# A Case Study of the CEFR Alignment in Language Assessment: Insights from EFL Instructors In Turkey

İsmail Fırat Altay[1]*, Nurgül Bekdemir[2]

[1]Hacettepe University, Faculty of Education, Ankara-Turkey
[2]Hacettepe University, Graduate School of Educational Sciences, Ankara-Turkey

## ABSTRACT

This study investigates the instructors' perspectives on the alignment of intermediate-level language tests (B1 & B2) applied in the 2023-2024 academic year at a preparatory school in Turkey. A qualitative case study was adopted in the study and qualitative data were collected from 6 lecturers in the testing office of the preparatory school through reflection papers and semi-structured interviews. The purposive sampling method was conducted in the study, in which all the instructors in the testing office participated. Moreover, all the tests throughout the academic year were purposefully included in the study, as sets of B1 and B2 levels would lead to more comprehensive results. Content analysis was conducted for data analysis. The results demonstrated diversity in the alignment rates of the B1 and B2 level tests with the help of factors that contributed to the alignment and detracted from the alignment with the target levels. The study is of importance with the suggestions of to enhance the tests at intermediate levels and provides practical implications for all the stakeholders in language assessment depending on the CEFR.

**How to cite this article:** Altay İF, Bekdemir N (2025).  A Case Study of the CEFR Alignment in Language Assessment: Insights from EFL Instructors In Turkey. International Online Journal of Education and Teaching, Vol. 12, No. 3, 2025, 56-74

## INTRODUCTION

Language assessment is a critical issue to assess the learning outcomes and to lead teaching practices in language teaching and learning procedures. The Common European Framework of Reference for Languages (henceforth CEFR), which has been developed in order to describe and standardize the language competences internationally, enables language assessment in an objective and comparable way. The CEFR aims to provide consistency in curriculum, materials, and assessment in the proficiency levels from A1 level to C2 level (CoE, 2020).

However, it is a complicated process to prepare target-level tests of English for language teachers as well as to determine the CEFR levels to which language tests correspond. Several factors, such as target competences, their objectives, and the tasks adopted for these purposes, have utmost importance to align the level of the adopted tests with the CEFR. In this context, instructors' perception of the level

of tests is critically important for the reliability and validity of the language tests in addition to the factors that had an effect on the alignment. However, the diversity in the instructors' perception could influence the actual level of the tests and even the academic achievement of the learners.

The current study aims to explore the perception of instructors on the alignment of the applied language tests aiming to test B1 and B2 levels based on the CEFR and the effective factors on the alignment. Thus, it is aimed to enhance the test development regarding the CEFR with in-depth understanding of the instructors who are responsible in the testing office. It also contributes to the language assessment practices being more reliable and transparent through the instructors' experiences in the test development procedure and the test design based on the institutional standards (CoE, 2020).

## LITERATURE REVIEW

Since it was published in 2001 (CoE, 2001), the CEFR has been guiding language education worldwide (Alderson, 2007; Byram & Parmenter, 2012; Hulstijn, 2007; Khalifa & French, 2008; North, 2014). Over approximately two decades, its effect on language education, especially in assessment, has dramatically increased. Thanks to the 20th session of the Standing Conference of the Ministers of Education of the CoE in Cracow, Poland, CEFR descriptors were decided to be used commonly to determine the proficiency level of the language learners (CoE, 2000). Common reference levels were mainly divided into three: (1) A (basic user), (2) B (independent user), and (3) C (proficient user). However, the development of

the language proficiency could be clearly seen in the figure below, as each level is comprised of the higher one (Figure 1) (CoE, 2020). Moreover, the plus levels were included in three levels: A2, B1, and B2. For instance, these are adopted when learners could not reach the requirements of the higher level even if they are highly competent at a level (CoE, 2001).

As for the purposes of such an important guide in language education, the main one is to "promote as well as to strengthen cooperation among different educational institutions in countries by providing a common reference standard for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc." (CoE, 2001, p.1). Furthermore, a description of proficiency levels for the standardization of the language tests to facilitate various systems of qualifications is another purpose that is supported by being comprehensive, transparent, and coherent (CoE, 2001). The complex nature of the language proficiency is clarified based on the learning objectives and outcomes for each level and includes a wide range of language competences, activities, and strategies indicating the comprehensiveness (CoE, 2018; 2020). The usage of clear and understandable language in the CEFR is of importance to its users, as its worldwide adoption supports its transparency function. Also, the absence of contradictory ideas throughout the whole document both enables reliability and shows its coherence.

### CEFR-Related Studies

The diversity in the knowledge and awareness of the instructors on the CEFR is of crucial significance in
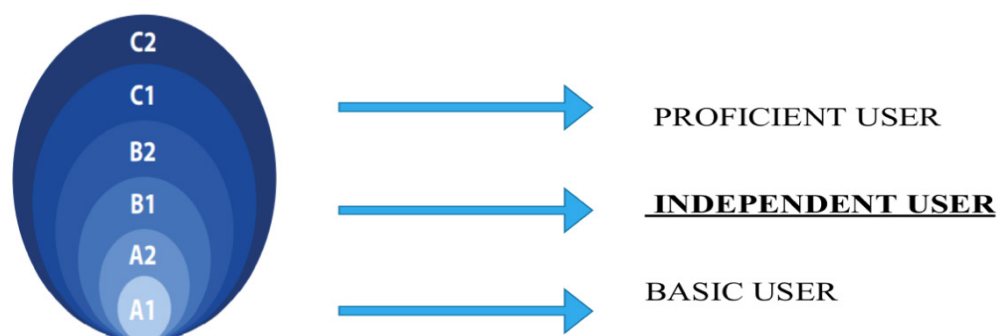


**Fig. 1: CEFR Common Reference Levels (CoE, 2020, p.36)**

its applications in the language teaching process. The research in Turkish context reveals the limited teachers' awareness regarding the CEFR and inadequate CEFR knowledge of teachers, especially in state institutions (Çağatay & Ünveren-Gürocak, 2016; Yılmaz-Yakışık & Ünveren-Gürocak, 2018). Despite the low level of knowledge, Ünlücan-Tosun and Glover (2020) proposed that they were highly motivated to adopt the CEFR in their teaching practices.

## The Effect of the CEFR on Language Assessment and Challenges in Implementation

The effect of the CEFR on language assessment and the challenges in its implementation have been revealed in different contexts. For more transparent and consistent assessment based on the CEFR, adopting clear objectives indicates the significant role of the CEFR in assessment (Mison & Jang, 2011). CEFR learning outcomes transform teachers' assessment practices; however, self-assessment and peer assessment are adopted more (Le, 2018). Similarly, CEFR-based language assessment is found to be pedagogically supportive and to contribute to the learning process of the learners (Sabbir, 2019). Despite these positive contributions, the challenges faced in the practice hinder the complete application of the CEFR. Moonen et al. (2013) describe the inconsistency between the teaching and assessment practices to be a problem and Baldwin and Apelgren (2018) emphasize that the actual practice of the CEFR descriptors in practice is limited and creates challenges during the assessment process. Le (2018) proposes that ignorance of the competence-based approach of the CEFR lessens its real potential while Sabbir (2019) reports the lack of appropriate material and technological sources in listening skills even if CEFR-based assessment is perceived positively. Studies in the Turkish context indicate the low level of awareness of CEFR-based language assessment, and pedagogical and institutional constructive support is required for the teachers who need in-service training (Çağatay & Ünveren-Gürocak, 2016; Kavaklı & Mirici, 2019; Ünlücan-Tosun & Glover, 2020). It could be seen that the CEFR has a potential to contribute to language assessment practices; however, teachers ought to have training on the CEFR and its applications in order to have an effective contribution. Also, this study is of importance to investigate the instructors' perspectives on the adoption and alignment of the CEFR in the language tests.

Regarding the research gap in the evaluation of language tests in relation to the CEFR in Turkey, this research addresses the following research questions:

1. What are the views of the instructors about the compatibility between the learning outcomes of the target levels according to the CEFR criteria and the applied exams?
   a. B1 Level Tests
      i. Listening Skill
      ii. Reading Skill
      iii. Language Use
      iv. Writing Skill
      v. Speaking Skill
   b. B2 Level Tests
      i. Listening Skill
      ii. Reading Skill
      iii. Language Use
      iv. Writing Skill
      v. Speaking Skill
2. What are the effective factors influencing the alignment of each skill at B1 and B2 levels?

## METHOD

### Research Design

A qualitative case study design, which is conducted to collect data to discover more about the research problem (Dörnyei, 2007), was adopted to explore the instructors' perspective on the alignment of the language tests designed at B1 and B2 levels with the CEFR. Through case studies, an in-depth understanding of a program, event, process, which are called a case, is provided (Creswell & Creswell, 2018). Thus, the case of the current study was a set of language tests at B1 and B2 levels which represented the institution's testing system, which indicates the collective case study as the method adopted in order to investigate the general condition (Stake, 2000).

### Instruments

Data collection was conducted through three instruments. First, an online form was utilized to collect

demographic information of the participants that could be correlated with the research results. It included close-ended questions about their educational background, experience in teaching languages and assessment. Also, during the 2023-2024 academic year, a total of 9 language tests were applied at the preparatory school, four of which were designed for B1-level competences and 5 of which were designed for B2-level testing. Instructors wrote a reflection paper for each test to demonstrate their opinions freely on the level of the skills and the factors contributing to or detracting from the alignment with the target level. After writing the reflection paper, semi-structured interviews with open-ended questions were administered online for each test to deeply explore their perspectives toward the rates of alignment with B1 level and the reasons for their opinions without any limitations (Creswell, 2015). For the interview questions, three faculty members, who were experienced and expert in the field, were consulted for their expert opinion in order to eliminate the misunderstandings during the interviews and improve the quality of questions.

### Participants

Instructors in the testing office of the SFL at a state university participated in the study. The purposive sampling method was adopted in order to achieve a rigorous focus on the phenomenon (Creswell & Creswell, 2018; Schoch, 2020). For the purpose of this research, the instructors who were responsible for the preparation of B1-level and B2-level tests in the testing office during the 2023-2024 academic year were selected, as they were knowledgeable about the testing system of the institution and could provide information about the applied tests. The number of the instructors was 6, all of whom participated in the study. The participants' educational background is given in Table 1. For BA degree, 4 participants (66.7%) graduated from the department of ELT whereas the number of participants who graduated from the department of ELL (n=1, 16.7%) and ACL (n=1, 16.7%) was equal. However, participants got MA degree from ELT (n=3, 60%) and ELL (n=2, 40%).

The other factor that could be effective was their experience in teaching and language testing (Table 2).

The ranges were 2-5 years, 6-10 years, and more than 10 years. There was an equality among the ranges of the experience in teaching, as there were 2 participants (33.3%) in each range. However, there was a variety in participants' experience in testing. 3 participants (50%) had 6-10 years of experience while 2 participants (33.3%) had 2-5 years of experience, and 1 participant (16.7%) had more than 10 years of experience in testing.

#### Table 1: Graduation Departments of the Participants

| Department of BA | n | % | Department of MA | n | % |
|---|---|---|---|---|---|
| English Language and Teaching | 4 | 66.7 | English Language and Teaching | 3 | 60 |
| English Language and Literature | 1 | 16.7 | English Language and Literature | 2 | 40 |
| American Culture and Literature | 1 | 16.7 | | | |
| Total | 6 | 100 | | 5 | 100 |

#### Table 2: Years of Experience in Teaching and Testing

| | Years of Experience in Teaching | | Years of Experience in Testing | |
|---|---|---|---|---|
| | n | % | n | % |
| 2-5 | 2 | 33.3 | 2 | 33.3 |
| 6-10 | 2 | 33.3 | 3 | 50 |
| More than 10 | 2 | 33.3 | 1 | 16.7 |
| Total | 6 | 100 | 6 | 100 |

## Data Collection Procedure

Data were collected throughout the 2023-2024 academic year and the data collection procedure was conducted in many phases. Firstly, participants completed the online form for demographic information. Then, participants wrote a reflection paper, and a semi-structured interview was conducted with them after each test. This process was applied nine times, as the set of tests included 9 tests. All the data from the interviews were transcribed for the data analysis.

## Data Analysis

The collected data were analyzed using qualitative approaches to address the research questions. Content analysis was conducted in order to explore the participants' perspectives towards the alignment rates of the tests and the factors influencing the alignment (Dörnyei, 2007). MaxQDA 2020 was utilized to analyze the qualitative data gathered through reflection papers and semi-structured interviews.

## Results

This section demonstrates the research findings emerged from the data collection procedure. First, the alignment rates of tests with target levels will be presented based on the perspectives of experts. Second, the perspectives of the participants will be elaborated with the factors that contributed to the alignment and distanced from the alignment with the target levels.

## Instructors' Perspectives towards the Alignment Rates

The perception of instructors on the alignment of target B1 level skills were scrutinized through two data collection instruments and Table 3 demonstrates the rates of alignment with B1 level based on instructors' perception.

First, listening skill was aligned with B1 level at a satisfactory level (f=29, 64.4%), but it was also revealed with the results of the reflection papers that listening skill was also found to be at a higher level than B1 level (f=16, 35.6%). These results were also supported by the interview results with slight changes, indicating B1 level (f=27, 57.4%) and

higher levels (f=20, 42.5%). Second, the results of reflection papers and interviews on the alignment of reading skill overlapped with a satisfactory rate of alignment with B1 level (f=33, 63.5%) and a low level of alignment with the levels beyond B1 (f=19, 63.5%). Next, language use, which comprises grammar and vocabulary, was moderately aligned with B1 level as the findings of both reflection papers (f=59, 58.4%) and interviews (f=59, 55.6%) indicated. Also, language use was reported to be beyond B1 level as a result of the reflection papers (f=42, 41.6%) and interviews (f=46, 43.5%). Besides, the interview results indicated that it was below B1 level with a small percentage (f=1, 0.9%). Then, writing skill was also aligned with B1 level at a moderate level based on the results of reflection papers (f=7, 58.3%) and interviews (f=8, 53.3%). However, a moderate rate of alignment was also reported for the levels beyond B1 level in the reflection papers (f=5, 41.7%) and interviews (f=7, 46.7%). Lastly, the reflection papers (f=8, 66.7%) and interviews (f=10, 62.5%) supported each other in terms of satisfactory rate of alignment of speaking skill with B1 level. It was also indicated in reflection paper (f=4, 33.3%) and interviews (f=6, 37.5%) that speaking skill was found to be beyond B1 level at a low rate.

The instructors' perception on the alignment rates of the skills with B2 level were shown in Table 4. The results demonstrated a high rate of alignment with B2 in receptive skills i.e., listening and reading skills, and language use. Satisfactory rates of alignment with the target level emerged in productive skills i.e. writing and speaking skills.

To start with listening skill, it was revealed that the alignment rate with B2 level was high as a result of reflection paper (f=39, 76.5%) and interviews (f=38, 76%). However, the results of reflection papers indicated several proficiency levels at a low rate of alignment: B1 level (f=5, 9.8%) and C1 level (f=6, 11.8%) in addition to the report that listening skill did not align with B2 level without reference to any proficiency levels (f=1, 1.9%). Also, interview results indicated similar results to the reflection papers i.e., B1 level (f=7, 14%), C1 level (f=3, 6%) and not at B2 level with no reference to any proficiency levels (f=2, 4%). As for reading skill, the results of both reflection

**Table 3: Perception of Instructors on B1-level tests**

| Skills | Reflection Papers | | | | | | Interviews | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | B1 | | > B1 | | Total | | B1 | | > B1 | | <B1 | | Total | |
| | f | % | f | % | f | % | f | % | f | % | f | % | f | % |
| Listening | 29 | 64.4 | 16 | 35.6 | 45 | 100 | 27 | 57.4 | 20 | 42.5 | | | 47 | 100 |
| Reading | 33 | 63.5 | 19 | 36.5 | 52 | 100 | 33 | 63.5 | 19 | 36.5 | | | 52 | 100 |
| LU | 59 | 58.4 | 42 | 41.6 | 101 | 100 | 59 | 55.6 | 46 | 43.5 | 1 | 0.9 | 106 | 100 |
| Writing | 7 | 58.3 | 5 | 41.7 | 12 | 100 | 8 | 53.3 | 7 | 46.7 | | | 15 | 100 |
| Speaking | 8 | 66.7 | 4 | 33.3 | 12 | 100 | 10 | 62.5 | 6 | 37.5 | | | 16 | 100 |

**Table 4. Perception of Instructors on B2-level tests**

| Instrument | Levels | f / % | Listening | Reading | LU | Writing | Speaking |
|---|---|---|---|---|---|---|---|
| Reflection Papers | B1 | f | 5 | 6 | 1 | 3 | 1 |
| | | % | 9.8 | 13.6 | 1.4 | 15 | 9 |
| | B2 | f | 39 | 32 | 62 | 14 | 7 |
| | | % | 76.5 | 72.7 | 89.9 | 70 | 63.7 |
| | B2X | f | 1 | - | 6 | 1 | 3 |
| | | % | 1.9 | - | 8.7 | 5 | 27.3 |
| | B2+ | f | - | 1 | - | - | - |
| | | % | - | 2.3 | - | - | - |
| | C1 | f | 6 | 5 | - | 2 | - |
| | | % | 11.8 | 11.3 | - | 10 | - |
| | TOTAL | f | 51 | 44 | 69 | 20 | 11 |
| | | % | 100 | 100 | 100 | 100 | 100 |
| Interviews | B1 | f | 7 | 3 | 2 | 5 | - |
| | | % | 14 | 7.7 | 2.8 | 22.7 | - |
| | B2 | f | 38 | 30 | 60 | 16 | 6 |
| | | % | 76 | 76.9 | 84.6 | 72.7 | 46.2 |
| | B2X | f | 2 | 1 | 4 | 1 | 5 |
| | | % | 4 | 2.6 | 5.6 | 4.5 | 38.5 |
| | C1 | f | 3 | 5 | 5 | - | 2 |
| | | % | 6 | 12.8 | 7 | - | 15.3 |
| | Total | f | 50 | 39 | 71 | 22 | 13 |
| | | % | 100 | 100 | 100 | 100 | 100 |

papers (f=32, 72.7%) and interviews (f=30, 76.9%) supported each other on the alignment with B2 level. However, alignment with different proficiency levels was also revealed. The reflection papers resulted in with low rate of alignment with B1 level (f=6, 13.6%), B2+ level (f=1, 2.3%) and C1 level (f=5, 11.3%) whereas the interviews demonstrated similar results to the alignment with B1 level (f=3, 7.7%) and C1 level (f=5, 12.8%) in addition to the report of no alignment with B2 level (f=1, 2.6%). Next, the alignment of

language use with B2 level was high as a result of both reflection papers (f=62, 89.9%) and interviews (f=60, 84.6%). However, instructors reported that language use was aligned with B1 level at a low rate in their reflections (f=1, 1.4%) and interviews (f=2, 2.8%). They also reported that it was not aligned with B2 level in both their reflections (f=6, 8.7%) and interviews (f=4, 5.6%) without reference to any other levels. The alignment with C1 level (f=5, 7%) was also revealed in the interviews. In terms of writing skill, the results demonstrated alignment with different proficiency levels according to instructors; however, both reflection papers (f=14, 70%) and interviews (f=16, 72.7%) yielded similar results with a satisfactory rate of alignment with B2 level. The results of reflection papers indicated the alignment with several proficiency levels such as B1 level (f=3, 15%), C1 level (f=2, 10%) and no alignment with B2 level (f=1, 5%). Similarly, the interview results indicated low rate of alignment with B1 level (f=5, 22.7%) and not alignment with B2 level (f=1, 4.5%). Lastly, the results demonstrated differences in the alignment rates of speaking skill. While reflection papers indicated a satisfactory level of alignment with B2 level (f=7, 63.7%), interviews demonstrated a moderate level of alignment (f=6, 46.2%). Also, the alignment with B1 level (f=1, 9%) and no alignment with B2 level (f=3, 27.3%) as the results of reflection papers indicated. The interviews supported no alignment with B2 level without referring to any other levels (f=5, 38.5%) and revealed a low level of alignment with C1 level (f=2, 15.3%).

## Instructors' Perspectives towards the Factors Influencing Alignment with Target Levels

Moreover, the factors that contributed to the alignment and detracted from the alignment with target levels were also revealed for each skill and language use. The factors for the alignment of listening skill with B1 level are shown in Table 5.

The contributing factors to the alignment with B1 level (f=160) outweighed the detracting factors from B1 level (f=74). Firstly, the factors that affected B1-level listening skill were discourse, question design and intelligibility as contributing and detracting

factors. Discourse was a contributing factor (f=66, 41.25%) which included several elements, such as familiar themes of the listening text with the learners, level-appropriate language used in the text, semantically organized text, lexical appropriateness, inclusion of interculturality in the listening texts, having conversational turn structure and various genres. However, it was also revealed to be a detracting factor from the alignment with B1 level (f=45, 60.8%). Its elements emerged as unfamiliar theme of the listening with the learners, higher level of language used in the text, semantically disorganized text, lexical inappropriateness, lack of interculturality, usage of same scenarios continuously, and problematic length of the listening text. Question design was the second factor that emerged as both contributing (f=66, 41.25%) and detracting ones (f=24, 32.4%). As for the contributing factor to the alignment with B1 level, there are four elements leading to the alignment e.g., appropriate objective of questioning, mode of questions, level-appropriate question roots and options, and instructional scaffolding. However, the first three elements were also reported to be detracting ones. Last factor in the alignment of B1-level listening skill was intelligibility. Comprehensible accent of speakers and their appropriate pace led to alignment with B1 level (f=28, 17.5%) whereas fast pace of the speakers and lack of face validity caused the listening skill to distance from B1 level (f=5, 6.8%).

Secondly, the factors affecting the alignment of B2-level listening skill are shown in Table 6.

The contributing factors to the alignment with B2 level (f=120) outweighed the detracting factors from B2 level (f=51). The effective factors were question design, discourse and intelligibility as contributing and detracting factors. First, question design was the most effective factor in B2-level listening skill. The same elements, objective of questioning, mode of questions and instruction, were reported to be on both sides: (a) contributing factor (f=61, 50.8%) and (b) detracting factor (f=30, 58.8%). Discourse both led to the alignment with B2 level (f=49, 40.8%) and caused the listening skill to distance from B2 level (f=17, 33.3%). Theme, language use and organization of the text were criticized by the instructors and

**Table 5: Effective Factors on the Alignment with B1-level Listening Skill**

| B1-Level Listening Skill | | | | | |
|---|---|---|---|---|---|
| Contributing Factors to the Alignment with Target Level | Theme | Discourse | Question Design | Intelligibility | Total |
| | Codes | Theme<br>Language Use<br>Organized Text<br>Lexical Appropriateness<br>Interculturality<br>Conversational Turn Structure<br>Genre | Objective of Questioning<br>Mode of Questions<br>Instructional Scaffolding<br>Question Roots and Options | Accent<br>Pace | |
| | f | 66 | 66 | 28 | 160 |
| | % | 41.25 | 41.25 | 17.5 | 100 |
| Detracting Factors from the Alignment with Target Level | Codes | Theme<br>Language Use<br>Semantic Organization<br>Scenario<br>Interculturality<br>Lexical Appropriateness<br>Length | Objective of Questioning<br><br>Mode of Questions<br><br>Question roots | Pace<br><br>Face Validity | |
| | f | 45 | 24 | 5 | |
| | % | 60.8 | 32.4 | 6.8 | |

**Table 6: Effective Factors on the Alignment with B2-level Listening Skill**

| B2-Level Listening Skill | | | | | | |
|---|---|---|---|---|---|---|
| Contributing Factors to the Alignment with Target Level | Theme | Question Design | | Discourse | Intelligibility | Total |
| | Codes | Objective of Questioning<br>Mode of Questions<br>Instructional Scaffolding | | Theme<br>Language Use<br>Organization of the Text<br>Extended Speech<br>Genre | Accent<br>Pace | |
| | f | 61 | 49 | | 10 | 120 |
| | % | 50.8 | 40.8 | | 8.3 | 100 |
| Detracting Factors from the Alignment with Target Level | Codes | Objective of Questioning<br>Mode of Questions<br>Unclear Instruction | Theme<br>Language Use<br>Organization of the Text<br>Length | | Pace<br>Accent | |
| | f | 30 | 17 | | 4 | 51 |
| | % | 58.8 | 33.3 | | 7.8 | 100 |

reported in both sides. However, inclusion of diverse genres and extended speeches had positive effect on the alignment whereas excessive length of the listening text had a negative effect. In terms of intelligibility, accent and pace were the elements emerged in contributing factors (f=10, 8.3%) and detracting ones (f=4, 7.8%).

The factors having an influence on the alignment of reading skill with B1 level are shown in Table 7.

The contributing factors to the alignment with B1 level (f=129) outweighed the detracting factors from B1 level (f=79). The effective factors in the alignment were discourse and question design. Instructors evaluated the language used in the reading texts and their themes in terms of discourse as contributing factor (f=63, 48.8%) and detracting ones (f=50, 63.3%). However, text design, length of the text and semantic organization were element to have been criticized as distancing factors from B1 level. Regarding question design, the contributing elements (f=66, 51.2%) outweighed the detracting ones (f=29, 36.7%). The shared elements were the mode of questions, instruction and objective of questioning, all of which may require a revision. On the other hand, mechanical mistake in the text and mismatch

between the test items and text led to distancing from B1 level even if roots of the test items were reported to be well-written.

As for B2-level reading alignment, the same factors were effective (Table 8).

The contributing factors to the alignment with B2 level (f=98) outweighed the detracting factors from B2 level (f=40). Discourse, as a contributing factor (f=40, 40.8%), included appropriate language used in the text, theme, well-designed text, diverse genre usage and the suitable length of the sentences in the text. However, language use, theme and the length of the text were criticized to be detracting factors from B2 level (f=21, 52.5%). Question design, as contributing factor (f=58, 59.2%), contained objective of questioning, well-designed distractors, appropriate length of the text and mode of questions whereas it included mode of questions, objective of questioning, redundant questions and not being cost effective which refers to the imbalance between the length of the text and the task requirement as detracting factor (f=19, 47.5%).

The factors on the alignment of language use were also revealed for B1 level (Table 9).

**Table 7: Effective Factors on the Alignment of Reading Skill with B1 Level**

| | | B1-Level Reading Skill | | Total |
|---|---|---|---|---|
| | **Theme** | **Discourse** | **Question Design** | |
| **Contributing Factors to the Alignment with Target Level** | **Codes** | **Language Use**<br>**Theme** | **Mode of Questions**<br>**Instruction**<br>**Objective of Questioning**<br>**Question Roots** | |
| | f | 63 | 66 | 129 |
| | % | 48.8 | 51.2 | 100 |
| Detracting Factors from the Alignment with Target Level | Codes | Theme<br>Text Design<br>Language Use<br>Length of the text<br>Organization of the text | Objectives of Questioning<br>Instruction<br>Mode of Questions<br>Mismatch with the Text<br>Mechanical Mistake in Text | |
| | f | 50 | 29 | 79 |
| | % | 63.3 | 36.7 | 100 |

**Table 8. Effective Factors on the Alignment of Reading Skill with B2 Level6**

| B2-Level Reading Skill | | | | |
|---|---|---|---|---|
| | **Theme** | **Discourse** | **Question Design** | **Total** |
| **Contributing Factors to the Alignment with Target Level** | **Codes** | **Language Use**<br>**Theme**<br>**Well-designed Text**<br>**Genre**<br>**Length of the sentences** | **Objective of Questioning**<br>**Well-Designed Distractors**<br>**Length of the Text**<br>**Mode of Questions** | |
| | f | 40 | 58 | 98 |
| | % | 40.8 | 59.2 | 100 |
| Detracting Factors from the Alignment with Target Level | Codes | Language Use<br>Theme<br>Length of the text | Mode of Questions<br>Objective of Questioning<br>Redundant Questions<br>Cost Effectiveness | |
| | f | 21 | 19 | 40 |
| | % | 52.5 | 47.5 | 100 |

**Table 9: Effective Factors on the Alignment of Language Use with B1 Level**

| | | B1-Level Language Use | | | |
|---|---|---|---|---|---|
| | **Theme** | **Question Design** | **Discourse** | **Task Complexity** | **Total** |
| Contributing Factors to the Alignment with Target Level | Codes | Text Validity<br>Objective of Questioning<br>Contextualization<br>Target Lexical Items<br>Length of the Text<br>Mode of Questions<br>Length of the Sentences | Contextualization<br>Text Validity<br>Language Use in the Texts<br>Theme | | |
| | f | 38 | 18 | | 66 |
| | % | 67.9 | 32.1 | | 100 |
| Detracting Factors from the Alignment with Target Level | Codes | Objective of Questioning<br>Mode of Questions<br>Semantic Distractors<br>Contextualization<br>Target Lexical Items<br>Balance between Target Lexical Items | Contextualization<br>Text Validity<br>Length of the Text<br>Theme<br>Language Use in the Texts | Cognitive Load | |
| | f | 38 | 12 | 1 | 51 |
| | % | 74.5 | 23.5 | 2 | 100 |

The contributing factors to the alignment with B1 level (f=66) and detracting factors from the alignment with B1 level (f=51) were similar in frequency.

The emerging factors for B1-level were discourse, question design and task complexity. Regarding question design, the positive effect on the alignment

(f=38, 67.9%) depended on many elements i.e., objective of questioning, mode of questions, level-appropriate target lexical items, contextualization, valid text usage, appropriate length of the text and sentences. However, the first four elements were also reported to be detracting factors from the alignment as well as the usage of higher-level semantic distractors and imbalance between the difficulty level of target lexical items (f=38, 74.5%). In terms of discourse, the same elements were reported to be contributing factors (f=18, 32.1%) and detracting ones (f=12, 23.5%), which were contextualization, validity of the texts, themes and language used in the texts. Additionally, the length of the text was reported not to be appropriate for B1 level. The task complexity (f=1, 2%) was only reported to be detracting factor from the alignment with B1 level as especially open-cloze tasks creates a huge cognitive load.

The factors for the alignment of language use with B2-level is demonstrated in Table 10.

The contributing factors to the alignment with B2 level (f=130) were more frequently reported than the detracting factors from the alignment with B2 level

(f=37) which were revealed to be question design and discourse. It could be seen that question design were more frequently reported to have a contributing factor on the alignment with B2 level (f=108, 83%) than detracting one from the alignment with B2 level (f=33, 89.2%). The mode of questions, objective of questioning, contextualization, target grammar and lexical items were reported on both sides. However, the language use task at B2 level were subjected to criticism in terms of the length of the test items, problematic placement of the gaps in contextualized tasks, and the difficulty discrepancy between the test items. The positive effect of discourse (f=22, 17%) were achieved by contextualization of the test items, appropriate themes of the texts and language used in the texts whereas the distancing effect (f=4, 10.8%) emerged from the usage of many private names in the texts and the excessive length of the text.

The factors on the alignment of writiskill reported by instructors were also revealed for B1 level (Table 11).

First, the frequencies of contributing factors to the alignment with B1 level (f=20) and detracting

**Table 10: Effective Factors on the Alignment of Language Use with B2 Level**

| B2-Level Language Use | | | | |
|---|---|---|---|---|
| Contributing Factors to the Alignment with Target Level | Theme | Question Design | Discourse | Total |
| | Codes | Mode of Questions<br>Contextualization<br>Target Grammar Structures<br>Target Lexical Items<br>Objective of Questioning | Theme<br>Language Use in the Text<br>Contextualization | |
| | f | 108 | 22 | 130 |
| | % | 83 | 17 | 100 |
| Detracting Factors from the Alignment with Target Level | Codes | Target Lexical Items<br>Target Grammar Structures<br>Objective of Questioning<br>Length of the Test Items<br>Decontextualization<br>Mode of Questions<br>Placement of the gaps<br>Difficulty Discrepancy | Private Names in the Text<br>Length of the Text | |
| | f | 33 | 4 | 37 |
| | % | 89.2 | 10.8 | 100 |

**Table 11: Effective Factors on the Alignment of Writing Skill with B1 Level**

| | | B1-Level Writing Skill | | | |
|---|---|---|---|---|---|
| | Theme | Question Design | Discourse | Assessment Parameters | Total |
| Contributing Factors to the Alignment with Target Level | Codes | Mode of Question Clear Instruction Valid Content | Valid Content Prompt Adequacy Language Use | | |
| | f | 15 | 5 | | 20 |
| | % | 75 | 25 | | 100 |
| Detracting Factors from the Alignment with Target Level | Codes | Mode of Question Objective of Questioning Valid Content | Invalid Content Prompt Inadequacy Language Use | Task Requirements Writing Constraints | |
| | f | 8 | 11 | 5 | 24 |
| | % | 33.3 | 45.8 | 20.8 | 100 |

factors from the alignment with B1 level (f=24) were similar. The effective factors for B1-level alignment were discourse, question design and assessment parameters. Regarding question design, the positive effect on the alignment (f=15, 75%) were based on three elements i.e., mode of questions, valid content of the writing prompts, and clear instructions. However, the detracting factors from the alignment with B1 level (f=8, 33.3%) included mode of questions, objective of questioning and invalid content of the texts in integrated reading and writing tasks. In terms of discourse, the same elements were reported to be contributing factors (f=5, 25%) and detracting ones (f=11, 45.8%), which were content of test items, language use in the test items, adequacy of writing prompts, which refers to comprehensibility. These elements were reported to be detracting factor more frequently than the contributing one. Assessment parameters (f=5, 20.8%) was only reported to be detracting factor from the alignment with B1 level. The task requirements and writing constraints were reported to be beyond B1 level.

The factors on the alignment of writing skill reported by instructors were also revealed for B2 level (Table 12).

The frequencies of contributing factors to the alignment with B2 level (f=9) and detracting

factors from the alignment with B2 level (f=7) were similar. B2-level writing skill was also evaluated by question design and discourse. It could be seen that question design were more frequently reported to have a contributing factor on the alignment with B2 level (f=5, 55.6%) than detracting one from the alignment with B2 level (f=2, 28.6%). The objective of questioning and clear instructions had a positive effect on the alignment; however, the objective of question was criticized to be a detracting one in addition to the difficulty discrepancy between the test items. The positive effect of discourse (f=4, 44.4%) were so close to detracting one (f=5, 71.4%). The valid content of the test items was reported to be on both sides, while the prompts' adequacy and topic relevance to the essay types were considered being distancing factors.

The factors having an influence on the alignment of speaking skill with the B1 level were presented in Table 13.

The detracting factors to the alignment with B2 level (f=36) were higher than the contributing factors from the alignment with B2 level (f=18). Instructors reported elements included in discourse, question design, learner-related issues and dynamics in assessment to be contributing factors (f=18) and detracting factors (f=36). Regarding discourse,

**Table 12: Effective Factors on the Alignment of Writing Skill with B2 Level**

| B2-Level Writing Skill | | | | |
|---|---|---|---|---|
| Contributing Factors to the Alignment with Target Level | Theme | Discourse | Question Design | Total |
| | Codes | Valid Content | Objective of Questioning Clear Instruction | |
| | f | 4 | 5 | 9 |
| | % | 44.4 | 55.6 | 100 |
| Detracting Factors from the Alignment with Target Level | Codes | Prompt Adequacy Valid Content Topic Relevance | Objective of Questioning Difficulty Discrepancy | |
| | f | 5 | 2 | 7 |
| | % | 71.4 | 28.6 | 100 |

**Table 13: Effective Factors on the Alignment of Speaking Skill with B1 Level**

| B1-Level Speaking Skill | | | | | | |
|---|---|---|---|---|---|---|
| | Theme | Discourse | Question Design | Learner-Related Issues | Dynamics in Assessment | Total |
| Contributing Factors to the Alignment with Target Level | Codes | Content of Test Items | Objective of Assessment Language Use Mode of Questions Equality in Difficulty Level | | | |
| | f | 11 | 7 | 0 | 0 | 18 |
| | % | 61.1 | 38.9 | 0 | 0 | 100 |
| Detracting Factors from the Alignment with Target Level | Codes | Unfamiliar Content of Test Items | Objective of Questioning Question Roots Difficulty Discrepancy Objective of Assessment | Content Knowledge of Learners | Role Inequity | |
| | f | 13 | 17 | 3 | 3 | 36 |
| | | 36.1 | 47.2 | 8.3 | 8.3 | 100 |

content of the test items was a controversial issue as its positive effect (f=11, 61.1%) was so close to its negative effect on the alignment with B1 level (f=13, 36.1%). Question design was also reported to be a distancing factor (f=17, 47.2%) rather than contributing one (f=7, 47.2%). The contribution arose from various elements: objective of assessment, which refers to production or interaction-based speaking test design, mode of questions based on the objective, language used in the test items, and equal difficulty level of them. On the other hand, the objective of questioning and assessment, roots of test items, and difficulty discrepancy between test items were criticized to be detracting factors from B1 level. Learner-related issues was only considered being a detracting factor (f=3, 8.3%) with the

**Table 14: Effective Factors on the Alignment of Speaking Skill with B2 Level**

| B2-Level Speaking Skill | | | | | | |
|---|---|---|---|---|---|---|
| | **Theme** | **Discourse** | **Question Design** | **Dynamics in Assessment** | **Learner-Related Issues** | **Total** |
| Contributing Factors to the Alignment with Target Level | Codes | Content of Test Items Prompt Adequacy | Objective of Questioning Mode of Questions Equality in the levels of prompts | Mode of Assessment | | |
| | f | 18 | 11 | 3 | 0 | 32 |
| | % | 56.4 | 34.3 | 9.3 | 0 | 100 |
| Detracting Factors from the Alignment with Target Level | Codes | Content of Test Items | Unclear Test Items Objective of Questioning Difficulty Discrepancy | Mode of Assessment | Content Knowledge of Learners | |
| | f | 10 | 17 | 2 | 1 | 30 |
| | | 33.3 | 56.7 | 6.7 | 3.3 | 100 |

element of content knowledge of learners, referring to the subjects of the test items being beyond the learners' general knowledge. Moreover, the dynamics in assessment was the last factor to detract speaking skill from B1 level (f=3, 8.3%). Specifically in interaction-based speaking tests, it was reported that the roles of the test-takers were not equal.

As for B2-level speaking tests, the factors having an influence on the alignment were presented in Table 14.

Their contribution to the alignment with B2 level (f=32) was found to be similar to their distancing effect (f=30). The effective factors were discourse, question design, dynamics in assessment and learner-related issues. In terms of discourse, the contributing elements (f=18, 56.4%) outweighed the detracting ones (f=10, 33.3%). The content of test items was reported on both sides even though the speaking prompts were found adequate for B2-level test-takers to comprehend. Regarding question design, the detracting elements (f=17, 56.7%) outweighed the contributing ones (f=11, 34.3%). Objective of questioning, mode of questions and equality in

difficulty levels of test items were the elements contributed to the alignment whereas unclear speaking test items and their problematic roots, objective of questioning and difficulty discrepancy between test items were the distancing elements. The factor of dynamics in assessment (f=3, 9.3%) was found to be more effective than detracting one through the appropriate mode of assessment which still signals a need for revision. Also, as a learner-related issue, the content knowledge of learners was reported to be a distancing factor (f=1, 3.3%).

## CONCLUSION AND DISCUSSION

The current research investigated the instructors' perception on the alignment of language tests with the target levels, B1 and B2 levels, based on skills and it revealed the factors that contributed to the alignment and detracted from the alignment with the target levels. To start with the B1 level tests, the results demonstrated moderate rate of alignment with B1 level in listening, language use and writing skills; however, the alignment rate of reading and speaking skills were at a satisfactory level. Secondly,

the results of B2 level tests demonstrated a high rate of alignment with B2 level in listening skill and language use whereas a satisfactory level of alignment was revealed in reading and writing skills. Speaking skill was moderately aligned with B2 level based on the instructors' perspective.

The factors that led instructors to determine whether the skills in the language tests align with the target level also emerged. Even if the CEFR does not notify what test methods should be adopted to achieve the described objectives (Alderson et al. 2006), the instructors conveyed their opinions on the factors that contributed to the alignment and detracted from the alignment based on their experiences, which may demonstrate the inconsistencies between the implementations.

Firstly, the effective factors on the alignment of listening skill with B1 and B2 levels were question design, discourse and intelligibility. Regarding question design, the objective of questioning, mode of questions, clarity in the instructions and level-appropriate roots of the test items and options were taken into consideration to have had more contributing effect on the alignment with B1 level than detracting one. B2-level listening tests also had appropriate objective of questioning, mode of questions and clear instructions, which may demonstrate tasks were designed based on the CEFR objectives (CoE, 2020) and tasks designed for the level. In terms of discourse, adoption of various genres at B1 and B2 level listening tests such as lectures, radio broadcasts, being in line with the CEFR (CoE, 2020) and the B1-level listening texts based on conversational turn structure, also supported by the CEFR (CoE, 2020), were the elements of contributing factors to the alignment with the target levels. However, theme, language use, semantic organization, lexical appropriateness, and inclusion of intercultural elements in the listening texts were criticized on both sides at B1-level listening tests and the length of the texts were found too long to test both B1-level and B2-level listening skill as it creates a processing demand (Alderson et al., 2006) even if the adoption of extended speech was revealed to be appropriate for B2 level. The attention span of the test takers ought to be considered for actual assessment of listening skill at the target level. These

could lead to a decline in the alignment rate as test-takers would not be able to construct meaning when there were inconsistencies in the discourse of the texts. As for intelligibility, the clear accents of speakers led to have a smooth understanding of the text at both target levels. The reason why instructors emphasized accent could be understood by the study of Kurtuldu and Özkan (2022), showing that the unfamiliar accents of the speakers make learners feel insecure, but the familiar ones help them be confident in listening skill. Also, their pace was an important issue for easy comprehension according to the target level in spite of the objections. These results could be supported the descriptor at B2 level: "Can understand … on concrete and abstract topics delivered in standard language or a familiar variety at normal speed" (CoE, 2020, p.51). Face validity was the last issue that requires revision at B1 level tests as it was important to test what was aimed. It was found to be a detracting factor because note-taking parts needed to include a guidance. Consequently, the contributing factors helped listening skill be at the target levels while all the detracting factors may have led to decrease the alignment rate.

Secondly, the alignment of reading skill at B1 and B2 levels was based on question design and discourse. At B1-level reading skill, the alignment was established through the mode of questions, objectives of questioning and instruction and roots of test items in terms of question design. However, the first three elements were also revealed to be detracting factors in addition to the mechanical mistake in text which is with the punctuation changing the meaning and the mismatch between test items and reading text. On the other hand, question design in B2-level reading skill depended on objective of questioning, mode of questions, well-designed distractors and appropriate length of the text even if mode of questions and objective of questioning were regarded as detracting factors. Besides, redundant questions which did not serve any objectives and lack of cost effectiveness led to decrease in the alignment rate which could also result in limited time to interact with the reading text making comprehension difficult for the test-takers. It could be said that the assessment of reading skill depends on not only the difficulty level

of the text but also the format and purpose of the test (Alderson et al., 2006). Regarding discourse, the alignment of B1-level reading skill was established through level-appropriate language use and familiar themes with the learners, which were also criticized to be distancing factors in addition to the complex text design, excessive length of the text and semantic organization. However, B2-level reading skill demonstrated that language use, theme, organized text structure, adoption of various genres, and the appropriate length of the sentences in the text were the elements to align. The CEFR also supports the usage of different genres with its scales (CoE, 2020). However, the language used in the text, themes and the length of the texts were found to be detracting factors, as well. However, it may be a signal to test whether B2-level test takers follow up the meaning in the context or not. Consequently, the instructors' perspective demonstrated that the factors that were effective on the alignment with the target levels were not only depended on the objectives described by the CEFR (CoE, 2020) but also many other elements regarding question design and discourse.

Thirdly, the alignment of language use at B1 level were founded by question design, discourse and task complexity. In terms of question design at B1-level language use, text validity, objective of questioning, mode of question, contextualization, target lexical items, length of the text and sentences played a significant role. The imbalance between target lexical items, semantic distractors and aforementioned contributing factors were revealed to be detracting factors as well. Regarding discourse, contextualized test items, language use in the texts, and themes of them were the elements leading to the alignment with B1 level. They also had distancing effect on the alignment with the target level in addition to the problematic length of the text. Cognitive load, as a detracting factor, was effective as part of the task complexity. Besides, the alignment of language use with B2-level was established through question design and discourse. Regarding question design, mode of questions, objective of questioning, target grammar structures and lexical items, and contextualization of test items were the elements leading to alignment even if they were also criticized to have distancing

effect. Furthermore, problematic length of the test items, placement of the test items in cloze test types, and difficulty discrepancy were elements distancing from B2 level. In discourse, theme, language used in the texts and contextualization were significantly effective on the alignment; however, the frequent usage of private names in the texts and excessive length of the texts led to detract from the alignment with B2 level. Consequently, linguistic competence in the CEFR (CoE, 2020) includes descriptors in terms of general linguistic range, grammatical accuracy, vocabulary control, vocabulary range, phonological control and orthographic control; however, these results based on the instructors' perspective demonstrate that alignment of language use with the target level could be achieved with the help of distinctive question design and discourse.

Next, the alignment of writing skill with B1 level was established through question design, discourse and assessment parameters. To start with question design, mode of questions, clear instruction and valid content of test items were effective on the alignment whereas mode of questions, objective of questioning and content of test items were criticized to have a distancing effect. Regarding discourse, content of the texts used in integrated reading and writing tests, adequacy of writing prompts and language used in the tasks led to have more distancing effect on B1 level. With regard to assessment parameters, task requirements like the number of paragraphs and writing constraints like word count were detracting factors from B1 level. As for B2-level writing skill, discourse and question design were effective on the alignment. Regarding discourse, the contribution of valid content of test items to the alignment was as influential as the detracting effect of inadequacy of prompts, their content and weak relevance of the topics with the essay types. The contribution of question design to the alignment was established through the objective of questioning and clear instructions; however, the objective of questioning was subjected to criticism as well as difficulty discrepancy between the test items. The results not only embrace the CEFR scales in terms of the context of writing (CoE, 2020) but also explain the alignment procedure in detail.

Lastly, the alignment of speaking skill with B1 level was established through discourse, question design, learner-related issues and dynamics in assessment. First, content of the test items was a controversial issue that requires a careful revision based on the objectives described by the CEFR (CoE, 2020) because the objective of questioning was one of the detracting factors in question design. To illustrate the other elements in the factor, mode of questions and language use positively affected the alignment with B1 level whereas the roots of test items were problematic, which creates a requirement of revision. The objective of assessment, production or interaction-based speaking test, and the discrepancy between difficulty levels of test items were controversial issues to be considered for validity and fairness issues in testing. Regarding learning-related issues, not having sufficient content knowledge of test takers on the test items may bring up a question of testing the speaking competence or general knowledge and result in not assessing the actual proficiency level of test takers. As part of dynamics in assessment, the imbalanced roles of test-takers, especially in interaction-based speaking tests, would lead to inequity, affecting the fairness in testing. The alignment of speaking skill with B2 level was established through discourse, question design, assessment parameters, dynamics in assessment and learner-related issues. Content of the test items, a part of discourse, was a controversial issue whereas the prompts were well-designed for B2-level test takers to comprehend. In terms of question design, the objective of questioning and difficulty discrepancy between the test items showed a dual-natured impact on the alignment. Moreover, the mode of questions had a contributing effect on the alignment; however, the test items could be clearer. These conflicts may prevent the standardization of the test items; thus, it may decrease the reliability and validity of the tests. The mode of assessment was the other controversial issue to be taken into consideration as well as the content knowledge of test takers. These problematic issues may hinder actual performance of test takers.

## Pedagogical Implications

This study presents valuable insights for EFL lecturers in the testing offices of the SFL at universities, the directors of the SFL at universities, ELT departments at universities, policymakers in language testing and CoHE in Turkey. A key takeaway for EFL lecturers working in the testing offices of the SFL at universities is to consider whether the descriptors of the CEFR illustrative scales at the target level overlap the objectives of the examinations. It should be noted that the examinations are prepared with different purposes, e.g., to test what they have learnt or not, to determine the test-takers' proficiency level. That's why lecturers at the testing offices are required to be knowledgeable of the CEFR descriptors and arrange their exam preparation procedure according to the target CEFR levels. The other major implication is for the directors of the SFL at universities who are responsible for designing the language curriculum and assessment. For an effective language testing and assessment process, they ought to take action to provide EFL lecturers with regular in-service training programs and workshops about the usage of the CEFR in language assessment. This is also particularly relevant for the ELT departments at universities, as a course related to the CEFR and its usage in language assessment could be integrated into the curriculum of ELT departments. Moreover, the lecturers who teach language assessment in ELT departments could also include CEFR-based assessment practices in their syllabus, both of which may provide the prospective teachers with the necessary knowledge of the CEFR for their future language teaching and assessment procedures. For policymakers, the current study highlights the significance of developing assessment frameworks which include diverse language skills. It should also be considered that to-be-created assessment frameworks enhance production-based and interaction-based language competencies, aligning with the target proficiency levels according to the CEFR. From a broader perspective, the current study recommends CoHE in Turkey to establish more standardized guidelines for language assessment at all universities as they are independent to prepare their language tests. Thus, the revelation of inconsistencies in test designs, test items in different skills, and assessment literacy among EFL lecturers in the testing offices could be lessened with the help of a standardized language testing system. All the implications of the current study would both enhance the reliability and

validity of language assessments and achieve greater alignment with the CEFR, which is considerably focused on in Turkish higher education policies

There are some limitations in the current study. Firstly, the data for this research was collected from a state university in Turkey and there were 6 participants who were active members of testing in preparatory classes, so the number of participants could be regarded as a limitation. Secondly, data collection took a long time because all the tests that were applied in the 2023-2024 academic year were included in the current study. Moreover, the tests did not have predetermined objectives before they were prepared, which led the data collection procedure to last long. Lastly, the results may not be generalized to all the language tests at B1 and B2 levels of English, as they were internal tests and taken from only one university. Despite the limitations mentioned above, this research is of importance in the study of the alignment of B1 and B2 level tests of English in accordance with the learning outcomes of the CEFR.

## Suggestions for Further Research

This study also presents valuable implications for the researchers in the field of language education and assessment for further research. First, further studies may involve an increased number of participants from diverse education settings both to enhance the generalizability of the findings and to obtain more diverse perspectives of experts. There is a need to further investigate the alignment process from different perspectives. Adoption of a mixed-method research design could enable a more comprehensive understanding of language assessment processes with the help of more generalizable quantitative data and in-depth information from qualitative data. Moreover, expanding the scope of research with the help of integration of curriculum design into the assessment practices could contribute to a more holistic approach in language teaching and assessment procedure.

## Acknowledgement

## REFERENCES

1. Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2006). Analysing tests of reading and listening in relation to the Common European Framework of Reference: The experience of the Dutch CEFR construct project. *Language Assessment Quarterly: An International Journal, 3*(1), 3-30. DOI: https://doi.org/10.1207/s15434311laq0301_2

2. Alderson, J. C. (2007). The CEFR and the need for more research. *The Modern Language Journal (MLJ), 91*(4), 659-663. https://www.jstor.org/stable/4626093

3. Baldwin, R., & Apelgren, B. M. (2018). Can Do and Cannot Do–CEFR inspired examination and assessment in a Swedish higher education context. *Apples: Journal of Applied Language Studies, 12(*2), 19-35. DOI: http://dx.doi.org/10.17011/apples/urn.201809144127

4. Byram, M., & Parmenter, L. (2012). *The common European framework of reference: The globalization of language education policy*. Bristol: Multilingual Matters.

5. Council of Europe (CoE). (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press. https://rm.coe.int/1680459f97

6. Council of Europe (CoE). (2018). *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment—Companion Volume with New Descriptors*. Strasbourg: CoE. https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989

7. Council of Europe (CoE). (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing, Strasbourg, https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

8. Creswell, J. W. (2015). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson.

9. Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. (5th ed). Sage publications.

10. Çağatay, S. & Ünveren-Gürocak, F. (2016). Is CEFR Really over There?. *Procedia-Social and Behavioral Sciences, 232*, 705-712. DOI: https://doi.org/10.1016/j.sbspro.2016.10.096

11. Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.

12. Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal (MLJ), 91(*4), 663-667. https://www.jstor.org/stable/4626094

13. Kavaklı, N., & Mirici, I. H. (2019). State-of-the-art of language testing and assessment in non-formal education: The case of English language schools in Turkey. *International Journal of Curriculum and Instruction, 11(*1), 65-88.

14. Khalifa, H., & French, A. (2008). *Aligning Cambridge ESOL examinations to the CEFR: Issues & practice*. Cambridge ESOL Research Notes, 37, 10-14.

15. Kurtuldu, E., & Özkan, Y. (2022). The Investigation of different English accents on listening comprehension. *The Reading Matrix: An International Online Journal, 22*(2), 50-62.

16. Le, H. T. T. (2018). Impacts of the CEFR-aligned learning outcomes implementation on assessment practice. Hue University Journal of Science: *Social Sciences and Humanities, 127*(6B), 87-99. DOI: https://doi.org/10.26459/hueuni-jssh.v127i6B.4899

17. Mison, S., & Jang, I. C. (2011). Canadian FSL teachers' assessment practices and needs: Implications for the adoption of the CEFR in a Canadian context. *Synergies Europe, 6*(1), 99-108.

18. Moonen, M., Stoutjesdijk, E., de Graaff, R., & Corda, A. (2013). Implementing CEFR in secondary education: impact on FL teachers' educational and assessment practice. *International Journal of Applied Linguistics, 23*(2), 226-246. DOI: https://doi.org/10.1111/ijal.12000

19. North, B. (2014). Putting the Common European Framework of Reference to good use. *Language Teaching, 47*(2), 228-249. DOI: https://doi.org/10.1017/S0261444811000206

20. Sabbir, F. (2019). Perceived view of teachers towards Pentaksiran Tingkatan Tiga (PT3)(Form Three assessment) English language: A case study. *Asian Journal of University Education, 15*(3), 34-44.

21. Schoch, K. (2020). Case Study Research. In G. J. Burkholder, et al. (Eds.), *Research Design and Methods: An Applied Guide for the Scholar-Practitioner* (pp. 245-258). SAGE Publications, Inc.

22. Stake, R. E. (2000). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (2nd ed., pp. 435-454). Thousand Oaks, CA: Sage.

23. Ünlücan-Tosun, F., & Glover, P. (2020). How Do School Teachers in Turkey Perceive and Use the CEFR?. *International Online Journal of Education and Teaching, 7*(4), 1731-1739. http://iojet.org/index.php/IOJET/article/view/1041

24. Yılmaz-Yakisik, B. Y., & Ünveren-Gürocak, F. Ü. (2018). A Comparative Study of Perceptions about the 'Common European Framework of Reference' among EFL Teachers Working at State and Private Schools. *International Online Journal of Education and Teaching, 5*(2), 401-417. http://iojet.org/index.php/IOJET/article/view/303/243