEL CE construct. After the Listening and Reading items were selected, examinee responses for the Speaking

and Writing tasks collected under operational conditions were selected. These responses were chosen based on how well they represented their respective scores (consistent ratings across operational raters) and how well these scores covered the possible range of scores (from low to high, exhausting the raw score scale).

Each day of the 4-day standard setting panel opened with a familiarization and training section. This introduction was followed by two rounds of judgements and ample opportunities for pair and group discussions. All four language components – listening, reading, speaking, and writing – utilized the Bookmark methodology (Cizek & Bunch, 2007). At the end of each day, feedback from the panellists, including their confidence about the judgements, was collected.

Analysis

After all participant judgements were collected, the final judgements were analyzed to recommend appropriate cut scores. First, it was determined that there were no excessively outlying judgements; this may be attributable to our usage of two rounds per component. The participants discussed their judgements and reconciled any misunderstandings or inconsistencies. As the bookmark method was used to establish quantitative links from the CEFR levels to the CAEL CE reporting scale, each participant's judgements produced a single value on the CAEL CE raw score scale for each of the components and CEFR levels. To arrive at one single cut score on the CAEL CE scale and CEFR level, the participants' individual judgements were averaged into one overall judgement. Note that the CAEL CE Overall score is a simple average, rounded to the nearest 10, of the four component scores.

Results

CEFR Level	CAEL CE Band
C1	80
B2	60
B1	40

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (AERA/APA/NCME). (2014). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Cizek, G. J., & Bunch, M. B. (2007). Standard setting: A guide to establishing and evaluating performance standards on tests. Thousand Oaks, CA: Sage.



Council of Europe (2009). Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). Strasbourg: Council of Europe. Accessed 08 December 2018. https://www.coe.int/en/web/common-european-frameworkreference-languages/tests-and-examinations