An Examination Turkish Proficiency Exams in Terms of CEFR (2020) Illustrative Descriptors
Türkçe Yeterlik Sınavlarının D-AOBM (2020) Örnek Tanımlayıcıları Açısından İncelenmesi

Gökhan Haldun DEMİRDÖVEN

Öğr. Gör. Dr, Sakarya Uygulamalı Bilimler
Üniversitesi
Rektörlük Ortak Dersler Bölüm Başkanlığı
Türk Dili Koordinatörlüğü
Lect. Dr., Sakarya University of Applied
Sciences, Rectorate – Department of
Common Courses, Turkish Language
Coordination Unit
gokhandemirdoven@subu.edu.tr
ORCID: 0000-0001-7892-5458

Alparslan OKUR

Prof. Dr, Sakarya Üniversitesi Türkçe ve
Sosyal Bilimler Eğitimi Bölümü,
Türkçe Eğitimi ABD
Prof. Dr., Sakarya University, Department of
Turkish and Social Sciences Education,
Division of Turkish Language Education
aokur@sakarya.edu.tr
ORCID: 0000-0002-2868-063X

DOI: 10.56720/mevzu.1795835

Abstract

This study aims to investigate the Turkish proficiency exams (TPEs) administered in Turkey from the perspective the language proficiency presented within the Common Framework for Foreign Languages Teaching (CEFR). Within this perspective, the TPEs (2020) conducted by some Turkish and Foreign Languages Application and Research Centers (TÖMERs), namely İstanbul University DİLMER, Gazi University TÖMER, Hacettepe University TÖMER, Dokuz Eylül University DEDAM and Sakarya University TÖMER were chosen as the sample and were examined regarding the illustrative descriptors emphasized in the CEFR. The qualitative research method was utilized due to its scope and content. In this respect, the cross-sectional screening model was utilized in the study in agreement with the nature of qualitative research. It was attempted to present a situational description of the contextual correlations of Turkish Proficiency Exams (2020) applied nationally in the country with the European Common Recommendations Framework (2020). In the related context, the test items first presented within the range of the TPEs in question were examined through content analysis in terms of the language proficiency they aimed to measure within the CEFR framework,, and they were divided into various categories in terms of illustrative descriptor scales, illustrative descriptors and common recommendation levels. As far the results are concerned, it was concluded that the test items utilized within the scope of the mentioned measurement tools exhibited an agglomeration towards certain illustrative descriptors presented in relation to certain illustrative descriptor scales related to the CEFR, and this situation negatively affected the content validity of the relevant TPEs. Based on the relevant context, increasing the number of illustrative descriptor scales and the diversity of illustrative descriptors that show alignment within the scope of testing instruments may be considered as a proposed solution.

Keywords: CEFR based testing and assessment, illustrative descriptor, illustrative descriptor scale, teaching Turkish as a foreign language, Turkish proficiency test.

Öz

Bu çalışma, Türkiye'de uygulanan Türkçe Yeterlik Sınavları'nı (TYS), Diller için Ortak Başvuru Metni (D-AOBM) kapsamında sunulan dil yeterlikleri açısından incelemeyi amaçlamaktadır. İlgili bağlamda, İstanbul Üniversitesi DİLMER, Gazi Üniversitesi TÖMER, Hacettepe Üniversitesi TÖMER, Dokuz Eylül Üniversitesi DEDAM ve Sakarya Üniversitesi TÖMER tarafından 2020 yılında gerçekleştirilen TYS'ler örneklem olarak seçilmiş ve D-AOBM'de vurgulanan örnek tanımlayıcılar açısından incelenmiştir. Araştırmanın kapsamı ve içeriği gereği nitel araştırma yöntemi kullanılmıştır. Bu doğrultuda, nitel araştırmanın doğasına uygun olarak çalışmada kesitsel tarama modeli uygulanmıştır. Ulusal olarak uygulanan Türkçe Yeterlik Sınavları'nın (2020), D-AOBM (2020) ile bağlamsal ilişkilerine ilişkin durumsal bir betimleme sunulmaya çalışılmıştır. İlgili bağlamda, söz konusu TYS'ler kapsamında sunulan test maddeleri öncelikle CEFR çerçevesinde ölçmeyi amaçladıkları dil yeterlikleri açısından içerik analizi yoluyla incelenmiş ve örnek tanımlayıcı ölçekler, örnek tanımlayıcılar ve ortak öneri seviyeleri açısından çeşitli kategorilere ayrılmıştır. Sonuçlar incelendiğinde, söz konusu ölçme araçları kapsamında kullanılan test maddelerinin, D-AOBM ile ilişkili belirli örnek tanımlayıcı ölçekler doğrultusunda belirli örnek tanımlayıcılara doğru bir kümelenme sergilediği ve bu durumun ilgili TYS'lerin kapsam geçerliliğini olumsuz etkilediği sonucuna varılmıştır. İlgili bağlamdan hareketle ölçme araçları kapsamında eşleşme gösteren örnek tanımlayıcı ölçek ve örnek tanımlayıcı çeşitliliğinin arttırılması, bir çözüm önerisi olarak ele alınabilir.

Anahtar kelimeler: D-AOBM tabanlı ölçme ve değerlendirme, örnek tanımlayıcı, örnek tanımlayıcı ölçek, Türkçe yeterlik testi, yabancı dil olarak Türkçe öğretimi.

Etik Beyan	Bu çalışma Prof. Dr. Alparslan Okur danışmanlığında 18.10.2022 tarihinde sunduğumuz/tamamladığımız Ulusal alanda yapılan Türkçe yeterlik sınavlarının Yabancı Diller Öğretimi Ortak Çerçeve Metni'nde belirtilen dil yeterlikleri açısından incelenmesi başlıklı doktora tezi esas alınarak hazırlanmıştır.
Yazar Katkıları	Çalışmanın Tasarlanması / *Conceiving the Study:* AO (%50), GHD (%50) Veri Toplanması / *Data Collection:* GHD (%100) Veri Analizi / Data Analysis: GHD (%100) Makalenin Yazımı / Writing up: GHD (%100) Makale Gönderimi ve Revizyonu / Submission and Revision: GHD (%100)

1.Introduction

The best way to define the concept of proficiency is to examine what it is not. At this point language proficiency exams designed to identify the existing linguistic skills of the candidates apart from any kind of educational process [1]are encountered as a crucial qualifier. It is simply because these exams are often confused with achievement tests or placement tests. While the skills of the individuals are significant in the proficiency exams, the skills acquired within a curriculum are taken into consideration in the achievement or placement tests. In this context, it would not be unfair to say that the proficiency exams are structured with regards to a certain theory and the level completion/achievement exams are constructed on the basis of a certain curriculum (Bachman, (1995)[2] language proficiency exams (henceforth TPE) which can be defined as conversion of grammar into performance [3] or the ability to apply

---

[1]   Hancock, Andy. "Attitudes and approaches to literacy in Scottish Chinese families" *Language and Education, 20* (5) (22 December 2006), 355-373.

[2]   Bachman, L.F. *Fundamental considerations in language testing.* Oxford: Oxford University Press, 1995.

[3]   Enric Llurda, "On Competence, Proficiency, and Communicative Language Ability", *International Journal of Applied   Linguistics* 10/1 (June 2000), 85-96.

the skills of the individual to real situations Harrison (1983)[4] refer to the process of measuring certain performance tasks created for real language skills and functions that an individual may need in the future (Heaton, 1990)[5]. A summarized and criterion-referenced testing and assessment application can be demonstrated as the most prominent features (Canale & Swain (1980)[6]. In these exams, which represent the individual's ability to perform in a second language, the language skills of the individual in a different language are judged by certain criteria (North, 2012)[7]. Testing and assessment procedures of these exams are explicated through a single scoring system in which each skill is attached to one another, and no feedback is given to the person regarding his/her weaknesses in line with the result. This exam type serves the purpose of comprehensively identifying the language skills of an individual. It is administered to reveal whether the individual has reached the minimum level of proficiency in a certain skill. The most distinctive characteristics are that they are summative, and criterion based. When we consider language, beyond being a communication tool, as a living organism that finds its existence within the scope of the individual's proficiency capacity and develops within its performative proficiencies, we can describe it, as [8] underlines, as a phenomenon that becomes a kind of self-realization effort on the focus of turning into action. This pursuit, which is the highest level in Maslow's (1943) [9]Pyramid of Needs, reflects the summative aspect of proficiency exams and acquires a certain form of reality when directly associated with the criteria that exemplify the life itself. The Common European Framework of Recommendations (CEFR) comes into use at this point.

Since Turkey is one the member states of the Council of Europe, Turkey is supposed to base the matters such as testing, assessment and exam prepara-

---

[4]   Andrew Harrison, *A language testing handbook.* (London: Macmillian, 1983), 27.

[5]   J.B. Heaton, *Classroom testing*. London: Longman, 1990), 68

[6]   Michael Canale- Merriill Swain, "Theoretical bases of communicative approaches to second language teaching and testing" *Applied Linguistics,* 1(1) (1980), 1-47. 26.

[7]   Brian North, *The development of a common framework scale of language proficiency.* New York: Peter Lang Publishing, 2012), 45.

[8]   Cahit Epçaçan, "Functional Linguistics as for Linguistic Applications in Turkish Language Teaching", *Procedia - Social and Behavioral Sciences* 116 (February 2014), 5063-5069.

[9]   Abraham Maslow, "A Theory of human motivation" *Psychological Review*, 50 (1943), 370-396.

tion on the criteria specified in the CEFR. The CEFR is an action-oriented system structuring built on the mechanics of 'being able to' descriptors related to organizing the foreign language teaching processes. This structure is shaped on the basis of illustrative descriptor, illustrative descriptor scales and common reference levels, which are structured on comprehensive authentic field studies. At this point, while the term 'illustrative descriptor' refers to the scope of authentic task content divided into the smallest decomposable building block, the term 'illustrative descriptor scale' represents the layered whole that these illustrative descriptors come together and form within the framework of certain contexts specific to skill areas. On the other hand, it is possible to describe the concept of 'common reference levels' as level ranges reflecting which step cluster the illustrative language competencies are positioned within the illustrative scales in terms of the relevant illustrative descriptor scales and calibrates them across the illustrative descriptor competencies. Mechanism is designed in a system mechanics, in which parts consisting of certain stages in the vertical direction are adapted to the representations of 'being able to' in the horizontal direction. The system-specific mechanism, on the other hand, becomes operational in the focus of performing certain actions/tasks that are likely to be encountered in real life, which we can define as 'performance'.

1.1. Statement of the problem

When the studies examining the problems encountered in teaching Turkish to foreigners (TTF) are examined, the fact that the TTF practices carried out nationally have not passed through an institutionalization process organized under a single roof is illustrated as the source of most problematic situations (Kocayanak, 2021[10]; see also Sertdemir, 2021[11]; Karagöl, 2020[12], Boylu,

---

[10]   Dilek Kocayanak, "*Yabancı Dil Olarak Türkçe Öğretenlerin Konuşma Becerisini Ölçme Sürecine Yönelik Uygulamaları/ Implementation of Turkish teachers as a foreign language in the process of testing   speaking skill"* (Ankara: Hacettepe University, Institute of Turkish Studies, Doktora Tezi, 2021), 135.

[11]   Ercan Sertdemir, "*Yabancı Dil Olarak Türkçe Öğretiminde Kullanılan Türkçe Yeterlik Sınavlarının Yenilenmiş Bloom Taksonomisine Göre İncelenmesi/ An analysis of Turkish proficiency exams used in teaching Turkish as a foreign language according to the Revised Bloom's Taxonomy"* (Çanakkale: Çanakkale Onsekiz Mart University, Institute of Educational Sciences, Yüksek Lisans Tezi, 2021), 78.

2019[13]; Gedik, 2017[14]; Boylu & Başar, 2016[15]; Bakır, 2014[16]; Işıkoğlu, 2015[17]; Durmuş, 2013[18]). This situation leads to significant discrepancies between TPEs, which are of great importance for international students, regarding the method of preparation, the way the TPEs are administered, the scope of content, the types of item formatting used, basic criteria and target skills, elements etc. Within this framework, the research problem statement is determined in that: "What is the level of validity between the Turkish Proficiency Exams (TPEs) implemented in Turkey and the illustrative descriptors presented in CEFR?".

1.2. Purpose of the research

Within the extent of this study, it was attempted to construct a cross-sectional state description based on categorical content analysis for the TPE (2020) implemente in Turkey. In relation to this goal, the measurement tools that constitute the sample have beeen addressed in terms of language profici-

---

[12] Efecan Karagöl, "Proficiency Exams in Teaching Turkish as a Foreign Language in TÖMER (Turkish and Foreign Languages Research and Application Centers)", *Dil ve Dilbilimi Çalışmaları Dergisi* 16/2 (28 June 2020), 930-947.

[13] Emrah Boylu, "*Yabancılara Türkçe Öğretiminde Ölçme ve Değerlendirme Uygulamaları ve Standart Oluşturma/Measurement assessment practices and standardization in teaching Turkish to foreigners.* (Çanakkale: Çanakkale Onsekiz Mart University, Institute of Educational Sciences, Doktora Tezi, 2016), 142.

[14] Enver Gedik, "*Yabancı Dil Olarak Türkçe Öğretiminde Ölçme ve Değerlendirme/An assessment and evaluation in teaching Turkish as a foreign language* (İstanbul: İstanbul Arel University, Institute of Social Sciences, Yüksek Lisans Tezi, 2017), 107.

[15] Emrah Boylu, "On the current situatıon of Turkish teaching centers and their standardization", *The Journal of Academic Social Sciences* 24/24 (01 January 2016), 309-309.

[16] Sıddık Bakır, "Turkish teaching centres to the foreigners in Turkey and Ataturk University Language Teaching Application and Research Centre (DILMER)]." *TAED*, 51 (23 April 2014), 435-456.

[17] Metin Işıkoğlu, "*Yabancı dil olarak Türkçe öğretiminde kullanılan yeterlik sınavlarının madde yazımı bakımından incelenmesi: Mersin ve Sakarya üniversiteleri örneği/Analysis of proficiency exams developed for teaching Turkish as a foreign language in terms of item writing samples of Mersin and Sakarya universities* (Erzurum: Atatürk University, Institute of Educational Sciences, Yüksek Lisans Tezi, 2015), 169.

[18] Mustafa Durmuş, "Teaching Turkısh to foreıgners: Problems, solutıons and ıdeas on the future of teachıng turkısh to foreıgners", *Adiyaman University Journal of Social Sciences* 11 (01 January 2013), 207-207.

ency presented in the CEFR and it was established which illustrative descriptor scale of the relevant test items match with which illustrative descriptors. In this way, the extent to which the relevant measurement tools reflect certain language skills was revealed. In this respect, we can list the sub-problem scopes under the scope of the study as follows:

1. What is the level of content validity alignment of the Gazi University (GU) TÖMER Turkish Proficiency Exam (TPE, 2020) with the CEFR?

2. What is the level of content validity alignment of the İstanbul University (IU) DILMER Turkish Proficiency Exam (TPE, 2020) with the CEFR?

3. What is the level of content validity alignment of the Hacettepe University (HU) TÖMER Turkish Proficiency Exam (TPE, 2020) with the CEFR?

4. What is the level of content validity alignment of the Dokuz Eylül University (DEU) DEDAM Turkish Proficiency Exam (TPE, 2020) with the CEFR?

5. What is the level of content validity alignment of the Sakarya University (SAU) TÖMER Turkish Proficiency Exam (TPE, 2020) with the CEFR?

1.3. Significance of the research

Since the TPEs administered nationally are not prepared as identical exams administered through a single center, there is an uncertainty regarding the reliability and validity of these exams. When the literature was reviewed, it was determined that there is no existing research that addresses TPEs within the framework of the illustrative descriptors presented in CEFR. Constructing a competency-based categorical content analysis for the context in question is important in order to produce a status description of the TPEs conducted nationally in Turkey. It is simply because the emergent situation will create an awareness about the issues to be considered while preparing measurement tools and will enable us to form an opinion about how TPEs can be made more qualified.

1.4. Assumptions of the research

It is assumed that the TPEs of the public institutions in 2020 constitute the study group. These exams are considered to represent the overall TPE universe.

1.5. Limitations of the research

The extent of this study was limited to the TPEs administered by five public institutions. The scope covered only the exams conducted in 2020.

2.Methodology

2.1. Design of the research

The study was planned within the framework of qualitative research method due to its scope and content. In this context, single screening model has been used within the framework of the general screening models.

2.2. Study group

The universe in the study were comprised of Turkish Proficiency Exams conducted nationally related to the discipline of teaching Turkish to foreigners. In order to draw a general framework within the extent of the study, the Turkish Proficiency Exams (2020) held at the GU TÖMER, IU DILMER, HU TÖMER and DEU DEDAM, SAU TÖMER, located in the metropolitan cities of Turkey, including Istanbul, Ankara, Izmir and Sakarya, were selected as samples. Within this framework, official permission was obtained to examine the examination documents from the year 2020 belonging to the institutions constituting the study group.

2.3. Data collection tools

The research data collection tools consist of the TPEs administered by GU TÖMER, IU DİLMER, HU TÖMER, DEU DEDAM, and SAU TÖMER. These exams were conducted in the year 2020.

2.4. Data Analysis

The results were acquired as a consequence of the categorical content analysis of the TPEs (2020) test items related to the relevant higher education institutions, conducted within the framework of illustrative descriptor scales. In this context, firstly, the test items presented within the extent of the TPEs in

question were examined on the basis of content analysis in terms of the language proficiency they aimed to measure within the CEFR framework, and the illustrative descriptors they matched with on this plane were accepted as the border frames, and the scales they were related to were divided into various categories in terms of language proficiency and levels.

Prior to this process, three field experts from the relevant institutions were selected as panelists—one from SAU, one from GU, and one from IU—and a coordinator was appointed in order to manage the panel. Subsequently, the panel become familiar with the CEFR itself, the approach to be followed in the panel, and the specific forms and grids to be used (Form A8: Initial Estimation of Overall CEFR Level; Form A24: Confirmed Estimation of Overall Examination Level; CEFR Content Analysis Grid for Listening & Reading and CEFR Content Analysis Grids for Speaking & Writing). Within this framework, two panels were conducted one week apart. In the first session, the field experts were presented with the Form A8 from *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) (2009)* and the CEFR Content Analysis Grids, and each panelist was asked to provide an initial estimation of the overall CEFR level for each item, offering a brief rationale and documenting references as intended. Following the analysis of TPEs in terms of CEFR categories, the content linkage results for each TPE were graphically profiled, revealing the structural content of these TPEs in relation to the CEFR subscales based on 'Communicative Language Activities' and 'Aspects of Language Competence'.

As the follow-up process, the second session was initiated, during which the field experts were encouraged to restructure the content analysis and connection-making processes. They were asked to review their justified predictions regarding each question item. Subsequently, the justified prediction analyses for each question item were compiled into a pool, allowing the contentious question items to be presented once again for analysis by the panelists. This eventually facilitated the shaping of the confirmed estimation of the overall CEFR level for each question item. In the final stage, the confirmed estimation of the overall examination level for each TPE was developed in accordance with Form A24, as outlined in *Relating Language Examinations to the*

Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) (2009).

To ensure reliability, the researchers conducted independent evaluations and coded the data; subsequently, they compared and discussed their findings and reached 100% agreement.

3.FINDINGS

3.1. GU TOMER TPE (2020) CEFR based data analyzes

Given the study results within the context of the first sub-problem suggested within the extent of the research, 40% (B1+ level) of the GU TOMER TPE (2020) listening skill test items; 36% (B2 level) of the test items in reading skills; 67% (B2+ level) of the writing skill test items and 66%(B2 (33%) and B2+ (33%) levels) of the speaking skill test items were structured on specific illustrative descriptors.

3.1.1. Data analysis of the GU TOMER (2020) listening skill test items

The GU TOMER TPE (2020) listening skill consisting of **A** (*4 multiple choice*), **B** (*5 yes-no*), **C** (*5 multiple choice*) and **D** (*6 true-false*) was structured under four sub-sections (20 test items). Table 1 illustrates the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 1: GU TOMER TPE's (2020) Listening Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B1	Understanding audio (or signed) media and recor-	Can understand key points and key details in stories and other narratives (e.g., a holi-	2	10	100

	dings	day story) provided the narrative is comp-rehensible (CEFR, 2020)[19].			
B1+	Overall oral comprehen-sion	Can understand un-complicated factual information, including general and specific messages related to everyday life or pro-fessional matters, usu-ally when presented in a familiar vernacular (CEFR, 2020, p. 48)[20].	8	40	100
B2	Overall oral comprehen-sion	Can follow a lengthy string of discourse and complex argu-ments/discussions, provided the subject is reasonably familiar and arguments are presented with clear indications (CEFR, 2020, p. 48)[21].	4	20	50
C1	Identifying cues and in-ferring (spo-ken, signed and written)	Has the ability to make inferences about atti-tude, mood, and goals and use contextual, grammatical and lexi-cal clues to make futu-re predictions (CEFR, 2020, p. 60)[22].	2	10	67
		Total:	16	80	

---

[19] Council of Europe, *Common European Framework of Reference for Languages: Companion Volume* (Namur: Council of Europe, 2020).

[20] Europe, *Common European Framework of Reference for Languages*.

[21] Europe, *Common European Framework of Reference for Languages*.

[22] Europe, *Common European Framework of Reference for Languages*.

When Table 1 is examined, it was revealed that 60% of the GU TÖMER TPE (2020) listening skill test items were aligned with the General Verbal Comprehension illustrative descriptor scale. Of these items, 40% were at the B1+ level and 20% at the B2 level, both formed on two illustrative descriptor axes.

3.1.2. Data analysis of the GU TOMER TPE (2020) reading skill test items

The GU TOMER TPE (2020) reading skill consisted of five parts (25 test items), including **A** (*5 multiple choice*), **B** (*4 multiple choice*), **C** (*2 multiple choice; 5 true-false*), **D** (*4 multiple choice*) and **E** (*5 true-false*) questions. Table 2 illustrates the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 2: GU TOMER TPE's (2020) Reading Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B1	Reading for information and argument	Can grasp plain, fact-based texts related to contexts related to fields of curiosity and research (CEFR, 2020, p. 57)[23].	4	16	100
B1+	Reading as a leisure activity	Can grasp the promotions such as movie, book, concert etc. (by understanding the basics) and periodicals (magazines, newspapers, etc.) produced to appeal to a wide rea-	2	8	100

---

[23] Europe, *Common European Framework of Reference for Languages*.

		dership (CEFR, 2020, p. 59)[24].			
B2	Reading for information and argument	Can identify opposing arguments, problem-solution transfer, cause-effect configurations in a text formatted based on a certain discourse (CEFR, 2020, p. 57)[25].	9	36	75
C1	Reading for information and argument	Can grasp in detail a wide range of large-scale and obscure texts, approaches, and implicit messages and clearly expressed ideas that are likely to be encountered in social, business, or scientific life (CEFR, 2020, p. 56)[26].	3	12	75
		Total:	18	72	

As presented in Table 2, the great majority (60%) of the GU TÖMER TPE (2020) reading skill test items were based on the illustrative descriptors. These descriptors corresponded to the B1, B1+, and B2 common reference levels.

3.1.3. Data analysis of the GU TOMER TPE (2020) writing skills test items

The GU TOMER TPE writing skill was structured within the framework of two sub-sections (3 open-ended test items); I (1 open-ended) and II (2 alternative open-ended). Table 3 presents the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

---

[24] Europe, *Common European Framework of Reference for Languages*.

[25] Europe, *Common European Framework of Reference for Languages*.

[26] Europe, *Common European Framework of Reference for Languages*.

Table 3: GU TOMER TPE's (2020) Writing Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B1+	Correspondence	Can create messages based on personal opinions about abstract or cultural topics (music, film, etc.) (CEFR, 2020, p. 83)[27].	1	33	100
B2+	Reports and Essays	Can evaluate a problem within the scope of different ideas or solutions (CEFR, 2020, p. 68)[28].	2	67	100
		Total:	3	100	

As depicted in Table 3, the extent of the writing skill was structured to reflect a certain illustrative proficiency majority. This majority (67%) was presented at the B2+ level.

3.1.4. Data analysis of the GU TOMER TPE (2020) speaking skill test items

The speaking skill of the GU TOMER TPE consisted of two parts (6 open-ended test items); I (3 open-ended alternative to each other) and II (3 open-ended alternative to each other). Table 4 illustrates the frequency distribution

---

[27] Europe, *Common European Framework of Reference for Languages*.
[28] Europe, *Common European Framework of Reference for Languages*.

data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 4: GU TOMER TPE's (2020) Speaking Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B2	Sustained monologue: putting a case)	Can create a clear argument by extending their viewpoints to a certain length and supporting it with auxiliary points and related examples (CEFR, 2020, p. 64)[29].	2	33	50
B2	Sustained monologue: putting a case)	Can explain a personal perspective on a current issue, considering the useful and inappropriate aspects of different alternative elements (CEFR, 2020, p. 64)[30].	1	16,5	25
B2	Overall oral production	Can make clear and detailed explanations/descriptions and presentations on a wide range of topics within his/her field of interest, expanding and supporting his/her thoughts with complementary po-	1	16,5	25

---

[29]   Europe, *Common European Framework of Reference for Languages*.
[30]   Europe, *Common European Framework of Reference for Languages*.

B2+	Overall oral pro-duction	ints and relevant examples (CEFR, 2020, p. 48)[31]. Can make clear, systematically developed explanations/descriptions and presentations with relevant supporting details appropriately highlighting key points (CEFR, 2020, p. 48)[32].	2	33	100
		Total:	6	66	

As highlighted in the table, 66.7% of the speaking skill test item pool was structured at the Continuous Single Speech level, corresponding to the B2 and above common reference levels. This structure was based on the illustrative descriptor *Making an argument (e.g., in a discussion)* and aligned with the General Verbal Production illustrative descriptor scales.

3.2. IU DILMER TPE (2020) CEFR based data analyzes

Given the study results in the context of the second sub-problem exhibited within the extent of the research, it was revealed that 75% (B2 level) of the listening skill test items, 32% (B1+ level) of the reading skill test items, 100% (B2+ level) writing skill of the test items and 33% (B2 level) of the speaking skill test items of the IU DILMER TPE (2020) were structured on specific illustrative descriptors.

3.2.1. Data analyzes of the IU DILMER TPE (2020) listening skill test items

The listening skill of IU DILMER TPE (2020) was presented under two parts (20 test items) as **I** (*10 multiple choice*) and **II** (*5 multiple choice; 5 true-false*). Table 5 illustrates the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

---

[31] Europe, *Common European Framework of Reference for Languages*.
[32] Europe, *Common European Framework of Reference for Languages*.

Table 5: IU DILMER TPE's (2020) Listening Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B1+	Overall oral comprehension	Can understand uncomplicated factual information, including general and specific messages related to everyday life or professional matters, usually when presented in a familiar language (CEFR, 2020, p. 48)[33].	4	20	80
B2	Overall oral comprehension	Can follow a lengthy string of discourse and complex arguments/discussions, provided the subject is reasonably familiar and arguments are presented with clear indications (CEFR, 2020, p. 48)[34].	15	75	100
		Total:	19	95	

As far as Table 5 is concerned, it was revealed that the IU DİLMER TPE (2020) listening skill test items were structured within the scope of two different illustrative descriptors on the same scale. These descriptors showed frequency distributions of 75% and 20%, respectively.

---

[33] Europe, *Common European Framework of Reference for Languages*.
[34] Europe, *Common European Framework of Reference for Languages*.

3.2.2. Data analyzes of the IU DILMER TPE (2020) reading skill test items

The IU DILMER (2020) reading skill was divided into three sub-sections: **I** (*8 pieces of matching*), **II** (*5 pieces of multiple choice; 4 pieces of true-false*) and **III** (*5 pieces of multiple choice; 3 pieces of true-false*). Table 6 presents the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 6: IU DILMER TPE's (2020) Reading Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B1	Reading for information and argument	Can grasp plain, fact-based texts related to contexts related to fields of curiosity and research (CEFR, 2020, p. 57)[35].	6	24	67
B1+	Reading for orientation	Can review simple, factual texts in a periodical, introductory booklet, or web, identify content information and come to a conclusion as to whether they have information that might be useful in practice (CEFR, 2020,	8	32	89

---

[35] Europe, *Common European Framework of Reference for Languages*.

		p. 56)[36].			
B2	Reading for information and argument	Can detect (contrasting arguments, problem-solution transfer, cause-effect configurations) in a text formatted based on a certain discourse (CEFR, 2020, p. 57)[37].	4	16	100
C1	Vocabulary range	Has extensive knowledge of technical and idiomatic language elements within the scope of his/her specialization and can make good use of these language elements (CEFR, 2020, p. 131)[38].	3	12	100
		Total:	21	84	

When we analyze Table 6, it is explicit that the reading skill test items of the IU DİLMER TPE (2020) were distributed across four different illustrative descriptor levels. These levels corresponded to B1 and above (56%), B2 (16%), and C1 (12%) common reference levels.

3.2.3. Data analyzes of the IU DILMER TPE (2020) writing skill test items

The writing skill of the IU DILMER TPE (2020) was formed within the extent of two alternative open-ended test items. Table 7 illustrates the frequency

---

[36] Europe, *Common European Framework of Reference for Languages*.
[37] Europe, *Common European Framework of Reference for Languages*.
[38] Europe, *Common European Framework of Reference for Languages*.

distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 7: IU DILMER TPE's (2020) Writing Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B2+	Reports and Essays	Can produce texts that systematically develop an argument, with appropriate emphasis on key points and supporting details (CEFR, 2020, p. 68)[39].	1	100	100
		Total:	1	100	

As described in Table 7, the scope of the writing skill of the IU DİLMER TPE (2020) was constructed around a specific illustrative proficiency. This proficiency was presented at the B2+ level.

3.2.4. Data analyzes of the IU DILMER TPE (2020) speaking skill test items

The speaking skill of the IU DILMER TPE (2020) was structured on the basis of two sub-sections: **I** (*2 open-ended*) and **II** (*1 open-ended*). Table 8 presents the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 8: IU DILMER TPE's (2020) Speaking Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework	Within the Framework

---

[39] Europe, *Common European Framework of Reference for Languages*.

				of Skills %	of Proficiency Level %
B2	Overall oral production	Can make clear and detailed explanations and presentations on a wide range of topics within his/her field of interest, expanding and supporting his/her thoughts with complementary points and relevant examples (CEFR, 2020, p. 48).[40]	1	33,3	33,3
B2	Sustained monologue: describing experience	Can comprehensively explain the value that events/facts and experiences represent for oneself (CEFR, 2020, p. 62)[41].	1	33,3	33,3
B2	Informal discussion (with friends)	Can play an active role in the discussion by making comments, expressing opinions clearly, evaluating alternative proposals and forming new hypotheses/assumptions against them within the context of an informal discussion about familiar contexts (CEFR, 2020, p. 64)[42].	1	33,4	33,4
		Total:	3	100	

---

[40] Europe, *Common European Framework of Reference for Languages*.
[41] Europe, *Common European Framework of Reference for Languages*.
[42] Europe, *Common European Framework of Reference for Languages*.

As highlighted in Table 8, the scope of the speaking skill of the IU DİLMER TPE (2020) was structured across three different illustrative descriptor levels. These levels corresponded to three different illustrative descriptor scales within the B2 common reference level.

3.3. HU TOMER TPE (2020) CEFR based data analyzes

As far as the results in the context of the third sub-problem revealed within study are concerned, it was revealed that 48% (B1+) of the listening skill test items, 36.4% (B1+ level) of the reading skill test items, 40% (B2 level) of the writing skill test items and 24% (B1 level) of the speaking skill test items of the HU TOMER TPE (2020) were structured on specific illustrative descriptors.

3.3.1. Data analyzes of the HU TOMER TPE (2020) listening skill test items

The HU TOMER TPE (2020) listening skill consisted of two parts (23 test items); **I** (*2 open-ended; 6 matching*) and **II** (*5 true-false; 5 yes-no; 5 matching*). Table 9 presents the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 9: HU TOMER TPE's (2020) Listening Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B1+	Overall oral comprehension	Can understand uncomplicated factual information, including general and specific messages related to everyday life or professional matters, usually when presented in a familiar vernacular (CEFR, 2020, p.	11	48	100

Level	Scale	Illustrative Descriptor			
		48)[43].			
B2	Overall oral comp-rehension	Can follow a lengthy string of discourse and complex argu-ments/discussions, pro-vided the subject is rea-sonably familiar and arguments are presented with clear indications (CEFR, 2020, p. 48)[44].	7	30	57
		Total:	18	78	

As described in Table 9, the scope of the HU TÖMER TPE (2020) writing skill was structured across four different illustrative descriptors from three different scales within the independent user (B2) common reference level. The majority of these descriptors (40%) were presented at the B2 level to form a specific illustrative proficiency.

3.3.2. Data analyzes of the HU TOMER TPE (2020) reading skill test items

The HU TOMER TPE (2020) reading skill was structured under 2 parts (22 test items); I (10 fill in the blanks) and II (2 open-ended; 2 multiple choice and 8 matching). Table 10 illustrates the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 10: HU TOMER TPE's (2020) Reading Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %

---

43 Europe, *Common European Framework of Reference for Languages*.
44 Europe, *Common European Framework of Reference for Languages*.

B1	Coherence and cohesion	Can present a series of short, discrete components as a connected linear sequence of points (CEFR, 2020, p. 141)[45].	3	14	50
B1+	Reading for orientation	Can review simple, factual texts in a periodical, introductory booklet, or web, identify content information and come to a conclusion as to whether they have information that might be useful in practice (CEFR, 2020, p. 56)[46].	8	36,4	66,6
B2	Reading for information and argument	Can detect (contrasting arguments, problem-solution transfer, cause-effect configurations) in a text formatted based on a certain discourse (CEFR, 2020, p. 57)[47].	2	9,1	50
		Total	13	59,5	

As described in Table 10, it is explicit that the HU TÖMER TPE (2020) reading skill was structured across three different illustrative descriptors within three different illustrative descriptor scales, covering 59.5% of the items. These descriptors formed a certain illustrative proficiency majority (36.4%) related to the B1+ level.

3.3.3. Data analyzes of the HU TOMER (2020) writing skill test items

The HU TOMER TPE (2020) writing skill was structured on two subsections (5 open-ended test items); **A** (*2 open-ended alternatives to each other*) and **B** (*3 open-ended alternatives to each other*). Table 11 illustrates the frequency distri-

---

[45] Europe, *Common European Framework of Reference for Languages*.

[46] Europe, *Common European Framework of Reference for Languages*.

[47] Europe, *Common European Framework of Reference for Languages*.

bution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 11: HU TOMER TPE's (2020) Writing Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B2	Creative Writing	Can produce clear and detailed descriptions/details of interests in different contexts.	2	40	50
B2	Creative Writing	Can examine a cinema product, written work or play (CEFR, 2020, p. 67)[48].	1	20	25
B2	Correspondence	Can write non-routine professional messages using appropriate language structure and rules, provided that they are limited to the facts.	1	20	25
B2+	Correspondence	Can produce a strong yet courteous grievance message, including supporting details and a statement of the desired outcome	1	20	25

---

[48] Europe, *Common European Framework of Reference for Languages*.

(CEFR, 2020, p. 83)[49].					
Total:			5	100	

As described in Table 11, the scope of the HU TÖMER TPE (2020) writing skill was structured across four different illustrative descriptors from three different scales within the independent user (B2) common reference level. The majority of these descriptors (40%) were presented at the B2 level to form a specific illustrative proficiency.

3.3.4. Data analyzes of the HU TOMER TPE (2020) speaking skill test items

The speaking skill of HU TOMER TPE (2020) was constructed within the extent of a question pool with 54 open-ended test items related to 16 different themes presented under various headings. Table 12 contains the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 12: HU TOMER TPE's (2020) Speaking Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B1	Sustained monologue: describing experience	Can make uncomplicated explanations on various, familiar topics within the scope of his/her interests (CEFR, 2020, p. 62)[50].	13	24	48
B1+	Sustained monologue: put-	Can offer simple reasons to support his/her point of view on a familiar	4	7	80

---

[49] Europe, *Common European Framework of Reference for Languages*.

[50] Europe, *Common European Framework of Reference for Languages*.

	ting a case	topic (CEFR, 2020, p. 64)[51].			
B2	General Verbal Production	Can make clear and detailed explanations/descriptions and presentations on a wide range of topics within his/her field of interest, expanding and supporting his/her thoughts with complementary points and relevant examples (CEFR, 2020, p. 48)[52].	6	11	50
B2+	Overall oral production	Can make clear, systematically developed explanations/descriptions and presentations with relevant supporting details appropriately highlighting key points (CEFR, 2020, p. 48)[53].	7	13	100
C1	Sustained monologue: putting a case	Can develop a thesis in a well-structured language by systematically considering the interlocutor's thoughts, emphasizing important points with supporting elements and concluding appropriately (CEFR, 2020, p. 64)[54].	3	5	100
		Total:	3	60	

51 Europe, *Common European Framework of Reference for Languages*.
52 Europe, *Common European Framework of Reference for Languages*.
53 Europe, *Common European Framework of Reference for Languages*.
54 Europe, *Common European Framework of Reference for Languages*.

As highlighted in Table 12, the HU TÖMER TPE (2020) speaking skill was structured across five different illustrative descriptors related to three common reference levels. The majority (60%) of these skills corresponded to B1 and above, B2 and above, and C1 levels.

3.4. DEU DEDAM TPE (2020) CEFR based data analyzes

As far as the results in the context of the fourth sub-problem revealed within study are concerned, it is explicit that 60% (B2 level) of the listening skills test items, 33.3% (B2 level) of the reading skill test items, 67% (B2+ level) of the writing skill test items and 20% (B2 level) of the speaking skill test items of the DEU DEDAM TPE (2020) listening skill test items were structured on specific illustrative descriptors.

3.4.1. Data analyzes of the DEU DEDAM TPE (2020) listening skill test items

The DEU DEDAM TPE (2020) listening skill was presented in the framework of two sub-sections (25 test items); **I** (*10 yes-no; 5 true-false*) and **II** (*5 true-false; 5 matching*). Table 13 contains the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 13: DEU DEDAM TPE's (2020) Listening Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B1+	Understanding audio (or signed) media and recordings Comprehension	Can understand the informational content of most recorded or published material using standard language on topics of perso-	10	40	100

		nal interest (CEFR, 2020, p. 51).[55]			
B2	Overall oral comprehension	Can follow a lengthy string of discourse and complex arguments/discussions, provided the subject is reasonably familiar and arguments are presented with clear indications (CEFR, 2020, p. 48)[56].	15	60	100
		Total:	25	100	

As illustrated in Table 13, the main part (60%) of the test items related to the listening skill of DEU DEDAM TPE (2020) demonstrated a specific illustrative descriptor match presented at the B2 level. Furthermore, the remaining test items overlapped with a specific illustrative descriptor presented as a sample at the B1+ level.

3.4.2. Data analyzes of the DEU DEDAM TPE (2020) reading skill test items

The DEU DEDAM TPE (2020) Reading Skills was structured in relation to two sub-headings (12 test items); **I** (*3 fill in the blanks; 7 multiple choice*) and **II** (*2 fill in the blanks*). Table 14 contains the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 14: DEU DEDAM TPE's (2020) Reading Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Profici-

---

55 Europe, *Common European Framework of Reference for Languages.*

56 Europe, *Common European Framework of Reference for Languages.*

				ency Level %
B1	Coherence and cohesion	Can present a series of short, discrete components as a connected linear sequence of points (CEFR, 2020, p. 141)[57].	1	8	100
B2	Vocabulary range	Can utilize a wide range of technical terms while communicating with other experts within the scope of his/her field of expertise (CEFR, 2020, p. 131)[58].	4	33,3	50
B2+	Conversation	Can establish relationships with interlocutors by using mutually understanding expressions of inquiry and participation, as well as by commenting on third parties and shared histories (CEFR, 2020, p. 73)[59].	1	8	100
C1	Reading for information and argument	Can grasp in detail a wide range of large-scale and obscure texts, approaches and implicit messages and clearly expressed ideas that are likely to be encountered in social,	1	8	100

---

[57] Europe, *Common European Framework of Reference for Languages*.

[58] Europe, *Common European Framework of Reference for Languages*.

[59] Europe, *Common European Framework of Reference for Languages*.

		business, or scientific life (CEFR, 2020, p. 57)[60].			
	Propositional precision	Can effectively use the linguistic channel of communication to indicate the strength of a claim, argument or position (CEFR, 2020, p. 141)[61].	1	8	100
		Total:	8	65,3	

As is clear in Table 14, the illustrative proficiencies of the DEU DEDAM TPE (2020) reading skill were matched across different levels. These proficiencies displayed a concentration towards the B2 common reference level (33%).

3.4.3. Data analyzes of the DEU DEDAM TPE (2020) writing skill test items

The DEU DEDAM TPE (2020) writing skill was created within the scope of three alternative open-ended test items. Table 15 illustrates the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 15: DEU DEDAM TPE's (2020) Writing Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B2+	Reports and essays	Can produce texts that systematically develop an argument, with appropriate emphasis on	2	67	67

---

[60] Europe, *Common European Framework of Reference for Languages*.

[61] Europe, *Common European Framework of Reference for Languages*.

		key points and supporting details (CEFR, 2020, p. 68)[62].			
B2+	Creative writing	Can produce detailed descriptions of authentic or unreal situations and experiences, following established conventions of the genre by specifying the relationship between ideas in linked texts (CEFR, 2020, p. 67)[63].	1	33	33
		Total:	3	100	

As illustrated in Table 15, the DEU DEDAM TPE (2020) writing skill was structured on the basis of two illustrative descriptors. Two of these descriptors (67%) represented the same illustrative proficiency offered at the B2+ intermediate reference level.

3.4.4. Data analyzes of the DEU DEDAM TPE (2020) speaking skill test items

The DEU DEDAM TPE (2020) speaking skill was structured under two subsections (21 open-ended test items); **I** (*1 open-ended*) and **II** (*20 open-ended test items*). Table 16 illustrates the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 16: DEU DEDAM TPE's (2020) Speaking Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of	Within the Framework of

---

62 Europe, *Common European Framework of Reference for Languages*.

63 Europe, *Common European Framework of Reference for Languages*.

				Skills %	Proficiency Level %
B1	Overall oral production	Can present an uncomplicated description of one of the various topics within its scope of interest in a series of linear items (CEFR, 2020, p. 48)[64].	3	15	30
B2	Sustained monologue: putting a case	Can create a clear argument by extending their viewpoints to a certain length and supporting it with auxiliary points and related examples (CEFR, 2020, p. 64)[65].	4	20	50
B2+	Overall oral production	Can make clear, systematically developed explanations/descriptions and presentations with relevant supporting details appropriately highlighting key points (CEFR, 2020, p. 48)[66].	3	15	67
		Total:	3	50	

As highlighted in Table 16, the DEU DEDAM TPE (2020) speaking skill consisted of a test item pool. Half of this pool (50%) was comprised of three different illustrative descriptors.

3.5. SAU TOMER TPE (2020) CEFR based data analyzes

As far as the results in the context of the fifth sub-problem revealed within study are concerned, it is clear that 37% (B2 level) of the reading skill test items, 36% (B2 level) of the writing skill test items, 50% (B2 and B2+ level) and

---

[64] Europe, *Common European Framework of Reference for Languages*.

[65] Europe, *Common European Framework of Reference for Languages*.

[66] Europe, *Common European Framework of Reference for Languages*.

20% (B2 level) of the speaking skill test items of the SAU TOMER TPE (2020) listening skill test items were structured on specific illustrative descriptors.

3.5.1. Data analyzes of the SAU TOMER TPE (2020) listening skill test items

The SAU TOMER TPE (2020) listening skill consisted of two sub-parts (12 test items); **I** (*6 multiple choice*) and **II** (*6 multiple choice*). Table 17 illustrates the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 17: SAU TOMER TPE's (2020) Listening Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B1+	Overall oral production	Can understand uncomplicated factual information, including general and specific messages related to everyday life or professional matters, usually when presented in a familiar language (CEFR, 2020, p. 48)[67].	2	20	100
B2	Identifying cues and inferring (spoken, signed and written)	Can use various strategies to realize comprehension, paying attention to key points and controlling comprehension using contextual clues (CEFR, 2020, p.	4	33,3	50

---

[67] Europe, *Common European Framework of Reference for Languages*.

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
		60)[68].			
C1	Vocabulary range	Has a command of common idiomatic expressions and every-day expression, can use words/signs quite well (CEFR, 2020, p. 131)[69].	2	20	100
		Total:	8	73,3	

The main part of the test items (73%) related to the listening skill of SAU TOMER TPE (2020), majority of which (53.3%) were at independent user (B1+ and B2) common reference levels, were structured within the scope of three different illustrative descriptors related to three different illustrative descriptor scales.

3.5.2. Data analyzes of the SAU TOMER TPE (2020) reading skill test items

The SAU TOMER TPE (2020) reading skill consisted of six parts (40 test items); **I** (*10 fill-in-the-blank*), **II** (*6 multiple choice*), **III** (*6 multiple choice*), **IV** (*6 multiple choice*), **V** (*6 multiple choice*) and **VI** (*6 multiple choice*). Table 18 contains the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 18: SAU TOMER TPE's (2020) Reading Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B1	Vocabulary range	Has a vocabulary that can be considered compre-	6	15	75

---

68 Europe, *Common European Framework of Reference for Languages*.
69 Europe, *Common European Framework of Reference for Languages*.

		hensive regarding everyday life and familiar contexts (CEFR, 2020, p. 131)[70].			
B1+	Reading for Information and Debate/Argument	Can grasp plain, fact-based texts related to contexts related to fields of curiosity and research (CEFR, 2020, p. 57)[71].	10	25	72
B2	Reading for information and argument	Can identify (contrasting arguments, problem-solution transfer, cause-effect configurations) in a text formatted based on a certain discourse (CEFR, 2020, p. 57)[72].	8	36	75
C1	Vocabulary range	Has extensive knowledge of technical and idiomatic language elements within the scope of his/her specialization and can make good use of these language	4	20	67

[70] Europe, *Common European Framework of Reference for Languages*.

[71] Europe, *Common European Framework of Reference for Languages*.

[72] Europe, *Common European Framework of Reference for Languages*.

	elements (CEFR, 2020, p. 131)[73].				
	Total:		22	96	

When we analyze Table 18, it is apparent that the central part (96%) of the test items of the SAU TÖMER TPE (2020) reading skill were structured within the scope of four different illustrative descriptors. The majority of these items (76%) corresponded to the independent user (B1+ and B2) common reference levels across two different scales.

3.5.3. Data analyzes of the SAU TOMER TPE (2020) writing skill test items

The SAU TOMER TPE (2020) writing skill was structured within the scope of 2 alternative open-ended test items. Table 19 contains the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 19: SAU TOMER TPE's (2020) Writing Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B2	Creative writing	Can produce clear and detailed descriptions/details of interests in different contexts (CEFR, 2020, p. 67)[74].	1	50	50
B2+	Reports and essays	Can produce detailed descriptions of authentic or unreal situations and experiences by following established con-	1	50	50

---

[73] Europe, *Common European Framework of Reference for Languages*.

[74] Europe, *Common European Framework of Reference for Languages*.

ventions of the relevant genre by specifying the relationship between ideas in linked texts (CEFR, 2020, p. 68)[75].				
Total:		2	100	

As illustrated in Table 19, the SAU TÖMER TPE (2020) writing skill was structured based on an illustrative descriptor. This descriptor was presented within the scope of the superiority common reference level (Vantage/B2).

3.5.4. Data analyzes of the SAU TOMER TPE (2020) speaking skill test items

The SAU TOMER TPE (2020) speaking skill was formed within the extent of a question pool with 10 open-ended test items. Table 20 contains the frequency distribution data within the framework of illustrative descriptors to which the relevant test items correspond.

Table 20: SAU TOMER TPE's (2020) Speaking Skill Frequency within the Framework of Illustrative Descriptors which the Relevant Test Items Correspond to

Level	Illustrative Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level %
B1	Sustained monologue: putting a case)	Can express thoughts on issues related to daily life using basic expressions (CEFR, 2020, p. 64)[76].	2	20	50
B2	Overall oral production	Can make clear and detailed explanations/descriptions and presentations on a wide	4	40	66,6

---

[75] Europe, *Common European Framework of Reference for Languages*.
[76] Europe, *Common European Framework of Reference for Languages*.

range of topics within his/her field of interest, expanding and supporting his/her thoughts with complementary points and relevant examples (CEFR, 2020, p. 48)[77].		
Total:	6	60

As highlighted in Table 20, 60% of the SAU TÖMER TPE (2020) speaking skill (independent user/B1 and B2) was constructed based on two illustrative descriptors. These descriptors were related to two different illustrative descriptor scales.

4. DISCUSSION, CONCLUSION AND RECOMMENDATIONS

4.1. Discussion

It is possible for the measurement tools to serve the purpose of evaluating language learning levels only by being associated with certain criteria. The proficiencies presented within the Framework Text are accepted as criteria units that enable this purpose to be materialized in today's foreign language education studies (Birol & Özbay, (2013)[78]and it becomes a prerequisite that the test items to be applied match the illustrative language proficiencies related to the common recommendation level that is aimed to be measured (Demirel, 2019)[79]. The context referred to here with the term 'illustrative language proficiency' is the phenomenon of the evolution of existing grammatical elements at the implicit level into 'possible' structures based on a certain authentic performance. Therefore, it would be true to characterize the language pro-

---

[77] Europe, *Common European Framework of Reference for Languages*.

[78] Cem Birol - Murat Özbay, *Yabancılara Türkçe öğretimi için ölçme ve değerlendirme soruları/Measurement and evaluation questions for teaching Turkish to foreigners* (Ankara: Pegem Academy, 2013), 96.

[79] Serkan Demirel, "*Türkçe Yeterlik Sınavı'nın uluslararası geçerliğe sahip bazı yeterlik sınavlarıyla karşılaştırılması/Comparison of Turkish Proficiency Exam (TPE) with international language proficiency exams"* (Nevşehir: Nevşehir Hacı Bektaş Veli University, Institute of Social Sciences, Yüksek Lisans Tezi, 2019), 108.

ficiency measurement as a situational description based on authentic performance.

Performance measurement specific to a particular discipline is considered within the scope of identifying the degree to which academic knowledge and skills related to that field are adapted to possible problem contexts in real life. Language, on the other hand, attains a certain value to the extent that it is transformed into a performance form. In other words, the individual data about language gains reality and reaches its measurable value to the extent that it is processed as output in a way to take action with a focus on problem, establish connections between relations, reason/purpose-justification, going from specific to general and making evaluations [80].

The level of a measurement tool acquires 'reliability' to the extent that it includes the performance illustrators of the common recommendation level to which it is intended to be associated. As unreliable measurement results cannot be considered valid, a measurement instrument is required to demonstrate a high level of reliability, that is, a minimal degree of random error (Demir, 2018)[81]. Therefore, the concept of 'scope validity', which is a basic testing and assessment element, comes into play and is at the forefront of the criterion elements that should be given importance in order to evaluate the quality of measurement tools.

The contextual association link between two the measurement tools that should measure the same construct is characterized as the content validity (Cronbach, 1980)[82]. One of the important source references that should be considered in the line of identifying the content validity of a measurement tool is the phenomenon of 'defining the impact domain. As far as this fra-

---

[80] İlhan Erdem, "Speech disorders encountered during speech therapy and therapy techniques", *Adiyaman University Journal of Social Sciences* 11 (01 January 2013), 415-415.

[81] Süleyman Demir, "*Çok kategorili bireyselleştirilmiş bilgisayarlı test uygulamalarının farklı madde seçim yöntemlerinde sonlandırma kuralları açısından incelenmesi/Investigation of different item selection methods in terms of stopping rules in polytomous computerized adaptive testing*" (Hacettepe University, Institute of Educational Sciences, Doktora Tezi, 2019), 108.

[82] Lee Cronbach, "*Validity on parole: how can we go straight?*" *New directions for testing and measurement: measuring progress over a decade in,* W. B. Schrader (Ed.) (Jassey-Bass, 1980) 99-108), 105.

mework is concerned, the impact domain of TPEs under the scrutiny within the present study is the B2, B2+, C1 and C1+ common reference levels specified in the CEFR.

Given the results of the study, we can say that the TPEs applied nationally in Turkey are administered to include B1 and B1+ levels in addition to the relevant language levels. At this juncture, it may seem natural that the relevant framework includes the B1+ level as well. It is simply because, as emphasized in CEFR (2020, p. 38)[83], the descriptive qualities of a higher language level begin to manifest themselves primarily at the plus levels of a lower level. This situation also overlaps with the 'rainbow' metaphor used when creating common reference levels. When structuring the sequential language proficiency levels, there should be a 'transitivity' situation, such as a rainbow color sequence, instead of sharp lines to indicate the relationship between borders.

As far as the impact domain definition phenomenon is concerned, on the other hand, at the level of the B1 common reference level, the assumption that there may be a 'transitivity state' operating in the opposite direction emerges this time. It is simply because the 'transitive state' is a phenomenon that can be associated with a higher language level as well as with the plus level of a lower common reference level. That is to say, we can say that a test item that matches the B1 common reference level also brings with it the negativity of approaching the A2+ language level limits.

Criterion validity, which is another sub-component of content validity, refers to the relationship structure that has a statistically significant value between '*measure*' and '*criterion*' phenomena[84] and takes responsibility in the context of deciding to what extent a measurement tool serves the output or product it is designed to measure (Piedmont, 2014)[85]. The criterion system of CEFR emerges when the vertical structuring of illustrative descriptors scales is constructed with 'can do' statements on horizontal planes. In this sense, each

---

83   Europe, *Common European Framework of Reference for Languages*.
84   Jum Nunnally - Ira Bernstein, Psychometric theory. (3rd ed.). (McGraw Hill, 1994), 8.
85   Ralp Piedmont, (2014). "Criterion validity", "*Encyclopedia of quality of life and well-being research, in*" eds. Michalos, A.C. (Springer, 2014), 1348.

of the illustrative descriptors can be considered as a functional criterion statement for a basic purpose. Therefore, each illustrative descriptor is a component of criteria-based measurement practices and constitutes the smallest decomposable particle of the measurement context in question. To the extent that these particles vary, they represent authentic real-life tasks and contribute to the transformation of measurement applications into a plural color mosaic form.

It is noticeable that in each skill area within the measurement tools analyzed within the extent of the study, there is a backlog especially on certain illustrative descriptors. This situation leads to associating the range line regarding the criterion range of the mentioned TPEs with a limited action network; in other words, being able to demonstrate certain language skills only in certain contexts and language competencies in the scope of *vantage/B2* and *effective operational proficiency/C1 common reference levels* bring together the negativity of retention. Therefore, perspective structuring in a depth ranging from basic language skills related to daily life to various contextual competences that require social, professional or expertise led it to be equated with a narrow criterion system. In this sense, it would be congruous to emphasize the fact that considering each test item within the scope of measurement tools as a different criterion item as the smallest decomposable particle (*quanta*) that constitutes the criterion system will increase the content validity of the proficiency exams.

Apart from the four basic language skills, there was also a section in the DEU TPE (2020) where grammatical proficiencies were measured. Table 21 is composed of statistical data analyzes structured based on grammatical proficiencies within the scope of this field.

Table 21: DEU DEDAM TPE (2020) Grammatical Skill Frequency within the Framework of Sample Illustrative Proficiencies which the Relevant Test Items Correspond to

Level	Scale	Illustrative Descriptor	Number of Items	Within the Framework of Skills %	Within the Framework of Proficiency Level

					%
A1	DEU DEDAM Grammar Curriculum	The '-(I)yor+ possessive suffix' dominates the functions of the present tense structure.	1	4	100
A2	DEU DEDAM Grammar Curriculum	Verbal (noun/adverb) structures dominate their functions.	2	8	50
B1	DEU DEDAM Grammar Curriculum	'-DIk+ possessive suffix' dominates the functions of the adjective verb structure.	4	16	100
B2	Propositional Precision	It can change the way the message is conveyed (CEFR, 2020, p. 141).	10	40	100
C1	DEU DEDAM Grammar Curriculum	'-mAktAnsA' dominates the functions of the adverb verb structure.	1	4	100
			Total	72	

Even though CEFR does not recommend any language teaching or pedagogical approach since it adopts impartiality in terms of the purpose of its emergence, it does not make any recommendation to abandon grammar teaching either. It treats the language user as a composite element equipped with action-oriented skills on multicultural social organization structures and offers general grammatical proficiencies within the scope of Communicative Language Competencies. When the DEU DEDAM TPE (2020) was considered in line with grammar field proficiencies, it is explicit that the relevant test items matched both Turkish-specific grammar skills and pragmatic competence proficiencies. As indicated in Table 21, these matches show an intensity on the B2 co-recommendation level axis of the Propositional Precision sample illustrative scale. In this regard, we can say that the points suggested above to be considered for basic skill areas may also apply to the test items to be prepared specifically for grammar proficiency.

4.2. Conclusion

The results obtained in this study, implemented in order to construct a CEFR-based categorical content analysis for the TPEs administered nationally in Turkey, can be itemized as follows in accordance with the order concerning the sub-problems:

➢ When the data analyzes implemented within the framework of the study are considered within the scope of the first sub-problem of the study, it was revealed that 40% (B1+ level) listening skill test items, 36% (B2 level) of the reading skill test items, 67% (B2+ level) of the writing skill test items and 66% (33% B2; 33% B2+ levels) of the speaking skill test items of the GU TOMER TPE (2020) were structured on a specific illustrative descriptor.

➢ When the data analyzes conducted within the framework of the study are considered within the scope of the second sub-problem of the study, it was found that 75% (B2 level) of the listening skill test items, 32% (B1+ level) of the reading skill test items, 100% (B2+ level) of the writing skill test items and 33% (B2 level) of the speaking skill test items of the IU DILMER TPE (2020) were structured on a specific illustrative descriptor.

➢ When the data analyzes implemented within the framework of the study are considered within the scope of the third sub-problem of the study, it was revealed that 48% (B1+) of the listening skill test items, 36.4% (B1+) level) of the reading skill test items, 40% (B2 level) of the writing skill test items and 24% (B1 level) of the speaking skill test items of the HU TOMER TPE (2020) were structured on a specific illustrative descriptor.

➢ When the data analyzes performed within the framework of the study are considered within the scope of the fourth sub-problem regarding the study, it was found that 60% (B2 level) of the listening skill test items, 33.3% (B2 level) of the reading skill test items, 67% (B2+ level) of the writing skill test items and 20% (B2 level) of the speaking skill test items of the DEU DEDAM TPE (2020) were structured on a specific illustrative descriptor.

➢ When the data analyzes executed within the framework of the study are considered within the scope of the fifth sub-problem of the study, it was concluded that 37% (B2 level) of the listening skill test items, 36% (B2 level) of

the reading skill test items, 50% (B2 and B2+ levels) of the writing skill test items and 20% (B2 level) of the speaking skill test items of the SAU TOMER TPE (2020) were structured on a specific illustrative descriptor.

4.3. Recommendations

Multilingual and multicultural society structures tend to evolve into plurilingual and pluricultural organic formations in today's world as a past aspect of the millennium line. The increase in foreign lineage within the borders of Turkey constitutes an indicator of the tendency of this spark to catch fire. As such, the Turkish Proficiency Examinations has been increasingly drawing attention as an important context and serve as a prerequisite for foreigners who have completed their language learning process within the scope of university education.

As the most important problem case regarding the Turkish Proficiency Exams administered in the national field within the study, the language proficiency matched within the scope of the test items demonstrates an accumulation on certain illustrative descriptors, and at this juncture the concept of content validity comes into play. The measurement tools provide valid data in so far as that they can broadly reflect the relevant scope. In other words, the proficiency exams are justifiable to the extent that the common reference level or levels to which they want to be associated include performance illustrators.

The CEFR has been structured on the measurement mechanics in which parts of certain sets of steps in the vertical direction are adapted to the expressions of 'can do' in the horizontal direction. Illustrative descriptors, on the other hand, are in the position of the smallest criterion component that can be differentiated regarding this mechanism. Therefore, each illustrative descriptor creates a sample of variance that can be evaluated on its own for the set of digits it is in and represents the smallest unit that can be scored for measurement tools. By way of explanation, we can say that in order to increase the scope validity of TPEs, first of all, the illustrative descriptor and illustrative descriptor scales diversity, which are matched on the axis of measurement tools, should be enriched.

The Companion Volume, which is the CEFR (2020) version, has undergone a modernization process with the updated studies on sample illustrative

scales and sample illustrative proficiency descriptions. In the updated and newly added illustrative scales, it is explicit that the illustrators in the expert language use common recommendation levels (C1-C2) are addressed in detail. It is noticeable that the target language speaker is preferred in the new illustrators instead of the native speaker used in the illustrative descriptions of the first version, especially for the B2 and C2 common recommendation levels. This change is valuable in that it clearly emphasizes that the concept of pluralingulism is prioritized in the CEFR, which is designed with contemporary language proficiencies and that the 'new world' language proficiencies should be addressed in the relevant context. Therefore, it is crucial vital to increase the diversity of sample illustrative scales and proficiency descriptions related to TPEs with the focus of CEFR-based innovations. It is because the performance factors, which are the parameters in the axis of measurement criteria, will get rid of the memorized content and the TPEs will transform into an organism of authentic proficiencies, which tends to be constantly updated.

References

Bachman, L.F. *Fundamental considerations in language testing.* Oxford: Oxford University Press, 1995.

Bakır, Sıddık. "Turkish teaching centres to the foreigners in Turkey and Ataturk University Language   Teaching Application and Research Centre (DILMER)." *TAED*, 51 (23 April   2014), 435-456.

Birol, Cem – Özbay, Murat. *Yabancılara Türkçe öğretimi için ölçme ve değerlendirme soruları. [Measurement   and evaluation questions for teaching Turkish to foreigners].* Ankara: Pegem Academy, 2013.

Boylu, Emrah- Başar, Umut. "On the current situatıon of Turkish teaching centers and their standardization" *The Journal of Academic Social Science*, *4724*   (01   January   2016),   309-324. https://asosjournal.com/?mod=tammetin&makaleadi=&makaleurl=8 75945058_1042%20Em   ah%20BOYLU.pdf&key=33557

Boylu, Emrah. *Yabancılara Türkçe öğretiminde ölçme değerlendirme uygulamaları standart oluşturma   [Measurement assessment practices and standardiza-*

tion in teaching Turkish to foreigners]. Çanakkale: Çanakkale   Onsekiz Mart University, Institute of Educational Sciences, Doktora Tezi, 2016.

Canale, Michael – Swain, Merrill. "Theoretical bases of communicative appro- aches to second language   teaching and testing" *Applied Linguistics, 1*(1)   (01   March   1980),   1-47. https://academic.oup.com/applij/article-abstract/I/1/1/181953

Cronbach, L. J. "Validity on parole: how can we go straight?" *New directions for testing and measurement: measuring progress over a decade in,* W. B. Schrader (Ed.) *99-108. Jassey-Bass, 1980.*

Demirel, Serkan. *Türkçe Yeterlik Sınavı'nın uluslararası geçerliğe sahip bazı yeterlik sınavlarıyla   karşılaştırılması.* [*Comparison of Turkish Proficiency Exam (TPE) with international languageproficiency exams*]. Nevşehir: Nevşehir Ha- cı Bektaş Veli University, Institute of Social Sciences, Yüksek Lisans Tezi, 2019.

Durmuş, Mustafa. "Teaching Turkısh to foreıgners: Problems, solutıons and ıdeas on the future of   teachıng turkısh to foreıgners" *Adıyaman Univer- sity Journal of Social Sciences*, *6*(11) (31 January 2013),   207-228. https://dergipark.org.tr/tr/pub/adyusbd/issue/1392/16419

Epçaçan, Cahit. "Functional Linguistics as for Linguistic Applications in Tur- kish Language Teaching"   *Procedia - Social and Behavioral Sciences* 116 (February   2014),   5063-5069.   https://www.sciencedirect.com/science/article/pii/S18770428140109 1X

Erdem, İlhan. "Speech disorders encountered during speech therapy and the- rapy techniques" *Adıyaman University Journal of Social Sciences* 11 (01 Janu- ary 2013), 415-415.   https://dergipark.org.tr/tr/download/article- file/15158

Europe, Council of. *Common European Framework of Reference for Languages: Companion Volume.*   Namur: Council of Europe, 2020.

Gedik, Enver. *Yabancı dil olarak Türkçe öğretiminde ölçme ve değerlendirme* [An *assessment and evaluation in teaching Turkish as a foreign language*]. İstanbul:

İstanbul Arel University, Institute of Social Sciences, Yüksek Lisans Tezi, 2017.

Hancock, Andy. "Attitudes and approaches to literacy in Scottish Chinese families" *Language and Education, 20* (5) (22 December 2006), 355-373. DOI: 10.2167/le641.0

Harrison, Andrew. *A language testing handbook.* London: Macmillian, 1983.

Heaton, J.B. *Classroom testing*. London: Longman, 1990.

Işıkoğlu, Metin. *Yabancı dil olarak Türkçe öğretiminde kullanılan yeterlik sınavlarının madde yazımı bakımından incelenmesi: Mersin ve Sakarya üniversiteleri örneği.* [*Analysis of proficiency exams developed for teaching Turkish as a foreign language in terms of item writing samples of Mersin and Sakarya universities*]. Erzurum: Atatürk University, Institute of Educational Sciences, Yüksek Lisans Tezi, 2015.

Karagöl, Efecan. "Proficiency Exams in Teaching Turkish as a Foreign Language in TÖMER (Turkish and Foreign Languages Research and Application Centers)". *Dil ve Dilbilimi Çalışmaları Dergisi 16/*2 (28 Haziran 2020), 930-947. https://doi.org/10.17263/jlls.759347

Kocayanak, Dilek. *Yabancı dil olarak Türkçe öğretenlerin konuşma becerisini ölçme becerisini ölçme sürecine yönelik uygulamaları* [*Implementation of Turkish teachers as a foreign language in the process of testing speaking skill*]. Ankara: Hacettepe University, Institute of Turkish Studies, Doktora Tezi, 2021.

Llurda, Enric. "On Competence, Proficiency, and Communicative Language Ability". *International Journal of Applied Linguistics* 10/1 (June 2000), 85-96. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1473-4192.2000.tb00141.x

Maslow, Abraham. "A Theory of human motivation" *Psychological Review, 50* (1943), 370-396. https://psycnet.apa.org/record/1943-03751-001

Nunnally, Jum- Bernstein, Ira. *Psychometric theory*. McGaw Hill, 1994.

North, B. *The development of a common framework scale of language proficiency.* New York: Peter Lang   Publishing, 2012.

Piedmont, Ralp. "Criterion validity". *Encyclopedia of quality of life and well-being research*, Michalos, A.C. (Eds). Springer, 2014.

Sertdemir, Ercan. *Yabancı dil olarak Türkçe öğretiminde kullanılan yeterlik sınavlarının yenilenmiş Bloom   taksonomisine göre incelenmesi.* [*An analysis of Turkish proficiency exams used in teaching Turkish as a   foreign   language   according to the Revised Bloom's Taxonomy*]. Çanakkale: Çanakkale Onsekiz Mart   University, Institute of Educational Sciences, Yüksek Lisans Tezi, 2021.