Language Teaching (2024), **57**, 616–619
doi:10.1017/S0261444824000223

CAMBRIDGE UNIVERSITY PRESS

RESEARCH IN PROGRESS

AMERICAN ASSOCIATION FOR APPLIED LINGUISTICS CONFERENCE, HOUSTON 2024
SELECTED POSTER PRESENTATIONS

Is this even B2? Challenges of selecting CEFR descriptors for young learner assessment linking studies[1]

David MacGregor, Mark Chapman, Fabiana MacMillan and Katie Schultz

University of Wisconsin-Madison, Madison, WI, USA
Corresponding author: David MacGregor; Email: macgregor3@wisc.edu

(Received 19 April 2024; accepted 13 June 2024)

1. Introduction

In this poster, we report on a project to link a language test designed for students in elementary school to the Common European Framework of Reference (CEFR) for Languages. Such a linking study presents challenges as CEFR descriptors are not always well suited to the language development of young learners nor to the assessments that are designed for them. While banks of supplementary descriptors, namely the collated representative samples of descriptors for young learners (Goodier & Szabo, 2018a, 2018b) ages 7–10 and 11–15, greatly help in linking young learner language assessments to the CEFR, many of these descriptors remain far removed from the language directly assessed on relevant language tests. We highlight the challenges of identifying appropriate CEFR descriptors for this purpose and propose methods for streamlining the descriptors presented to linking study panelists in advance of meetings.

WIDA MODEL Online (hereinafter MODEL) is an English language proficiency assessment for students in grades 1–12 that tests the four language domains of Listening, Reading, Speaking, and Writing. MODEL assesses the language that students need to access the content of the classroom with test items targeting the WIDA English Language Development Standards: Language of Language Arts, Language of Mathematics, Language of Science, Language of Social Studies, and Social Instructional Language. The test is structured by grade-level clusters with distinct test forms for students in grades 1–2, 3–5, 6–8, and 9–12.

MODEL can be administered at any time during the school year and is often used by international schools where English is the language of instruction to identify students who may be candidates for language support services, determine whether students meet admissions criteria for a school or program, gather immediate English language proficiency level scores for students, inform instructional planning, and monitor student progress with English language proficiency development over time. Outside of the United States, MODEL is used in over 500 English language medium schools in over 100 countries.

Also used internationally, the CEFR was designed to describe language proficiency across six main levels of proficiency (A1, A2, B1, B2, C1, and C2) in a wide variety of languages and contexts. Over time, plus levels (i.e., A2+, B1+, and B2+) and a new Pre-A1 level have also become more fully developed within the framework. Today, the CEFR comprises 1,832 descriptors at several levels of organization, in addition to the ten levels of proficiency. The CEFR defines five "Descriptor Schemes" (e.g., Communicative Language Activities, Signing Competencies), four modes of communication (e.g., Interaction, Reception), 16 Activities, Strategies, or Competences (e.g., Oral Comprehension,

© The Author(s), 2025. Published by Cambridge University Press

https://doi.org/10.1017/S0261444824000223 Published online by Cambridge University Press

Pragmatic Competence), and 93 Scales (e.g., Addressing Audiences, Flexibility, Prosodic Features); each of these categories is further subdivided.

When linking an assessment to the CEFR, it is the responsibility of the study designers to select a subset of those descriptors that are relevant to the construct of the assessment. Previous studies (Baron & Papageorgiou, 2014; Tannenbaum & Baron, 2015) have reported on linking assessments for young language learners to the CEFR using a curated sample of descriptors relevant to the construct of the assessments under consideration. However, the current study focuses on an assessment designed for multilingual learners as young as those in grades 1 and 2, which is a younger population of test-takers than that of previous CEFR linking studies. In addition, the assessment context of English language medium international schools, where this assessment is often administered, is different from assessment contexts previously studied in CEFR linking studies. In this context, students receive instruction in the content classroom in English, student populations are typically very diverse, and families are commonly of high socioeconomic status with high rates of literacy in both their home languages and in English.

In preparation for the linking study, we worked with experts in young learner language development and CEFR subject matter experts to create a curated list of appropriate CEFR descriptors in the four language domains assessed by MODEL. We started with the collated representative samples of descriptors for young learners ages 7–15, and conducted several rounds of pruning. The first step was to eliminate whole categories in CEFR that do not correspond to the construct of the assessment. For example, because MODEL is a test of English language proficiency, we eliminated all descriptors in the category "Plurilingual and pluricultural competence". Next, we eliminated all descriptors for "plus levels" (e.g., A2+) because we wanted panelists to focus on the main levels of proficiency. We then determined a ceiling that represented the highest level tested for each of the domains in each grade-level cluster (e.g., B2 for 3–5 Reading). Finally, we examined the remaining descriptors and eliminated any that we deemed irrelevant to the construct of MODEL (e.g., "Can exploit format, appearance, and typographic features to identify the type of text"; while this is an important skill for students to develop, it is not a feature of English language proficiency as defined by the WIDA ELD Standards). As shown in Table 1, after going through this process, we were left with 251 descriptors spanning Levels A1 to B1 (for Reading) or B2 (for all other domains) for the grade 1–2 cluster; 258 descriptors spanning Levels A1 to B2 for the grade 3–5 cluster; and 315 descriptors spanning Levels A1–C1 for the grades 6–8 and 9–12 clusters.

2. The linking study

2.1 Participants

For the linking study, we recruited 20 educators from international schools where MODEL is administered to serve as panelists. Each of the educators indicated familiarity with both MODEL and CEFR. The educators came from 15 countries across five continents. Each educator was assigned to one of four grade-level cluster panels, with a total of five educators per panel.

2.2 Pre-meeting

Before the meeting, panelists were asked to review both MODEL test content and the curated list of CEFR descriptors for their grade-level cluster. They were then asked to write short descriptions of what students in grade 2, 5, 8, or 10 (as appropriate for their cluster) could do at each CEFR level as covered by the curated list in the four domains.

2.3 Procedures used

The panelists met on the campus of Colegio Internacional SEK-El Castillo in Villafranca del Castillo, Spain, from January 27–29, 2023. Panelists were trained in the methods used for the linking study, which followed widely accepted standard-setting procedures (Cizek & Bunch, 2007; Council of

https://doi.org/10.1017/S0261444824000223 Published online by Cambridge University Press

Table 1. Final descriptors used in the study

Grade-level cluster	1–2	3–5	6–8, 9–12
CEFR level descriptors covered	Pre-A1–B2 (L, S, W) Pre-A1–B1 (R)	Pre-A1–B2	Pre-A1–C1
Number of descriptors used	251	258	315

Europe, 2009; Tannenbaum & Baron, 2015). This included a Bookmarking method for the Listening, Reading, and Speaking domains, and a Body of Work method for Writing. For the Bookmarking method, panelists examined test items ordered by difficulty and, for Listening and Reading, were asked to consider where a minimally qualified student at CEFR level (e.g., A2) would fall below a 50% chance of answering an item correctly; for Speaking they were asked to consider what level of proficiency would be required to meet the linguistic demands of the item. For the Body of Work method, they examined student responses to the writing tasks and were asked to consider the following questions:

- What CEFR level does the writing best represent?
- How confident are you (50%–100%)?
- If not 100% confident, what other adjacent level does the writing represent (higher or lower)?

Detailed descriptions of the methods used and the results obtained can be found in MacGregor et al. (2023).

3. Results

The results from the linking study showed that panelists were able to link MODEL scale scores up to CEFR level B2 for all domains in the 1–2 and 3–5 clusters (with the exception of 1–2 Reading, which reached a ceiling at B1) and up to level C1 for all domains for the 6–8/9–12 clusters.

4. Discussion

The current study shows that the CEFR is flexible enough that it can be applied to assessments for young learners, even if the assessment was not originally designed to measure proficiency according to CEFR levels, as long as the designers of the linking study can identify a sufficient set of descriptors from the framework that fit the construct of the assessment. When linking an assessment for young language learners to the CEFR, it is also worth noting that the banks of supplementary descriptors for young learners can also serve as a helpful resource and starting point.

It should be emphasized that because the CEFR encompasses such a broad range of proficiency levels and language use contexts, no assessment can or should try to assess the full range. In the context of the current study, for example, the fact that the highest level to which we were able to link scores on the grades 9–12 test was C1 should not be taken to imply that learners in that age range cannot reach level C2; rather, it is a limitation of the assessment that it cannot provide evidence for proficiency at the C2 level. Similarly, as mentioned above, the fact that MODEL does not assess "Plurilingual and pluricultural competence" should not be taken to mean that WIDA does not value such competence, but rather that such competence is not directly relevant to the test construct.

Understanding MODEL test scores in terms of CEFR levels provides educators with information about their students' language development that is internationally meaningful. This information is valued by students, their families, and other stakeholders.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S0261444824000223.

https://doi.org/10.1017/S0261444824000223 Published online by Cambridge University Press

Note

1 A reproduction of the poster discussed is available in the supplementary material published alongside this article on Cambridge Core.

References

Baron, P., & Papageorgiou, S. (2014). *Mapping the TOEFL primary test onto the common European framework of reference*. (Research Memorandum ETS RM-14-05). Educational Testing Service. https://www.ets.org/Media/Research/pdf/RM-14-05.pdf

Cizek, G., & Bunch, M. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Sage Publishing.

Council of Europe. (2009). *Relating language examinations to the common European framework of reference for languages: Learning, teaching, assessment. Strasbourg*. Council of Europe.

Goodier, T., & Szabo, T. (2018a). *Collated representative samples of descriptors of language competences developed for young learners: Resource for educators* (Vol. *1)*, Ages 7–10, 2018 Edition. Council of Europe. https://rm.coe.int/16808b1688

Goodier, T., & Szabo, T. (2018b). *Collated representative samples of descriptors of language competences developed for young learners: Resource for educators* (Vol. *2*) Ages 11–15, 2018 Edition. Council of Europe. https://rm.coe.int/collated-representative-samples-descriptors-young-learners-volume-2-ag/16808b1689

MacGregor, D., Chapman, M., & Cook, H. G. (2023). *Linking scores from WIDA MODEL online to the common European framework of reference (CEFR)*. (Report No. RR-2023-2). Wisconsin Center for Education Research, University of Wisconsin–Madison. https://wida.wisc.edu/sites/default/files/resource/Linking-Scores-from-WIDA-MODEL-Online-to-CEFR.pdf

Tannenbaum, R. J., & Baron, P. A. (2015). *Mapping scores from the TOEFL junior comprehensive test onto the common European framework of reference (CEFR)*. (Research Memorandum ETS RM-15-13). Educational Testing Service. https://www.ets.org/Media/Research/pdf/RM-15-13.pdf.

Cite this article: MacGregor, D., Chapman, M., MacMillan, F., & Schultz, K. (2024). Is this even B2? Challenges of selecting CEFR descriptors for young learner assessment linking studies. *Language Teaching*, *57*(4), 616–619. https://doi.org/10.1017/S0261444824000223

https://doi.org/10.1017/S0261444824000223 Published online by Cambridge University Press