METAMETRICS RESEARCH BRIEF

Aligning the *Lexile*® Framework for Reading to the Common European Framework of Reference

Jing Wei, Director of New Product Innovation; Alistair Van Moere, Chief Product Officer

AUGUST 2021

OBJECTIVE

The Common European Framework of Reference (CEFR) is the most widely-used framework to describe students' language learning progress on tests and in classrooms. International language tests align student test scores to CEFR levels so that students' performances can be interpreted with CEFR can-do descriptors, therefore making test scores more meaningful (Green, 2018).

The Lexile Framework for Reading utilizes the psychometric concept of conjoint measurement and places both reader ability measures and text complexity measures on the same scale. The Lexile scale was calibrated in such a way that when the text complexity is controlled within a certain range of the reader ability measure, the reader will find the text within their zone of optimal development, that is, "challenging but not frustrating" to read. Through this approach, the Lexile Framework for Reading has brought meaning to test scores and made the scores actionable.

Because of this unique feature, over 23 states within the U.S. and many more testing institutions internationally have linked their test scores to the Lexile Framework and added student Lexile® reading measures to their score reports. Numerous publishers also license the Lexile® Text Analyzer (i.e., an automated text leveling software) to measure their books and use Lexile® text measures to guide readers to select books at the right level. As CEFR and the Lexile Framework share the same goal of bringing meaning to measures on either the text or the reader side, it is useful to understand the relationship between these two frameworks.

METHODS

Dataset

This study drew on data that was aggregated from a number of linking studies that MetaMetrics conducted in collaboration with leading testing organizations. A total of nine internationally reputable English language tests were selected for this study, each of which have robust links to the Lexile Framework. Linking studies are the standard practice for connecting a test's score scale to the Lexile scale. MetaMetrics conducted separate linking studies in collaboration with each of these nine tests by following procedures as recommended by the research literature (Kolen & Brennan, 2014). This enabled each test to align their score scale with the Lexile scale and report Lexile measures from their scores.

Each of the nine English language tests' publishers had previously conducted their own standard-setting studies in order to align their test scores to CEFR. As a result, the linking studies show (i) the correspondence between Lexile measures and each tests' scale scores, and (ii) the correspondence between Lexile measures and CEFR levels, as estimated by each tests' own CEFR benchmark study (see Figure 1).

Figure 1. Illustration of the link between the Lexile Framework and nine English language tests, and by extension, to the CEFR.

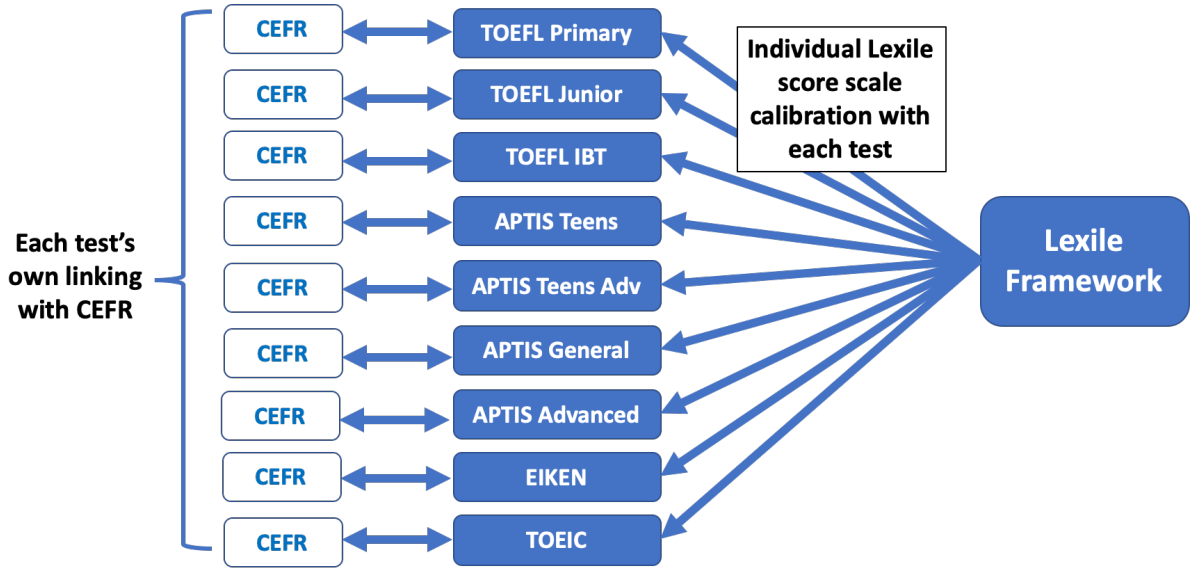


Table 1 shows the number of students in each of the nine Lexile linking studies. Also shown in the table is the number of students within each CEFR band, as determined by each tests' CEFR benchmark study. The total number of students involved is 34,477. Some cells contain zeros

because not every test spans all CEFR bands; for example, TOEFL Primary (a test for young learners) has not been benchmarked to the CEFR above B1, while TOEFL iBT (a test for university entrance) is rarely attempted by students with below B1 proficiency.

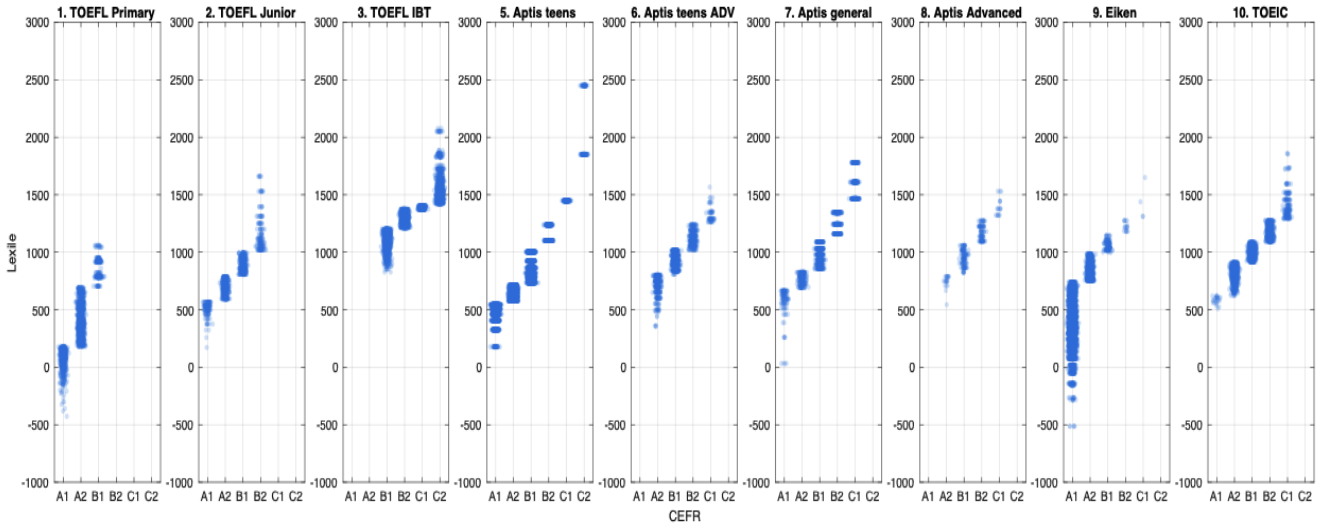
Table 1. The number of students per CEFR band, as classified by each of the nine tests.

Test	*Number of students within each CEFR band*						*Total*
	A1-/A1	*A2*	*B1*	*B2*	*C1*	*C2*	
TOEFL Primary	578	1,388	254	0	0	0	2,220
TOEFL Junior	145	448	403	175	0	0	1,171
TOEFL iBT	0	0	1,676	2,144	434	1,558	5,812
Aptis Teens	2,672	3,874	3,223	543	155	131	10,598
Aptis Teens Adv	0	383	707	414	67	0	1,571
Aptis General	115	260	791	531	338	0*	2,035
Aptis Advanced	0	22	134	97	18	0*	271
Eiken	6,881	989	137	18	4	0*	8,029
TOEIC	26	889	1,043	626	186	0*	2,770
Total	10,417	8,253	8,368	4,548	1,202	1,689	34,477

**Note: TOEIC, EIKEN, Aptis General and Aptis Advanced do not report C2 separately but rather report "C1 and above".*

Figure 2 shows the distribution of student Lexile measures at each CEFR level from each test. The blue dots show the distribution of students on the Lexile scale, as categorized by CEFR bands. Inspection of the figure reveals some variations from test to test – students categorized as A1 from the EIKEN exam are more similar in proficiency to those categorized as A2 by TOEFL Primary, for example. This is not entirely surprising. Although there are industry-recommended guidelines for conducting standard setting studies, in practice two tests could follow somewhat different methods and use different panelists. Combined with qualitatively different test tasks that are designed for different ages of test-takers, this explains the lack of consistency in CEFR mappings from test to test.

Figure 2. Scatterplots showing the distribution of students (blue dots) for each of the nine tests.



Note: The x-axis shows the CEFR placement of students according to each tests' CEFR benchmarking study. The y-axis shows the students' Lexile measures, according to the link between each test and the Lexile scale.

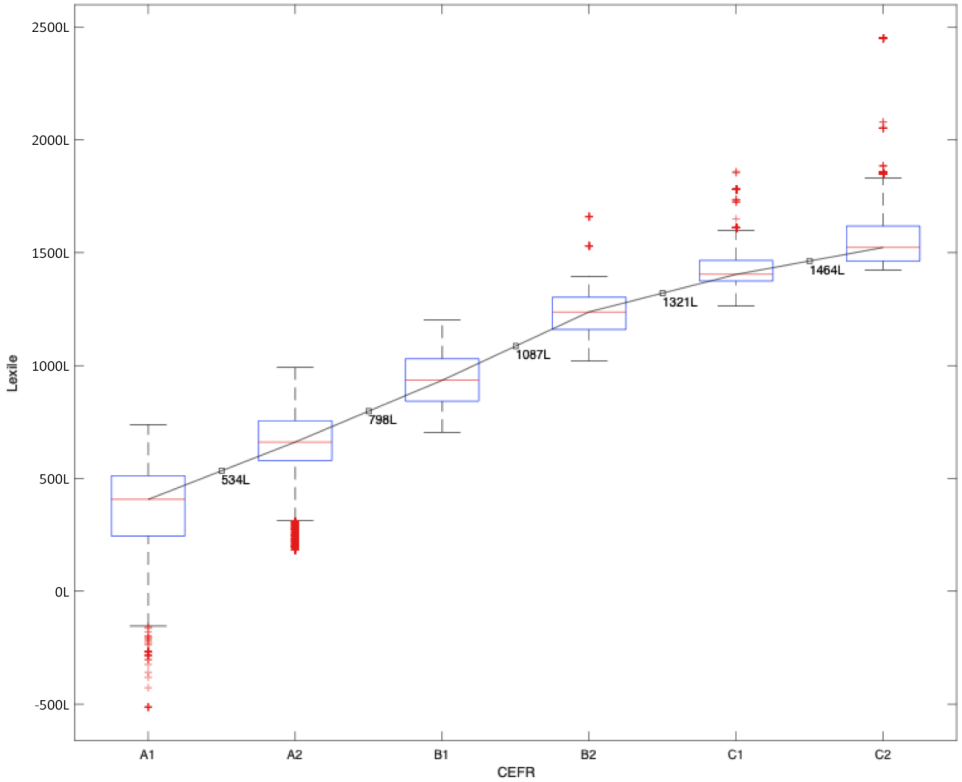
ANALYSES

In order to resolve these inconsistencies of CEFR classifications across different tests, the entire set of 34,477 students' records were aggregated and listed out in two columns: Lexile measure and CEFR band. It was hypothesized that the variations in CEFR cut scores due to different tests would be "washed out" once they were combined into a single dataset.

Figure 3 shows the box-whisker plot for Lexile measures: the red horizontal line shows the median Lexile measure at each CEFR band, the horizontal lines forming the top and bottom of each box show the interquartile range (therefore, 50% of all students at that CEFR band are within the box), and the upper and lower horizontal "whiskers" show the maximum and minimum Lexile measures for students within each CEFR band. The red dots below and above the whiskers are individual students that are considered outliers.

The Lexile cut scores are shown as points in between each box. They were established by taking the midpoint between the two median scores of each of the adjacent CEFR levels. When taking this aggregated approach, it can be seen that students are clearly separated into CEFR bands as expressed by Lexile measures, regardless of which test they took.

Figure 3: Box-whisker plot showing the distribution of students on the Lexile scale and their CEFR classification.



Note: This plot shows the distribution of students on the Lexile scale and their CEFR classification according to the nine tests, using the aggregate dataset. The boundary between CEFR bands on the Lexile scale are represented by points between boxes.

RESULTS & DISCUSSION

The cut scores were evaluated using a classification consistency approach. We asked the question: what proportion of students that were classified as A1 by the test they took were also classified as A1 using the assigned Lexile ranges in Figure 2? The same question was asked for students classified as A2, B1, B2, C1 and C2.

Overall, 72% of students were consistently classified; that is, 72% of students were classified as belonging to the same CEFR band by both the test they took and by the Lexile mapping. Conversely, 28% of students were classified as being in adjacent CEFR bands, for example, they were assigned as A2 by the test's CEFR mapping and as B1 by the Lexile CEFR mapping. Classification disparities into CEFR levels more than 1 band apart were practicality non-existent except for a few students that were categorized as B2 by their test and as C2 by the Lexile mapping.

Table 2 shows the CEFR band-by-band classification consistency. Each band shows high classification consistency except for C1, at 60%. However, there are two reasons why this percentage may not represent the full picture. First, several of the tests - namely TOEIC, EIKEN, Aptis General and Aptis Advanced – do not report C2 separately but rather report "C1 and above". Therefore, many students classified as C1 may actually be C2 students. Also, the Lexile C1 mapping draws heavily on TOEFL iBT test-takers, which, as shown in Table 1 and Figure 1, is similarly represented by a comparatively small number of students (n=434) that fall into a narrow score range between B1 (n=2,144) and C2 (n=1,558). For these reasons, the Lexile range for C1 is consistent with the available data, but the 60% classification consistency is artificially low and the true consistency is likely be higher.

Table 2. The number and percent of students that were consistently classified into CEFR bands by both the test they took and by the Lexile-CEFR benchmarking.

CEFR	*Lexile measure*	*Number of students correctly classified by the Lexile range*	*Number of students incorrectly classified by the Lexile range*	*% correctly classified*
A1	< 535L	7,957	2,460	76%
A2	540L – 800L	5,794	2,459	70%
B1	805L – 1090L	5,802	2,566	69%
B2	1095L – 1320L	3,272	1,276	72%
C1	1325L – 1460L	722	480*	60%*
C2	> 1465L	1,205	484	71%

**Many of these students were classified by their tests as "C1 and above", meaning that the true classification consistency of the C1 band is higher than shown here.*

Concordance Table

Using the cut measures from the box-and-whisker plot, a concordance table was created between students' Lexile measures and the corresponding CEFR levels. For example, if a student's Lexile measure is 550L, she will be classified as an A2 student.

Table 3: Concordance table for student Lexile measures and CEFR.

CEFR	*Student and Text Lexile range*
A1	<535L
A2	540L - 800L
B1	805L - 1090L
B2	1095L - 1320L
C1	1325L - 1460L
C2	>1465L

Note: Lexile measures were rounded to the nearest 5L by following MetaMetrics rounding conventions.

Recommended Uses and Next Steps

This study represents a first attempt to establish the alignment between the Lexile Framework for Reading and CEFR. Such an alignment will make test scores more interpretable and actionable. After receiving a Lexile score on a score report, educators, students and parents will be able to select the most appropriate reading materials for their students. Lexile measures also add a numerical layer to CEFR can-do descriptors and make the descriptors more objective and actionable. For example, instead of knowing that an A2 student "can understand the main idea of newspaper articles with the support of visual aids", a teacher will be able to say: "That student can understand the main idea of newspaper articles falling between 500L to 800L. To facilitate optimal reading growth, she will need to read materials ranging from 400L to 850L."

For the next step, a standard-setting study that involves panel judgment will be conducted. Results from the statistical aggregation and the panel judgment approaches will be compared, which will provide additional validity evidence to the cut scores from this round.

REFERENCES

Green, A. (2018). Linking Tests of English for Academic Purposes to the CEFR: The Score User's Perspective, *Language Assessment Quarterly, 15*:1, 59-74, DOI:10.1080/15434303.2017.1350685

Kolen, M., & Brennan, R. (2014). Testing Equating, Scaling and Linking: Methods and Practices. Springer: Iowa City