

## Use of Rubric in the Assessment of Writing Skills in Teaching Turkish as a Foreign/Second Language: B1-B2 Level

Aybegüm Aksak \*

### ARTICLE INFO

Received: 11.09.2024  
Revised form: 15.03.2025  
Accepted: 23.03.2025  
Doi: 10.31464/jlere.1548666

### Keywords:

*Turkish as a foreign language  
Turkish as a second language  
writing skills  
assessment, rubric*

### ABSTRACT

Rubrics offer a solution to assessment inconsistencies frequently observed among evaluators at Turkish teaching centers, both within and across institutions. Particularly for productive language skills, rubrics can mitigate subjective evaluation by providing structured assessment frameworks. This research aimed to develop valid and reliable analytic rubrics for assessing B1-B2 level writing skills in Turkish language instruction. Using an exploratory sequential mixed methods design, the study followed systematic rubric development protocols. Validity was established through expert consultation with seven specialists using the Davis technique, while reliability was determined by having 20 evaluators score six writing samples, analyzed using Kendall's W coefficient of concordance. The study also examined whether evaluation experience (Kruskal Wallis-H test) or instructor background (foreign language vs. second language Turkish instruction; Mann Whitney U test) influenced scoring patterns when using the rubrics. Results confirmed the developed rubrics' validity and reliability. No significant differences emerged between ratings from experienced versus inexperienced instructors, or between foreign language and second language Turkish instructors, demonstrating the rubrics' effectiveness in standardizing assessment across evaluator profiles. These findings suggest that properly constructed rubrics can successfully reduce assessment inconsistencies regardless of evaluator experience or instructional background.

### Acknowledgments

The study was produced from the doctoral thesis prepared in the Department of Teaching Turkish as a Foreign Language under the supervision of Assoc. Prof. Dr. Meltem EKTİ at Hacettepe University Institute of Turkology Research.

### Statement of Publication Ethics

The study was conducted in accordance with scientific publication ethics. Ethics Committee approval was obtained for this study: Committee Name: Hacettepe University Rectorate Date and Decision number: 31.03.2023, E-35853172-000-00002773855

### Authors' Contribution Rate

If there is more than one author, please indicate the contribution rates of the authors.

### Conflict of Interest Reference

There is no conflict of interest in this study.

Aksak, A. (2025). Use of rubric in the assessment of writing skills in teaching Turkish as a foreign/second language: B1-B2 level. *Journal of Language Education and Research*, 11 (1), 369-403.

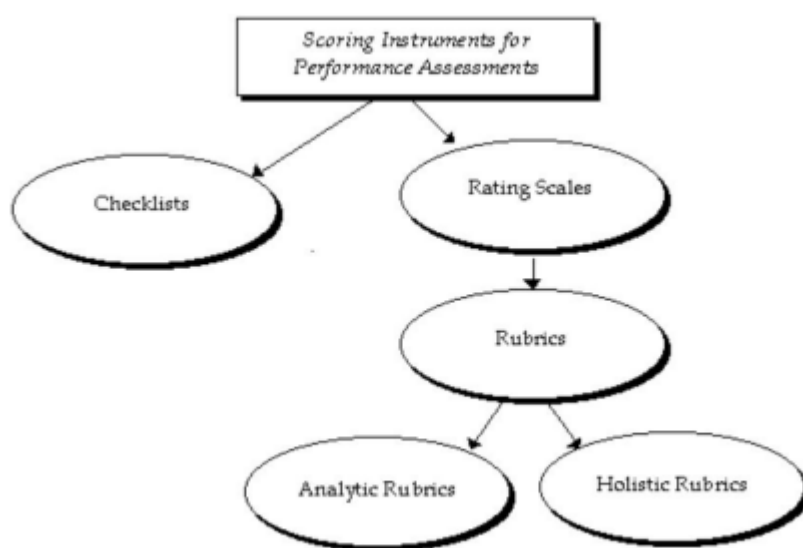
\* Öğr. Gör. Dr., ORCID ID: 0000-0002-5367-6608, Ankara Yıldırım Beyazıt Üniversitesi, DİLMER, aybegumce@gmail.com

## Introduction

The lack of standardization in the field of teaching Turkish as a foreign/second language causes various issues in practical applications. This problem affects the entire teaching process from beginning to end and leads to differences in practices among teaching centers and instructors. One of the areas where this lack of standardization is commonly observed is in assessment and evaluation. The inconsistency in determining the levels of learners completing Turkish courses indicates this problem. Similarly, inconsistent evaluations by different evaluators within the same institution highlight the issue of objective assessment. To overcome these issues in assessment and evaluation, it is necessary to establish and implement standards from specific to general.

Objective and consistent assessment is crucial for determining whether the aims and objectives of teaching are being achieved, for the contribution of teaching outcomes to the process, for making teaching more efficient, and for identifying the strengths and weaknesses of learners. Some assessment tools exist to ensure objective and consistent evaluation. Rating scales and checklists are widely used to prevent inconsistencies in the assessment of learner performances (Özen, 2025, p. 75)

**Figure 1.** Performance Evaluation Classification (Mertler, 2001, p. 1)



Rubrics are long-term tools that have accompanied education systems for decades and are now widely used at all levels of education (Andrade, 2023). They limit evaluators in assessments and contribute to the consistent evaluation of learner performances and to obtain diagnostic information about a skill/performance (Durmuş, 2022, p. 22). Providing feedback is important to increase learner motivation in assessment (Hyland, 2003). Thanks to the diagnostic information obtained through rubrics, it becomes easier for evaluators to identify learners' successes, weaknesses, and strengths (Yokouchi, 2025) and the opportunity to provide feedback with diagnostic information is obtained.

To date, criteria have shown their usefulness as assessment tools and have supported instructors in making more consistent scores (Mahdi & Alkhateeb, 2025, p. 2). Rubrics play an important role in shaping educational evaluation and provide clear criteria for both teaching and learning (Panadero et al., 2024, p. 31). The use of rubrics by all evaluators in language teaching centers will contribute to ensuring consistency among evaluators. Similarly, rubrics based on certain basic standards will also support consistency and objectivity in evaluations made in different centers.

As can be seen in Figure 1, there are two types of rubrics: holistic and analytical. Holistic rubrics provide a single score and a general evaluation. Holistic scoring according to this assessment tool is controversial as it involves giving a single score to the entire text and combining the measurement and evaluation of different features that make up the written expression (Mahdi & Alkhateeb, 2025, p. 3). Analytical rubrics, on the other hand, provide the opportunity to evaluate the learner's performance in detail and by dividing it into sections and limit the evaluator more, thus allowing for more consistent results in the evaluation and additional feedback (Weigle, 2002). Especially in cases where a well-structured rubric is not used, the evaluations made may remain subjective and superficial.

Rubrics are especially used in the assessment of productive skills. Writing skills are generally assessed based on learners' performance in writing independent texts, making objective evaluation challenging. Due to the multidimensional nature of writing, which includes various elements, rubrics are important for fair and valid assessment of written production skills. The writing process, involving planning, drafting, and revising, culminates in the final draft (Harmer, 2004, p. 5). Therefore, writing has a multidimensional structure, encompassing elements from title to conclusion and from punctuation to style (Aksak Gönül & Ekti, 2022, p. 51). From the perspective of language teaching, grammar, vocabulary, punctuation, organization, and other aspects must be considered. According to a study by Lee (2010), most instructors focus on spelling and grammar when evaluating writing performance. Instead of focusing solely on linguistic structures, the communicative aspects of writing, language use as required by communication with the reader, and the purposes achieved in writing come to the fore (Yaylı, 2015, p. 303).

The Common European Framework of Reference for Languages (CEFR) suggests the following criteria for assessing language proficiency and error in writing:

- Pronunciation errors;
- Spelling (punctuation) errors;
- Vocabulary errors;
- Grammatical errors;
- Syntactic errors;
- Sociolinguistic and sociocultural errors (Council of Europe, 2013, p. 155).

Similarly, the CEFR Companion Volume recommends the following dimensions for assessing written production skills: general language/vocabulary, coherence, accuracy,

description, and argumentation (Council of Europe, 2021, p. 191). Analytic rubrics allow for the measurement of different aspects of the text produced by the candidate, such as organization, lexical resources, grammar, coherence, and cohesion (Viñas, 2022). Wiggins (1993) states that rubrics should reflect assessment criteria that support learners' more authentic, outcome-focused, and contextual writing skills. These skills include discourse awareness, language skills, goal setting, sociocultural knowledge, and memory management strategies (Flower & Hayes, 1980; Kellogg & Whiteford, 2009). Therefore, when the multidimensional structure of writing is not considered, and the focus is solely on spelling and grammar, the evaluation of writing performance will lack components such as vocabulary and choice, organization of the text, textuality criteria, and the communicative functions of language as an action.

There is a lack of commonly used shared rubrics in the field of teaching Turkish as a foreign language. Some language teaching centers use their own developed holistic or analytic rubrics that have not been tested for validity and reliability, while some instructors conduct evaluations without rubrics. In a study conducted with 167 instructors by Göktaş and Aksak Gönül (2022, p. 142), only 35.93% of the participants reported using rubrics in writing assessments. The low use of rubrics raises questions about the objectivity of evaluations. Therefore, it is essential for instructors to have the skills to recognize the components of good writing and be familiar with the elements of textuality to include them in assessments. One of the most important ways to prevent these issues and achieve consistent results is by using comprehensive rubrics. Additionally, rubrics serve as a guide for instructors to analyze and interpret data obtained from assessments. Due to these features, the study aims to develop analytic rubrics for the objective assessment of writing skills in the teaching of Turkish as a foreign/second language.

The effectiveness of rubrics is substantially directly proportional to the instructors' commitment to rubrics (Jeong, 2015). Optimizing and appropriating use of rubrics involves the continuous professional development of evaluators in rubric design and application (Mace & Pearl, 2021). Considering that evaluator guides for rubrics should be created for this commitment and development, it is also aimed to prepare evaluator guides for the prepared rubrics.

### **Research Objective and Sub-objectives**

In the field of teaching Turkish as a foreign/second language, there are different competencies expected of learners at each level. Therefore, the expected teaching outcomes and the quality of learner performances vary at each level. While short, simple texts are expected at the basic level, longer and more detailed texts are expected at intermediate and advanced levels. As a result, starting from the intermediate level, evaluations of written texts should consider various aspects beyond spelling and grammar errors. For this reason, separate analytic rubrics should be developed for each level to ensure accurate assessment. The study aims to develop separate analytic rubrics for B1 and B2 levels, as these levels are transition levels that are difficult to distinguish from each other and represent the levels

where metacognitive skills begin to come into play. The use of separate analytic rubrics for each level is necessary for assessing the different elements of writing. As a sub-objective of this research, an attempt was made to determine whether experience influences evaluations conducted using rubrics. Another sub-objective is to examine whether there is a significant difference in evaluations based on rubrics between evaluators teaching in multilingual classrooms in Turkey, where the target language is the native language, and those teaching in monolingual classrooms abroad, where the target language is not the native language.

### **Method**

The stages of rubric development involve both qualitative and quantitative applications, therefore a mixed-method was applied in the study. The stages of rubric development are as follows:

1. Determining the learning objectives to be evaluated
2. Identifying the expected competencies and undesired features from products, processes, or performances
3. Determining dimensions and defining criteria
4. Deciding the type of rubric and defining the descriptors for the highest and lowest performance levels for each criterion
5. Defining the descriptors for the middle levels of criteria
6. Conducting sample learner performance assessments
7. Reviewing the rubric (Mertler, 2001).

Since these stages are compatible, an exploratory sequential design was applied. In this design, findings obtained from qualitative data are followed by the creation of the quantitative phase and the collection of quantitative data. The qualitative phase can be used to determine the measurement tool or variables to be applied in the quantitative phase (Creswell, 2017).

In the qualitative part of the research, a literature review was conducted first. In the relevant literature, the rubric developed within the scope of the research article titled “Development Study of B1 Level Writing Skills Rubric for Turkish Learners as a Foreign Language” by Yorgancı and Baş (2021) includes four dimensions and twenty criteria. The dimensions are “Organization and Content, Language and Expression, Vocabulary, Spelling and Grammar”. The dimensions of vocabulary, spelling, and grammar overlap with the dimensions of the research. Within the scope of the research, 30 writing performances were scored by 5 raters.

In the research article titled “Writing Skills Rubric for Narrative/Demonstrative Texts in Teaching Turkish as a Foreign Language” by İltar and Karataş (2022), it was aimed to develop a genre-specific rubric. It differs from the study in terms of being genre-specific, but some items were also included in the evaluator guides prepared within the scope of the study to be used when evaluating narrative texts. There is no statement in the study that includes any level distinction.

A rubric was developed in the research article titled “Development of Analytical Rubric for A2 Level Writing Skills in Teaching Turkish as a Foreign Language” conducted by Tekin and Sallabaş (2022). The criteria and criterion descriptors are quite superficial according to the study. The main reason for the differences between the two studies can be shown as

being prepared at different levels. 26 writing performances were evaluated by 5 evaluators in the study.

The rubric developed in the research article titled “Development of A2 Level Writing Skills Rubric for Those Who Teach and Learn Turkish as a Foreign Language” conducted by Harmankaya, Sallabaş and Toker (2022) is also prepared for the A2 level, so it differs from the study in terms of size and criteria.

When the studies conducted in the literature are examined, it is seen that the studies differ from the levels addressed in the relevant study and the dimensions and criteria included in the rubric. The intermediate level B1 and B2 levels are related to each other and at the same time have different outcomes. Therefore, it is thought that it would be more useful for the rubrics to be developed for these levels to be complementary to each other and to be created together. Sample rubrics used in various Turkish teaching centers and institutions in the field of teaching Turkish as a foreign language and CEFR (Council of Europe, 2013, 2021) achievements were examined. After the examinations, an item pool was created. The dimensions to be included in the rubric to be prepared based on the item pool were determined, and the criteria were selected. The dimensions to be included in the B1 and B2 level analytical rubrics were determined as grammar, vocabulary, spelling and punctuation, and organization. In addition, for the criteria to be included in the grammar dimension, Gazi University TÖMER, Yedi İklim Türkçe, Yeni İstanbul and İzmir teaching sets, which are widely used in the field of teaching Turkish as a foreign language, were examined and common grammar structures were determined. After all these scans, the criteria for the grammar dimension of the rubric were as follows: The criteria for the vocabulary dimension were determined as content-oriented word selection and function-oriented word selection; and in the criteria for the spelling and punctuation dimension, spelling and punctuation were included as two separate criteria. In the organization section, the text plan, content, and expression sub-criteria were selected.

The developed rubrics were designed to be four-tiered with "4 (excellent), 3 (good), 2 (needs improvement), 1 (insufficient)." The rubric, consisting of 25 items and four tiers, can be scored on a scale of 100 and 25 points (when scored as 1, 0.75, 0.50, and 0.25). Descriptors for each criterion were written to guide evaluators on which score to assign in different situations. To make the rubrics more effective and efficient for evaluators, assessment guides were prepared to ensure that all criteria of the rubrics are understood correctly and consistently by the assessors and that the rubrics are used more easily and effectively.

After the rubric was developed, expert opinions were obtained to determine the content validity. Two of the expert opinion forms were filled out by individuals who had completed their doctoral education in the field, and five were filled out by individuals who were continuing their doctorate and had at least 10 years of experience in the field. The validity indexes of the items were calculated using the expert opinion forms obtained through the application of the Davis (1992) technique. The final form of the rubric was given by taking into account expert opinions.

To determine the reliability of the rubrics, six B1 and six B2 learner writing performances were evaluated by 20 experienced and inexperienced teachers of Turkish as a foreign/second language. The Kendall's W concordance coefficient was calculated to measure the agreement among evaluators.



Additionally, the Kruskal-Wallis test was conducted to determine whether there was a significant difference in the scoring based on evaluator experience. Evaluators were divided into three groups: 0-5 years, 6-9 years, and over 10 years of experience. Furthermore, the Mann-Whitney U test was conducted to see if there was a difference between the scores given by Turkish teachers teaching as a foreign language abroad and those teaching as a second language domestically.

## Publication Ethics

**Committee Name:** Hacettepe University Ethics Committee

**Decision Date:** March 31, 2023

**Number:** E-35853172-000-00002773855

## Participants

The study involves three different participant groups: field experts, learner performances, and evaluators. Information regarding the field experts consulted for determining the validity index of the study is provided below.

**Table 1.** Information About Experts

Expert	Field Experience	Education	Title
Expert 1	10 years	PhD graduate	Assoc. Prof.
Expert 2	10 years	PhD graduate	Dr. Lecturer
Expert 3	10 years	Pursuing PhD	Lecturer
Expert 4	12 years	Pursuing PhD	Lecturer
Expert 5	12 years	Pursuing PhD	Lecturer
Expert 6	10 years	Pursuing PhD	Lecturer
Expert 7	11 years	Pursuing PhD	Lecturer

In selecting the field experts, the prerequisites were “having at least 10 years of experience in teaching Turkish as a foreign/second language” and “having or pursuing a PhD in the field.” Criterion sampling, one of the purposive sampling methods, was chosen for selecting the experts. Criterion sampling involves selecting a sample based on predetermined criteria set by the researcher (Yıldırım & Şimşek, 2021, p. 120).

For the writing performances of learners, 12 performances (6 at B1 level and 6 at B2 level) were selected from the writing exams of students studying at Gazi University TÖMER. Extreme case sampling, a purposive sampling method, was used to select 6 writing performances each that were considered weak, average, and good. It is believed that using extreme case sampling allows the researcher to better observe the variability related to the problem being studied (Koç Başaran, 2017, p. 490).

To determine the reliability coefficient of the study, 20 evaluators, both experienced and inexperienced in the field, were asked to evaluate the selected learner performances. Since determining whether there are differences in the scoring of experienced and inexperienced evaluators is one of the sub-objectives of the research, no training was provided to the evaluators and evaluator guides were shared. However, it is the evaluator's responsibility to read or not read these guides. The evaluators were chosen using the maximum variation sampling method, which aims to obtain experiences from different stakeholders in various

contexts (Yağar & Dökme, 2018, p. 5). Selection is based on different qualifications and criteria (Patton, 2018, p. 235). The evaluators, a mix of experienced and inexperienced, were chosen from those teaching at different levels both domestically and abroad. Information about the evaluators is provided in the table below.

**Table 2.** Information About Evaluators

Evaluator	Gender	Education Level	Major	Teaching Location	Years of Experience
Evaluator 1	Female	Bachelor's	Turkish Education	Domestic	12 years
Evaluator 2	Female	Pursuing PhD	Turkish Education	Domestic	12 years
Evaluator 3	Female	Master's	Turkish Education	Domestic	5 years
Evaluator 4	Female	Pursuing PhD	Urdu Language and Literature	Abroad	10 years
Evaluator 5	Female	Pursuing PhD	Turkish Education	Abroad	7 years
Evaluator 6	Female	Master's	Contemporary Turkish Dialects and Literature	Domestic	8 years
Evaluator 7	Female	PhD	Turkish Education	Abroad	6 years
Evaluator 8	Female	Master's	Turkish Education	Domestic	3 years
Evaluator 9	Female	Pursuing Master's	Turkish Language and Literature	Abroad	7 years
Evaluator 10	Female	Pursuing PhD	English Language and Literature	Domestic	3 years
Evaluator 11	Female	Pursuing PhD	Turkish Education	Abroad	11 years
Evaluator 12	Male	Master's	Turkish Education	Abroad	10 years
Evaluator 13	Male	Master's	Turkish Language and Literature	Domestic	5 years
Evaluator 14	Male	Master's	Turkish Education	Abroad	9 years
Evaluator 15	Male	PhD	Turkish Language and Literature	Domestic	5 years
Evaluator 16	Male	Pursuing PhD	Turkish Education	Domestic	6 months
Evaluator 17	Female	Pursuing PhD	Turkish Language and Literature	Domestic	7 years
Evaluator 18	Female	Pursuing PhD	Translation and Interpreting	Domestic	8 years
Evaluator 19	Female	Pursuing PhD	Turkish Language and Literature	Domestic	11 years
Evaluator 20	Female	Pursuing PhD	Turkish Education	Domestic	10 years

### Data Analysis and Collection

Following the stages of the exploratory sequential design, qualitative data were first obtained, and then quantitative data were derived from the qualitative findings.



Qualitative data were collected, and quantitative data were obtained with the B1 and B2 level analytic rubrics prepared using these qualitative data.

In the qualitative phase of the study, expert opinions were sought to determine the content validity indices of the developed rubrics. Reviewers with expertise in structural aspects of tool design, such as developing and using rating scales or matching tool items with statistical measures for testing the study hypothesis, can also provide valuable contributions (Davis, 1992). Data were collected using a form developed with the Davis (1992) technique to reach expert opinions.

In the quantitative phase of the study, 20 evaluators voluntarily participated in evaluating the learner performances with the developed rubrics, and the data obtained were analyzed using the Kendall-W technique. To measure the effect of experience on the results obtained from the same data, Kruskal Wallis H analysis was conducted, and Mann Whitney U analysis was performed to determine whether there was a difference between the scores given by those teaching Turkish as a foreign language and those teaching it as a second language. All the data were analyzed using SPSS 27 software.

## Findings

### 1. Findings on Content Validity:

To test the validity of the analytic rubrics developed for B1 and B2 levels, expert opinions were obtained from seven instructors with at least 10 years of experience using the Davis technique. Based on the data obtained, it was determined that the rubrics were valid.

**Table 3.** Content Validity Index for the B1 Level Rubric

Dimensions	Criteria	Sub-Criteria	A	B	C	D	VI
Grammar	Tense Usage	-	7	-	-	-	1
Voice Usage	-	6	1	-	-	1	
Compound Sentences	-	6	1	-	-	1	
Verbal Nouns	-	6	1	-	-	1	
Syntax	-	6	-	1	-	0.85	
Appropriateness to Level	-	7	-	-	-	1	
Vocabulary	Content-Oriented Vocabulary Choice	-	7	-	-	-	1
Function-Oriented Vocabulary Choice	-	7	-	-	-	1	
Spelling and Punctuation	Spelling	-	7	-	-	-	1
Capitalization	-	7	-	-	-	1	
Punctuation	-	6	1	-	-	1	
Organization	Text Plan	Title	6	1	-	-	1
Introduction	7	-	-	-	1		
Development	7	-	-	-	1		
Conclusion	6	1	-	-	1		
Formal Features	6	1	-	-	1		
Word Count	7	-	-	-	1		

Dimensions	Criteria	Sub-Criteria	A	B	C	D	VI
Content	Main Idea and Supporting Ideas	7	-	-	-	1	
Development of Ideas	7	-	-	-	1		
Expression	Language of the Text	7	-	-	-	1	
Coherence	7	-	-	-	1		
Genre and Sociolinguistic Appropriateness	6	1	-	-	1		
<b>Total</b>	-	-	-	-	-	-	.994

According to Davis (1992), the content validity index should be .80 or above. Based on expert opinions for the B1 level, the total content validity index was determined to be 0.994, with item indices ranging from .85 to 1. Therefore, it can be said that the B1 level analytic rubric developed is valid.

**Table 4.** Content Validity Index for the B2 Level Rubric

Dimensions	Criteria	Sub-Criteria	A	B	C	D	VI
Grammar	Compound Tenses and Voice Usage	-	7	-	-	-	1
Indirect Speech	-	7	-	-	-	1	
Compound Sentences	-	6	1	-	-	1	
Verbal Nouns	-	6	1	-	-	1	
Syntax	-	6	1	-	-	1	
Appropriateness to Level	-	7	-	-	-	1	
Vocabulary	Content-Oriented Vocabulary Choice	-	7	-	-	-	1
Function-Oriented Vocabulary Choice		Metaphors	-	6	1	-	-
Incorrect Words	-	7	-	-	-	1	
Conjunctions and Prepositions	-	6	-	1	-	0.85	
Spelling and Punctuation	Spelling	-	6	1	-	-	1
Punctuation	-	7	-	-	-	1	
Organization	Text Plan	Title	7	-	-	-	1
Introduction	7	-	-	-	1		
Development	7	-	-	-	1		
Conclusion	7	-	-	-	1		
Formal Features	7	-	-	-	1		
Word Count	7	-	-	-	1		
Content	Main Idea and Supporting Ideas	6	-	1	-	0.85	
Development of Ideas	7	-	-	-	1		
Expression	Language of the Text	6	1	-	-	1	
Coherence	7	-	-	-	1		
Repetition of Meaning	7	-	-	-	1		
Genre and Sociolinguistic Appropriateness	7	-	-	-	1		
<b>Total</b>	-	-	-	-	-	-	.988

Based on expert opinions for the B2 level, the total content validity index was determined to be .988, with item indices ranging from .85 to 1. Therefore, it can be said that the B2 level analytic rubric developed is valid.

## 2. Findings on Reliability:

For the reliability tests of the developed rubrics, six B1 and six B2 learner writing performances from exams conducted at Gazi University TÖMER were evaluated by 20 instructors, and the data obtained were analyzed using the Kendall's W technique. The whole rubric and dimension-specific Kendall's W coefficients for the B1 level are presented below.

**Table 5.** Kendall-W Concordance Coefficient for B1 Level

Level	W	P
All Criteria	.856	0.000

**Table 6.** Kendall-W Concordance Coefficients for B1 Level Dimensions

Dimensions	W	P
Grammar	.815	0.000
Vocabulary	.679	0.000
Spelling and Punctuation	.714	0.000
Organization	.766	0.000

For the B1 level, Kendall-W concordance coefficient for the all six writing performances evaluated by 20 evaluators using the 25-item analytic rubric was found to be .856, with a P value of 0.000. These data indicate that the B1 level analytic rubric has strong agreement and significant results.

The Kendall-W coefficients for the B1 level dimensions are as follows: Grammar dimension .815 (strong agreement), Vocabulary .679 (moderate agreement), Spelling and Punctuation .714 (strong agreement), Organization .766 (strong agreement). Therefore, it can be said that the analytic rubric also has strong agreement at the dimension level, with all P values showing significant results.

**Table 7.** Kendall-W Concordance Coefficient for B2 Level

Level	W	P
All Criteria General	.829	0.000

For the B2 level, Kendall-W concordance coefficient for the all six writing performances evaluated by 20 evaluators using the 25-item analytic rubric was found to be .829, with a P value of 0.000. These data indicate that the B2 level analytic rubric has strong agreement and significant results.

**Table 8.** Kendall-W Concordance Coefficients for B2 Level Dimensions

Dimensions	W	P
Grammar	.775	0.000
Vocabulary	.772	0.000
Spelling and Punctuation	.621	0.000
Organization	.817	0.000

The Kendall-W coefficients for the B2 level dimensions are as follows: grammar dimension .775 (strong agreement), vocabulary .772 (strong agreement), spelling and punctuation .621 (moderate agreement), organization .817 (strong agreement). Therefore, it can be said that the analytic rubric also has strong agreement at the dimension level, with all P values showing significant results. Based on these data, it can be concluded that the analytic rubrics developed for the B1 and B2 levels are valid, reliable, and ready for use.

### 3. Effect of Experience on Evaluation Results:

Additionally, the data were analyzed using the Kruskal-Wallis technique to determine whether evaluator experience had any impact on scoring, and it was found that there was no significant difference. Evaluator experience was divided into three groups: 0-5 years, 6-9 years, and over 10 years. For each level, 6 evaluators with 0-5 years of experience, 7 with 6-9 years of experience, and 7 with over 10 years of experience were compared. In the Kruskal-Wallis technique, the P value is first examined to determine the significance of the data; if  $P < 0.05$ , there is a significant difference in the group mean ranks (Cevahir, 2020, p. 54). The data related to the analysis results are presented below.

**Table 9.** Effect of Experience on Evaluation Results B1 Level Kruskal-Wallis H Analysis

Writing Paper	X2	P
B1.1	1.258	.553
B1.2	.621	.733
B1.3	1.790	.409
B1.4	1.559	.459
B1.5	4.392	.111
B1.6	.112	.945

In the data obtained for the B1 level learner performances, Kruskal-Wallis H analysis was conducted on the scores given by 6 evaluators with 0-5 years, 7 with 6-9 years, and 7 with over 10 years of experience using the analytic rubric. The P values were as follows: .553 for B1.1, .733 for B1.2, .409 for B1.3, .459 for B1.4, .111 for B1.5, and .945 for B1.6. Since all values were greater than 0.05, it was concluded that there was no significant difference in the scores given by evaluators with different experience levels using the analytic rubric for B1 level, so the X2 data did not need to be considered.

**Table 10.** Effect of Experience on Evaluation Results B2 Level Kruskal-Wallis H Analysis

Writing Paper	X2	P
B2.1	2.403	.301
B2.2	5.431	.066
B2.3	.419	.811

Writing Paper	X2	P
B2.4	1.754	.416
B2.5	.544	.762
B2.6	2.005	.367

In the data for the B2 level learner performances, Kruskal-Wallis H analysis was conducted on the scores given by 6 evaluators with 0-5 years, 7 with 6-9 years, and 7 with over 10 years of experience using the analytic rubric. The P values were as follows: .301 for B2.1, .066 for B2.2, .811 for B2.3, .416 for B2.4, .762 for B2.5, and .367 for B2.6. Since all values were greater than 0.05, it was concluded that there was no significant difference in the scores given by evaluators with different experience levels using the analytic rubric for B2 level, so the X2 data did not need to be considered.

#### 4. Findings on Differences in Evaluation Results Between Foreign and Second Language Teachers

Within the scope of the study, it was aimed to determine whether there were any differences in the evaluations conducted with the analytic rubric between Turkish teachers teaching as a foreign language (abroad) and as a second language (domestically). For this purpose, the scores given to learner performances by evaluators in these two groups were analyzed using the Mann-Whitney U test. The Mann-Whitney U test compares the mean ranks of two independent/unrelated groups and determines whether there is a statistically significant difference between the scores of the two groups (Büyüköztürk vd., 2020, p. 187; Cevahir, 2020, p. 40).

To determine the significance of the analysis data, the P value is first examined, and if there is a significant difference, the U value is considered. If  $P > 0.05$ , the H0 hypothesis is accepted, indicating that there is no significant difference in the mean ranks of the groups; if  $P < 0.05$ , the H1 hypothesis is accepted, indicating that there is a significant difference in the mean ranks of the groups (Cevahir 2020, p. 54). The findings related to the analysis are presented in the tables below.

**Table 11.** Differences in Evaluation Results Between Foreign and Second Language Teachers B1 Level Mann-Whitney U Analysis

Writing Paper	U	P
B1.1	45.000	1.000
B1.2	24.500	.097
B1.3	61.500	.211
B1.4	50.000	.757
B1.5	49.000	.817
B1.6	66.000	.115

In the data obtained for the B1 level learner performances, the Mann-Whitney U analysis was conducted on the scores given by 7 foreign language Turkish teachers and 13 second language Turkish teachers. The P values were as follows: 1.000 for B1.1, .097 for B1.2, .211 for B1.3, .757 for B1.4, .817 for B1.5, and .115 for B1.6. Since all values were greater than 0.05, it was concluded that there was no significant difference in the evaluations conducted

with the analytic rubric for B1 level between teachers teaching abroad and domestically, so the U data did not need to be considered.

**Table 12.** Differences in Evaluation Results Between Foreign and Second Language Teachers B2 Level Mann-Whitney U Analysis

Writing Paper	U	P
B2.1	38.000	.588
B2.2	54.500	.485
B2.3	57.500	.351
B2.4	59.500	.275
B2.5	53.500	.536
B2.6	59.500	.275

In the data for the B2 level learner performances, the Mann-Whitney U analysis was conducted on the scores given by 7 foreign language Turkish teachers and 13 second language Turkish teachers. The P values were as follows: .588 for B2.1, .485 for B2.2, .351 for B2.3, .275 for B2.4, .536 for B2.5, and .275 for B2.6. Since all values were greater than 0.05, it was concluded that there was no significant difference in the evaluations conducted with the analytic rubric for B2 level between teachers teaching abroad and domestically, so the U data did not need to be considered.

### Discussion and Conclusion

One of the main issues in the field of teaching Turkish as a foreign/second language is the lack of standardization. This problem is particularly evident in assessment and evaluation studies, which reveal the learning outcomes. Differences are observed in assessment and evaluation practices and results among Turkish teaching centers, as well as differences in scoring by evaluators within the same teaching center. Therefore, having common standards in measurement and evaluation criteria is important for obtaining objective results and achieving standardization. Rubrics play an important role in standardization in the field of evaluation. According to the results of a master's research titled "Examination of Rater Reliability When Using Scoring Guides (Rubrics)" prepared by Atmaz (2009, pp. 77–78), evaluators using rubrics achieved more consistent results compared to those not using rubrics. Therefore, the use of rubrics in evaluation leads to more objective results, highlighting the importance of rubrics in language teaching centers and making them a subject of academic research.

The evaluations conducted with 20 evaluators using the prepared B1 and B2 level rubrics yielded consistent results. Based on the findings, which are limited to twenty evaluators and a total of twelve writing performances at the B1 and B2 levels, it was determined that the developed rubrics exhibit a strong capacity for objective assessment. The number of evaluators involved is one of the distinguishing features of this study compared to other research with similar content. Evaluator experience in performance assessment is a key factor influencing the evaluation process. In the absence of a standardized assessment tool, differences may arise between the judgments of experienced and inexperienced evaluators. However, the findings indicate that when using the prepared rubrics, there was no significant difference between the evaluations of experienced and inexperienced evaluators. This suggests that rubrics can be effectively utilized by inexperienced evaluators and play a



crucial role in minimizing subjectivity. Furthermore, it is believed that the evaluator guides accompanying the rubrics are particularly effective in assisting inexperienced evaluators. Additionally, the study found no significant difference between the evaluations of instructors teaching the target language as a foreign language abroad, in monolingual classroom settings, and those teaching it as a second language in Turkey, in multilingual classrooms where the target language is spoken as a native language. This finding further supports the validity and reliability of rubrics across different instructional contexts. Another notable aspect of this study is the inclusion of 25 detailed assessment criteria within the rubrics, along with the development of evaluation guides to ensure a consistent and accurate understanding of the criteria and descriptors by all evaluators. A review of the literature revealed no similar guide in previous rubric development studies. It is believed that these evaluation guides significantly enhance the reliability of the rubrics, enabling instructors to objectively assess learner performance and, when necessary, refine their teaching strategies.

Another aspect that distinguishes this study is the inclusion of 25 items in the rubrics, being quite detailed, and the preparation of evaluation guides to ensure that the criteria and descriptors are correctly understood and applied by all evaluators. Literature reviews did not find a similar guide in other rubric development studies. It is believed that the evaluation guides prepared significantly enhance the reliability of the rubrics. This can assist instructors in objectively measuring learner achievements and, when necessary, updating teaching strategies.

Writing is not only a skill area but also one of the best indicators of a learner's language proficiency. Therefore, assessment and evaluation studies need to analyze the learner's language proficiency comprehensively. The valid and reliable analytic rubrics prepared for B1-B2 levels enable the determination of the extent to which learners have mastered the expected achievements at these levels. In this respect, rubrics can be used as an effective tool to enhance learner performance and improve instructor evaluations. It is believed that these results will provide valuable contributions to educators and researchers aiming to enrich and develop the practice of teaching Turkish.

### **Recommendations**

In addition to rubrics prepared according to levels, it is important to develop different rubrics for task and type-focused applications. For example, narrative texts and informative texts contain different features. Therefore, rubrics to be used in evaluating these types may contain different dimensions and criteria. Hence, in the study in question, which focuses more on evaluating informative texts, the characteristics of narrative texts were also taken into account and descriptors were included in the evaluator guides regarding the replacement of some criteria with narrative text criteria, and an attempt was made to ensure that the analytical rubrics prepared could be used for both types. However, in order to be useful, valid and reliable, it would be beneficial to prepare separate graded scoring according to types, and to conduct validity-reliability tests regarding rubrics. In exam and in-class activity applications, students are made to write for different tasks such as writing e-mails, postcards, writing texts based on graphics or pictures. For this reason, it is thought that it would be useful to prepare and apply different rubrics that include criteria specific to certain tasks for task-based applications.

One of the most important issues encountered in the application of rubrics is whether evaluators use the rubrics carefully and meticulously. To ensure that the criteria in the

rubrics are understood similarly and correctly by evaluators, evaluation guides were prepared in addition to the criterion descriptors. However, there is no definitive information on whether these guides and rubrics were read carefully. Evaluators should be made more aware of using rubrics.

It is believed that planning in-service training programs focused on measurement and evaluation practices would be beneficial. These programs should not only cover topics such as measurement criteria, exam preparation and administration, and evaluation standards, but also include content on the advantages of using rubrics, their development, and strategies for utilizing them more effectively in the evaluation process.

Analytic rubrics can help identify the dimensions and criteria where errors are concentrated among learners with different native languages or characteristics (age, sociocultural differences, foreign language, second language, monolingual, multilingual, etc.), providing a basis for instructors to classify learners and errors and develop applications according to the target audience, as well as laying the groundwork for academic studies.

### References

- Aksak Gönül, A., ve Ekti, M. (2022). Yabancı dil olarak Türkçe öğretiminde yazma becerisini değerlendirici ölçme araçlarının hazırlanması. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, 31(31), 47–60. <https://doi.org/10.29000/rumelide.1220228>
- Andrade, H. L. (2023). What is next for rubrics? A reflection on where we are and where to go from here. C. Gonsalves (Ed.), *Improving Learning Through Assessment Rubrics: Student Awareness of What and How They Learn* içinde (13. baskı, ss. 314–326). King's Academy, King's College London, UK. <https://doi.org/10.4018/978-1-6684-6086-3.ch017>
- Atmaz, G. (2009). *Puanlama yönergesi (rubrik) kullanılması durumunda puanlayıcı güvenilirliğinin incelenmesi*. [Yayımlanmamış yüksek lisans tezi]. Mersin Üniversitesi.
- Büyüköztürk, Ş., Çokluk, Ö. ve Köklü, N. (2020). *Sosyal bilimler için istatistik*. Ankara: Pegem Akademi.
- Cevahir, E. (2020). *SPSS ile nicel veri analizi rehberi*. İzmir: Kibele Yayınları.
- Council of Europe. (2013). *Common European framework of reference for languages*. Cambridge: Cambridge University Press.
- Council of Europe. (2021). *Common European framework of reference for languages complementary volume*. Cambridge: Cambridge University Press.
- Creswell, J. W. (2017). *Araştırma deseni: nitel, nicel ve karma yöntem yaklaşımları* (S. B. Demir (çev. ed.). Ankara: Eğiten Kitap.
- Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Clinical Methods Research*, 5(4), 194–197.
- Durmuş, A. O. (2022). *Dereceli puanlama Anahtarına Dayalı Geribildirim Türkçenin Yabancı Dil Olarak Öğretiminde Yazma Becerisinin Geliştirilmesine Etkisi*. [Yayımlanmamış doktora tezi]. Ankara Üniversitesi.
- Flower, L. S. ve Hayes, J. R. (1980). The dynamics of composing: making plans and juggling constraints. L. W. Gregg ve E. R. Steinberg (Ed.), *Cognitive processes in writing* içinde. New Jersey: Lawrence Erlbaum.
- Göktaş, B. ve Aksak Gönül, A. (2022). Yabancı dil olarak Türkçe öğreticilerinin yazma becerisinde geri bildirim alışkanlıkları. *Uluslararası Türkçe Öğretimi Araştırmaları Dergisi*, 2(2), 124–147.

- Harmankaya, M. Ö., Sallabaş, M. E., ve Toker, T. (2022). Türkçeyi yabancı dil olarak öğreten ve öğrenenler için a2 düzeyi yazma becerisi dereceli puanlama anahtarı geliştirme çalışması. *IBAD Sosyal Bilimler Dergisi*, (13), 511–524. <https://doi.org/10.21733/ibad.1212818>
- Harmer, J. (2004). *How to teaching writing*. London: Pearson Longman.
- Hyland, F. (2003). Focusing on form: student engagement with teacher feedback. *System*, 31(2), 217–230. [https://doi.org/10.1016/S0346-251X\(03\)00021-6](https://doi.org/10.1016/S0346-251X(03)00021-6)
- Jeong, H. (2015). Rubrics in the classroom : do teachers really follow them? *Language Testing in Asia*, 5(6), 2-14. <https://doi.org/10.1186/s40468-015-0013-5>
- Kellogg, R. T., ve Whiteford, A. P. (2009). Training advanced writing skills : the case for deliberate practice training advanced writing skills: the case for deliberate practice. *Educational Psychologist*, 44(4), 250–266. <https://doi.org/10.1080/00461520903213600>
- Koç Başaran, Y. (2017). Sosyal bilimlerde örnekleme kuramı. *Akademik Sosyal Araştırmalar Dergisi*, 5(47), 480–495.
- Latif, İ., ve Karataş, A. G. (2022). Türkçenin yabancı dil olarak öğretiminde anlatmaya/göstermeye dayalı metinler için yazma becerisi dereceli puanlama anahtarı. *Okuma Yazma Eğitimi Araştırmaları*, 10(2), 194–213. <https://doi.org/https://doi.org/10.35233/oyea.1177730>
- Lee, I. (2010). Writing teacher education and teacher learning : testimonies of four EFL teachers. *Journal of Second Language Writing*, 19(3), 143–157. <https://doi.org/10.1016/j.jslw.2010.05.001>
- Mace, M. K., ve Pearl, D. (2021). Rubric development and validation for assessing comprehensive internationalization in higher education. *Journal of Studies in International Education*, 25(1), 51–65. <https://doi.org/10.1177/1028315319865790>
- Mahdi, H. S., ve Alkhateeb, A. (2025). Revolutionising essay evaluation: a cutting-edge rubric for ai-assisted writing. *International Journal of Computer-Assisted Language Learning and Teaching*, 15(1), 1–19. <https://doi.org/10.4018/IJCALLT.368226>
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research, and Evaluation*, 7(25), 1–8.
- Özen, O. (2025). Türkçe dersi uygulama sınavlarının öğrencilerin konuşma ve dinleme becerilerine etkileri. *International Symposium On Measurement, Selection And Placement*, January, 71–85.
- Panadero, E., Fernández Ortube, A., Krebs, R. ve Roelle, J. (2024). Analysis of online rubric platforms: advancing toward erubrics. *Assessment and Evaluation in Higher Education*, 50(1), 31–49. <https://doi.org/10.1080/02602938.2024.2345657>
- Patton, M. Q. (2018). *Nitel araştırma ve değerlendirme yöntemleri*. Ankara: Pegem Akademi.
- Tekin, E., ve Sallabaş, M. E. (2022). Yabancı dil olarak Türkçe öğretiminde a2 seviyesi yazma becerisine yönelik analitik rubrik geliştirme. *ANASAY*, 6(22), 81–95. <https://doi.org/10.33404/anasay.1181729>
- Viñas, L. F. (2022). Testing the reliability of two rubrics used in official english certificates for the assessment of writing. *Alicante Journal of English Studies*, 36, 85–109.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wiggins, G. P. (1993). *Assessing student performance: exploring the purpose and limits of teaching*. CA: Jossey-Bass Publishers.
- Yağar, F. ve Dökme, S. (2018). Niteliksel araştırmaların planlanması: araştırma soruları, örneklem seçimi, geçerlik ve güvenirlik. *Gazi Sağlık Bilimleri Dergisi*, 3(3), 1–9.
- Yaylı, D. (2015). Yazma becerisinin öğretimi. N. Bekleyen (Ed.), *Dil Öğretimi* içinde (ss. 287–308). Ankara: Pegem Akademi.

- Yıldırım, A., ve Şimşek, H. (2021). *Sosyal bilimlerde nitel araştırma yöntemleri*. Ankara: Seçkin Yayıncılık.
- Yokouchi, Y. (2025). Revisiting the effectiveness of a performance decision tree-style rubric compared to a grid-style rubric. *Language Testing in Asia*, 15(8), 2-26. <https://doi.org/10.1186/s40468-024-00338-5>
- Yorgancı, K., ve Baş, B. (2021). Yabancı dil olarak Türkçe öğrenenler için b1 düzeyi yazma becerisi dereceli puanlama anahtarı geliştirme çalışması. *Okuma Yazma Eğitimi Araştırmaları*, 9(1), 67–80. <https://doi.org/10.35233/oyea.934684>

## Appendices

### Appendix A: B1 Level Writing Skill Analytic Rubric

Boyutlar	Ölçütler	Ölçüt Tanımlayıcıları				PUAN
		4 Puan (Çok İyi)	3 Puan (İyi)	2 Puan (Geliştirilmeli)	1 Puan (Yetersiz)	
1.Dil Bilgisi	1.1.Kip Kullanımı	Kip eklerini kullanırken neredeyse hiç anlam, işlev ve yapı hatası yapmamıştır. (Gereklilik, dilek, şart kipleri)	Kip eklerini kullanırken anlamsal, yapısal veya işlevsel olarak nadiren hata yapmıştır.	Kip eklerinin anlamsal, yapısal veya işlevsel kullanımında sık sık hata yapmıştır.	Kip eklerinin anlamsal, yapısal veya işlevsel kullanımında çok fazla hata yapmıştır veya cümle kurulumunda kip eklerinden yararlanmamıştır.	
	1.2.Çatı Kullanımı	Cümleler yapı, anlam ve işlev bakımından uyumludur. Özne yüklem uyumsuzluğu yapmamıştır.	Yapı, işlev veya anlam bakımından az sayıda hata görülse de bunlar metin geneline yayılmamıştır ve anlamı bozmamaktadır. Özne yüklem uyumsuzluğu yoktur.	Çatı kullanımında anlam karışıklığına/değişikliğine sebep olabilen hatalar vardır ve nadiren özne yüklem uyumsuzluğu yapmıştır.	Çatı kullanımında sıklıkla anlamı ve işlevi etkileyen hatalar vardır ve özne yüklem uyumsuzlukları söz konusudur.	
	1.3.Fiilimsi Kullanımı	Fiilimsileri kullanırken neredeyse hiç anlam, işlev ve yapı hatası yapmamıştır.	Fiilimsileri kullanırken anlamsal, yapısal veya işlevsel olarak nadiren hata yapmıştır.	Fiilimsilerin anlamsal, yapısal veya işlevsel kullanımında sık sık hata yapmıştır.	Fiilimsi kullanımında anlamsal, yapısal veya işlevsel çok fazla hata yapmıştır veya fiilimsi kullanmamıştır.	
	1.4.Birleşik Cümle	Amaca uygun uzun, bağlı, sıralı ve birleşik yapı cümleleri neredeyse hiç hata yapmadan kurmuştur.	Uzun cümleler kurmuş, bağlı, sıralı ve birleşik yapı cümle kurulumunda yaptığı hatalar anlam karışıklığına/değişikliğine yol açmaz.	Bağlı, sıralı, birleşik yapı cümleler kurarken nadiren anlam karışıklığına yol açabilecek hatalar yapmıştır.	Genellikle kısa cümleler vardır, uzun cümle kurmaya çalıştığında anlam karışıklığına/değişikliğine sebep olabilecek hatalar yapmıştır.	
	1.5.Söz Dizimi	Kurallı cümle kurulumunda ve öğelerin dizilişinde Türkçeye uygunluk bakımından neredeyse hiç hata yapmamıştır. (Özne, nesne tümleş, yüklem, tamlamalar)	Kurallı cümle kurabilmiş, öğelerin dizilişinde uzun cümleler kurarken nadiren hata yapmıştır. (Tümleşlerin yanlış yerde kullanımı gibi)	Kurallı cümle kurabilse de öğelerin diziliminde sıklıkla hata yapmıştır. (Özne ve nesnenin, sıfatların vb. yanlış yerde kullanımı gibi)	Kurallı cümle kurulumunda sorunlar vardır ve öğe diziliminde yanlışlıklar vardır.	
	1.6.Düzeye Uygunluk	A1 ve A2 düzeyindeki dil bilgisi kurallarında neredeyse hiç hata yapmamıştır.	A1 ve A2 düzeyindeki dil bilgisi kurallarında nadiren hata yapmıştır.	A1 ve A2 düzeyindeki dil bilgisi kurallarında sıklıkla hata yapmıştır.	A1 ve A2 düzeyindeki dil bilgisi kurallarında hata sayısı oldukça fazladır. Ancak metin anlaşılabilir.	

2.Sözcük Bilgisi	2.1.İçerik Odaklı Sözcük Seçimi	Temaya uygun sözcük seçimi yapabilmış, bağlama ve anlama uygun doğru sözcükleri kullanabilmiştir.	Temaya uygun sözcük seçimi yapabilse de nadiren bağlamsal veya anlamsal hata yapmıştır.	Sözcük seçimi temayla kısmen uyumludur, sıklıkla anlamsal veya bağlamsal hata yapmıştır.	Temaya uygun seçilen sözcükler temel düzey ile sınırlıdır. Anlamsal ve bağlamsal hatalar sınırlıdır.	
		Yanlış sözcük seçmemiş veya sözcüğü yanlış anlamda kullanmamış, benzer anlamlı sözcüklerden kullanım uygunluğuna göre seçim yapabilmıştır. (kara- siyah, gönül-yürek-kalp vb.)	Anlam bozulmasına yol açmayacak şekilde nadiren yanlış sözcük kullanımı yapmış ve benzer anlamlı sözcüklerin kullanım yerlerindeki farkları ayırt edebilmiştir.	Sözcük seçiminde zaman zaman hatalar vardır (yanlış fiil, yanlış sıfat, yanlış zamir seçimi, yanlış yardımcı fiil vb.) ve benzer anlamlı sözcüklerin kullanım yerlerindeki farkları ayırt etmekte zorlanmıştır ancak metin anlaşılabilir.	Yanlış seçilen sözcükler metnin anlaşılmasını güçleştirmiştir.	
		Söz dağarcığını çeşitlendirebilmiş ve kalıp söz, deyim vb. ile zenginleştirebilmiştir.	Söz dağarcığı çeşitli olsa da aynı sözcükleri nadiren tekrar etmiştir.	Söz dağarcığında yeterince çeşitlilik yoktur, zaman zaman aynı sözcükleri tekrar etmiştir.	Sözcük dağarcığı çok sınırlıdır, cümleler ve sözcükler sık sık tekrar etmektedir.	
	2.2.İşlev Odaklı Sözcük Seçimi	Cümlelerinde anlatmak istediği duruma uygun olarak sözcüklerin yan ve mecaz anlamlarını doğru ve bağlama uygun şekilde kullanmıştır. (ağır=zor, bağ= bağlı olma)	Sözcüklerin yan ve mecaz anlamlarını kullanırken nadiren anlamsal ve bağlamsal hatalar yapmıştır.	Sözcüklerin yan ve mecaz anlamlarını bilse dahi bağlama uygun kullanımlarda hata yapmış veya yan ve mecaz anlam kullanımlarına çok az başvurmuştur.	Sözcüklerin yan ve mecaz anlamlarının kullanımında çok fazla anlam ve bağlam hatası yapmış veya yan ve mecaz anlama başvuramamıştır.	
		Temel bağlaçları/edatları doğru anlamda ve işlevde kullanmış ve aynılarını tekrar etmek yerine çeşitlendirmiştir (fakat, sonuç olarak, bu sebeple, bu yüzden, bundan dolayı, böylece vb.).	Temel bağlaçları/edatları doğru anlamda ve işlevde kullanmıştır ancak bağlaç çeşitliliği çok fazla değildir.	Genellikle aynı bağlaç veya edatları kullanmış, zaman zaman bağlam ve işlev hatası yapmıştır.	Bağlaç ve edat kullanımı çok sınırlıdır, hatalı kullanımlar oldukça çoktur.	
3. Yazım ve Noktalama	3.1. Yazım	Metin boyunca sözcüklerin yazımında neredeyse hiç hata yapmamıştır.	Sözcüklerin yazımında nadiren hata yapmıştır ancak bu hatalar anlam bozulmasına yol açmamıştır.	Sözcüklerin yazımında sıklıkla hata yapsa da metin anlaşılabilir.	Metinde yer alan sözcüklerin yazımında oldukça hata vardır ve bu hatalar metnin anlaşılmasını güçleştirmektedir.	
		Büyük/küçük harf kullanımlarında neredeyse hiç hata yapmamıştır.	Büyük/küçük harf kullanımlarında yazım kurallarının detaylarını bilmeyi gerektiren nadir hatalar yapmıştır.	Büyük/küçük harf kullanımlarında en temel kuralları uygulamıştır (özel isim, cümle başı vb.). Ancak detay gerektiren kurallarda sık hata yapmıştır.	Büyük/küçük harf kullanımları hakkında bilgisi yoktur, çok fazla hata yapmıştır.	
	3.2. Noktalama	Temel noktalama işaretlerini doğru ve yerinde kullanmıştır. (nokta, virgül, soru işareti, ünlem işareti, tırnak işareti, noktalı virgül, kesme işareti, iki nokta)	Temel noktalama işaretlerinin kullanımında detay bilmeyi gerektiren nadir hatalar yapmıştır. Söz konusu hatalar anlam karışıklığına/değişikliğine yol açmamıştır.	Temel noktalama işaretlerine dair sıklıkla hata yapsa da metnin anlaşılabilirliği bozulmamıştır.	Temel noktalama işaretlerinin kullanımında sıklıkla hata yapmış ve bu hatalar metnin anlaşılmasını da zorlaştırmıştır.	

4.Organizasyon	4.1. Metin Planı	Başlık içerikle uyumludur, ilgi çekicidir ve metnin bütünü kapsamaktadır.	Başlık içerikle büyük ölçüde uyumludur ve metnin genelini kapsamaktadır.	Başlık içerikle kısmen uyumludur ve metnin bir kısmını kapsamaktadır.	Başlık <u>yoktur/içerikle uyumsuzdur.</u>	
		Konuya uygun iyi bir giriş ve tanımlamalar yapmıştır. Giriş bölümü metnin geneli hakkında fikir verecek niteliktedir ve bir sonraki bölüm ile ilişkilidir.	Konuya çok genel bir giriş yapmış, metnin genelinden ziyade özel bir hususa odaklanmıştır.	Konuya giriş çok kısa ve yetersizdir.	Metne yönergede verilen konuya uygun bir giriş yapmamıştır.	
		Metni uygun örneklerle zenginleştirmiştir.	Örneklerle anlatmak istediklerini desteklese de düşüncelerini kanıtlama veya kurgulama noktasında yeterli değildir.	Metinde örnek vardır, söz konusu örnekler yetersiz veya konu ile ilgisizdir.	Metinde neredeyse hiç örnek yoktur.	
		Konuyu varılmak istenen noktaya getirip değinilen konuları sentezleyerek sonuca bağlamıştır.	Yardımcı fikirleri sentezleme noktasında eksiklikler olsa da içerikle ilgili genel bir yargıya varabilmiştir.	Metnin sonuç bölümü çok kısa ve yetersizdir.	Metni bir sonuca bağlamadan bitirmiştir.	
		Metnin türünün gerektirdiği tüm biçim ve içerik özelliklerine metinde yer vermiştir. (mektup, dilekçe, tebrik kartı, rapor vb.).	Biçim ve içerik özelliklerini uygularken nadiren hata yapmıştır.	Biçim ve içerik özelliklerinin sadece bir kısmını uygulamıştır.	Biçim ve içerik özelliklerine dikkat etmeden yazmıştır.	
		Yönergeyle belirtilen sözcük sayısına uygun yazmıştır.	Yönergede belirtilen sözcük sayısının yaklaşık %20 altında yazmıştır.	Yönergede belirtilen sözcük sayısının yaklaşık %40 altında yazmıştır.	Yönergede belirtilen sözcük sayısının yarısından az yazmıştır.	
	4.2. İçerik	Metnin tamamı yönergede belirtilen konu ile doğrudan ilişkilidir, konuya uygun ana fikir ve yardımcı fikirleri iyi bir şekilde işlemiştir, konu dışına çıkmamıştır.	Nadiren konu dışına çıksa da metnin geneli konuyla ilişkilidir. Ana fikir ve yardımcı fikirlere yer vermiştir.	Zaman zaman konu dışına çıksa da sonuç bölümünde düşüncelerini toparlayabilmiş, metnin ana fikrine yer verse de yardımcı fikirleri işleyememiştir.	Konu çok fazla dağılmıştır, bütünlük zayıftır, ana fikri yeterince işlememiş ve yardımcı fikirlere yer vermemiştir.	
		Metnin türüne ve ihtiyaca göre birden çok düşüncüyü geliştirme yolundan yararlanmıştır (tanımlama, örnekleme, karşılaştırma, benzetme).	Düşüncüyü geliştirme yollarından yalnızca birinden yararlanmıştır veya yerinde yararlanmamıştır ancak konuyu iyi şekilde işlemiştir.	Düşüncüyü geliştirme yollarından çok az yararlanmıştır ve konuyu yeterince işlememiştir.	Düşüncüyü geliştirme yollarından yararlanmamıştır.	
	4.3. Anlatım	Metnin dili açık ve anlaşılırdır. Metinde anlam tekrarına sebep olan cümleler kullanmamıştır.	Metnin dili açık ve anlaşılır olsa da akıcılığı azdır. Metinde anlam tekrarına sebep olabilecek birbirini içeren az sayıda cümle kullanmıştır.	Anlatımın akışını bozan ifadeler yer almaktadır. Zaman zaman aynı muhtevaya sahip cümleler kurmuş ve aynı sözcükleri tekrar etmiştir.	Metnin dili akıcılıktan uzaktır, metinde kopukluklar fazladır. Konu ile alakalı veya alakasız benzer cümleleri ve sözcükleri tekrar etmiştir.	
		Metin tutarlıdır ve mantıksal şekilde ilerlemiştir. Paragraflar ve cümleler birbiriyle ve	Metin genel olarak tutarlı görünse de nadiren mantık sıralamasında bozulmalar söz konusudur. Paragraflar	Metin genel bir tutarlılık çerçevesi içerisinde ilerlese de sık sık bir fikirden diğer bir fikre atlamıştır. Metinde	Metin tutarlılıktan yoksundur ve mantıksal bir sıralama söz konusu değildir. Paragraflar ve	



	kendi içinde bağlantılıdır.	arasında genel bir bağlantı söz konusudur ancak nadiren cümleler arasında kopmalar mevcuttur.	bütünlük vardır ancak sık sık paragraflar arasında kopukluklar ve cümle bağlantılarında zayıflıklar vardır.	cümleler arasında bütünlük yoktur.	
	Metnin türüne ve/veya toplumdilbilimsel uygunluğa sahip üslupta yazmıştır.	Metnin türüne uygun yazsa da nadiren toplumdilbilimsel olarak uygun olmayan ifadeler yer vermiştir.	Metnin türüne uygun bir üslupta yazsa da sıklıkla toplumdilbilimsel olarak uygun olmayan ifadeler yer vermiştir.	Metnin türüne uygun üslup geliştirmemiştir ve toplumdilbilimsel açıdan uygun olmayan ifadeler yer vermiştir.	
				<b>TOPLAM PUAN:</b>	

### Appendix B: B2 Level Writing Skill Grading Rubric

Boyutlar	Ölçütler	Ölçüt Tanımlayıcıları				PUAN
		4 Puan (Çok İyi)	3 Puan (İyi)	2 Puan (Geliştirilmeli)	1 Puan (Yetersiz)	
1. Dil Bilgisi	1.1. Birleşik Zaman ve Çatı Kullanımı	Birleşik çatı ve/veya birleşik zamanları doğru yapı, anlamda ve işlevde kullanarak hatasız cümleler kurmuştur.	Birleşik çatı ve/veya zamanları kullanırken nadiren hata yapmıştır.	Birleşik çatı ve/veya zamanları kullanırken sık sık hata yapmıştır.	Birleşik çatı ve/veya zamanları kullanırken çok fazla hata yapmıştır ya da ihtiyaç olmasına rağmen birleşik zamanları kullanmamıştır.	
	1.2. Dolaylı Anlatım	Kişi, zaman, fiil (söyle-, de-) ve belirtme eki kullanımlarında hata yapmadan dolaylı anlatım yapabilmektedir.	Kişi, zaman, fiil ve belirtme kullanımlarında nadiren hata yapmıştır, bu hatalar genellikle bu unsurlardan sadece biri ile ilgilidir.	Kişi, zaman, fiil ve belirtme kullanımlarında sık sık hata yapmıştır, bu hatalar genellikle bu unsurlardan birkaçı ile ilgilidir ve çok belirgin hatalardır (söyle- fiili yerine de-kullanımı gibi).	Dolaylı anlatımdan yararlanmamıştır veya anlamayı güçleştiren çok fazla hata yapmıştır.	
	1.3. Fiilimsi Kullanımı	Fiilimsileri kullanırken neredeyse hiç anlam, işlev ve yapı hatası yapmamıştır.	Fiilimsileri kullanırken anlamsal, yapısal veya işlevsel olarak nadiren hata yapmıştır.	Fiilimsilerin anlamsal, yapısal veya işlevsel kullanımında sık sık hata yapmıştır.	Fiilimsi kullanımında anlamsal, yapısal veya işlevsel çok fazla hata yapmıştır veya fiilimsi kullanmamıştır.	
	1.4. Birleşik cümle	Cümlelerini çeşitlendirerek (basit, sıralı, bağlı; eksilteli, devrik cümle) uzun ve karmaşık cümle yapılarını hata yapmadan kurmuştur.	Uzun cümleler kurabilmiş; bağlı, sıralı ve birleşik yapı cümle kurulumunda nadiren hata yapmıştır.	Bağlı, sıralı, birleşik yapı cümleler kurarken sık hata yapmıştır.	Genellikle kısa cümleler kurmuş, uzun cümle kurmaya çalıştığında fazla hata yapmıştır.	
	1.5. Söz Dizimi	Kurallı ve devrik cümle kurulumunda ve öğelerin dizilişinde hata yapmamıştır. (Özne, nesne, tümleç, yüklem)	Kurallı ve devrik cümle kurabilmiş, öğelerin dizilişinde uzun cümleler kurarken nadiren hata yapmıştır. (Tümleçlerin yanlış yerde kullanımı gibi)	Kurallı ve devrik cümle kurabilmiş, öğelerin diziliminde hata yapmıştır. (özne ve nesnenin, sıfatların vb. yanlış yerde kullanımı gibi)	Kurallı ve devrik cümle kurulumunda sorunlar ve öğe diziliminde yanlışlıklar vardır.	
	1.6. Düzeye Uygunluk	A1, A2 ve B1 düzeyindeki dil bilgisi kurallarında neredeyse hiç hata yapmamıştır.	A1, A2 ve B1 düzeyindeki dil bilgisi kurallarında nadiren hata yapmıştır.	A1, A2 ve B1 düzeyindeki dil bilgisi kurallarında sıklıkla hata yapmıştır.	A1, A2 ve B1 düzeyindeki dil bilgisi kurallarında hata sayısı oldukça fazladır.	
		Yönergede verilen konunun gerektirdiği	Metni oluşturan sözcükleri içeriğe	Söz dağarcığı içerikle uyumlu olsa da	Sözcük dağarcığı çok sınırlıdır, cümleler ve	

2.Sözcük Bilgisi	2.1.İçerik Odaklı Sözcük Seçimi	bağlama ve anlama uygun sözcükleri seçebilmiş, çeşitlendirebilmiş, zenginleştirebilmiştir. Soyut sözcükleri kullanabilme yeterliğine sahiptir.	uygun seçebilmiştir ancak zaman zaman aynı sözcükleri tekrar etmiştir. Soyut-somut sözcük ilişkisi kurabilmiştir.	yeterince çeşitlilik bulunmamaktadır. Düzeyinin gerektirdiği sözcük bilgisine sahip değildir ve metnin genelinde somut sözcükler kullanmıştır.	sözcükler tekrar etmektedir. Seçilen sözcükler metnin teması ile uyumlu değildir.	
	2.2. İşlev Odaklı Sözcük Seçimi	Cümlelerinde anlatmak istediği duruma uygun olarak sözcüklerin yan, mecaz, terim anlamlarını doğru ve bağlama uygun şekilde kullanabilmiştir (kol= organ, dal, kapı kolu, meslek dalı vb.).	Sözcüklerin yan, mecaz, terim anlamlarını kullanırken nadiren anlamsal ve bağlamsal hatalar yapmıştır.	Sözcük seçiminde zaman zaman hatalar vardır (yanlış fiil, yanlış sıfat, yanlış zamir seçimi, yanlış yardımcı fiil vb.). Sözcüklerin yan, mecaz ve terim anlamlarını bilse dahi bağlama uygun kullanımlarda hata yapmıştır veya yan, mecaz ve terim anlam kullanımlarına çok az başvurmuştur.	Yanlış seçilen sözcükler metnin anlaşılmasını oldukça güçleştirmiştir. Sözcüklerin yan, mecaz ve terim anlamlarının kullanımında çok fazla anlam ve bağlam hatası yapmıştır veya yan, mecaz ve terim anlama başvurmamıştır.	
		Yanlış sözcük veya sözcüğü yanlış anlamda kullanmamıştır, kullanım uygunluğuna dikkat etmiştir.	Sözcük seçiminde nadiren hata vardır (yanlış fiil, yanlış sıfat, yanlış zamir seçimi, yanlış yardımcı fiil vb.).	Sözcük seçiminde sıklıkla hata vardır (yanlış fiil, yanlış sıfat, yanlış zamir seçimi, yanlış yardımcı fiil vb.).	Yanlış sözcük veya sözcüğün yanlış anlamda kullanılmasından kaynaklanan hatalar çok fazladır ve metnin anlaşılabilirliğini bozmaktadır.	
		Bağlaçları ve edatları doğru anlamda ve işlevde kullanmış ve aynılarını tekrar etmek yerine çeşitlendirebilmiştir. Bağlaçlarla uzun cümleler kurabilmiş ve cümleleri bağlayabilmiştir.	Bağlaçları ve edatları doğru anlamda ve işlevde kullanabilmiştir, ancak kullanımlarında çeşitlilik çok fazla değildir.	Genellikle aynı bağlaç veya edatları kullanmış, zaman zaman bağlam ve işlev hatası yapmıştır.	Bağlaç ve edat kullanımı çok sınırlıdır ve hatalı kullanımlar vardır.	
3.Yazım ve Noktalama	3.1. Yazım	Metin boyunca sözcüklerin yazımında neredeyse hiç hata yapmamıştır.	Sözcüklerin yazımında nadiren hata yapmıştır, bu hatalar genellikle harf hatalarıdır ve anlam bozulmasına yol açmamıştır.	Sözcüklerin yazımında sıklıkla hata yapsa da metin anlaşılabilir.	Metinde yer alan sözcüklerin yazımında oldukça hata vardır ve bu hatalar metnin anlaşılmasını güçleştirmektedir.	
	3.2. Noktalama	Detay gerektiren noktalama işaretleri kullanımlarında dahi (örneğin, tırnak içi alıntı yapma) hata yapmamıştır.	Temel noktalama işaretlerinin kullanımında hiç hata yapmamıştır, detay gerektiren noktalama işaretleri kullanımlarında nadiren hata yapmıştır.	Temel noktalama işaretlerine dair nadiren hata yapmıştır, detay gerektiren noktalama işaretleri kullanımlarında ise hata oldukça fazladır.	Temel noktalama işaretlerinin kullanımında sıklıkla hata yapmıştır ve bu hatalar metnin anlaşılmasını zorlaştırmıştır.	
4.Organizasyon		Başlık içerikle uyumludur, metnin bütünü kapsamaktadır ve özgündür.	Başlık içerikle uyumludur, metnin genelini kapsar ancak sıradandır.	Başlık içerikle kısmen uyumludur ve metnin bir kısmını kapsar.	Başlık yoktur/içerikle uyumsuzdur.	
		Konuya uygun iyi bir giriş ve tanımlamalar yapmıştır. Giriş bölümü metnin geneli hakkında fikir verecek niteliktedir	Konuya çok genel bir giriş yapmış, metnin genelinden ziyade özel bir hususa odaklanmıştır.	Konuya giriş çok kısa ve yetersizdir.	Metne yönergede verilen konuya uygun bir giriş yapmamıştır.	

	4.1. Metin Planı	ve bir sonraki bölüm ile ilişkilidir.				
		Metni örneklerle zenginleştirmiş, görüşlerini gerekçelendirmiş, desteklemiş, çözüm üretmiş veya tartışıp eleştiri yapmıştır.	Örneklerle anlatmak istediklerini desteklemeye çalışmıştır. Ancak ayrıntılandırma yeterli değildir.	Metinde az sayıda örnek vardır, söz konusu örnekler yetersiz ya da konu ile ilgisizdir.	Metinde; neredeyse hiç örnek yoktur.	
		Metinde yer verilen tüm yardımcı fikirleri ve örnekleri sentezleyerek sonuca bağlayabilmiştir. Sonuca varma anlamı içeren bağlama araçlarından yararlanmıştır (sonuç olarak, özetle, bundan dolayı vb.).	Yardımcı fikirleri sentezleme noktasında eksiklikler olsa da içerikle ilgili genel bir yargıya varabilmiştir.	Metnin sonuç bölümü çok kısa ve yetersizdir.	Metni bir sonuca bağlamadan bitirmiştir.	
		Metnin türünün gerektirdiği tüm biçim ve içerik özelliklerine metinde yer vermiştir. (mektup, dilekçe, makale, rapor vb.).	Biçim ve içerik özelliklerini uygularken nadiren hata yapmıştır.	Biçim ve içerik özelliklerinin sadece bir kısmını uygulamıştır.	Biçim ve içerik özelliklerine dikkat etmeden yazmıştır.	
		Yönerge ile belirtilen sözcük sayısına uygun yazmıştır.	Yönergede belirtilen sözcük sayısının yaklaşık %20 altında yazmıştır.	Yönergede belirtilen sözcük sayısının yaklaşık %40 altında yazmıştır.	Yönergede belirtilen sözcük sayısının yarısından az yazmıştır.	
	4.2. İçerik	Metnin tamamı yönergede belirtilen konu ile doğrudan ilişkilidir, konuya uygun ana fikir ve yardımcı fikirler iyi bir şekilde işlemiştir ve geliştirmiştir, konu dışına çıkmamıştır.	Nadiren konu dışına çıksa da metnin geneli konuyla ilişkilidir. Ana fikir ve yardımcı fikirlere yer vermiştir.	Zaman zaman konu dışına çıksa da sonuç bölümünde düşüncelerini toparlayabilmiştir, metnin ana fikrine yer verse de yardımcı fikirleri işleyememiştir.	Konu çok fazla dağılmıştır, bütünlük zayıftır, ana fikri yeterince işlememiştir ve yardımcı fikirlere yer vermemiştir.	
		Her bir paragrafta yalnızca bir ana fikir işlenmiştir ve tüm bu fikirler birbiriyle, metnin ana fikriyle ve verilmek istenen mesaj ile ilintilidir.	Nadiren paragraflarda birden fazla ana fikir yer alsa da bu fikirler metnin akışını bozmamıştır, kendi içinde bütünlük taşımaktadır ve kısmen metnin ana fikriyle ve birbirleriyle ilişkilidir.	Paragrafların bir kısmında birden fazla ana fikir işlenmiştir, bu fikirler metnin ana fikriyle ve birbirleriyle uzak ilişkilidir.	Paragrafların çoğunda birden fazla ana fikir yer almaktadır ve gerek birbirlerinden gerekse metnin ana fikrinden bağımsızdır.	
		Metnin türüne ve ihtiyaca göre birden çok düşüncüyü geliştirme yolundan yararlanmıştır (tanımlama, örnekleme, karşılaştırma, tanık gösterme ve alıntı yapma, somutlama, soyutlama, sayısal verilerden yararlanma, benzetme).	Düşüncüyü geliştirme yollarından yalnızca birinden yararlanmıştır veya yerinde yararlanmamıştır ancak konuyu iyi şekilde işlemiştir.	Düşüncüyü geliştirme yollarından çok az yararlanmıştır ve konuyu yeterince işlememiştir.	Düşüncüyü geliştirme yollarından yararlanmamıştır.	
		Metnin dili; açık, anlaşılır ve akıcıdır. Metin boyunca	Metnin dili genellikle açık ve anlaşılırdır, akıcılığı azdır.	Metnin dili kısmen açık ve anlaşılırdır, metinde anlatımın	Metnin dili açıklıktan ve akıcılıktan uzaktır,	

		okuyucunun ilgisini canlı tutmayı başarmıştır.		akışını bozan ifadeler vardır.	metinde kopukluklar fazladır.	
	4.3. Anlatım	Gerek metnin bütünü, gerekse paragraflar ve cümleler birbiriyle bağlantılı ve kendi içinde tutarlıdır.	Metinde bütünsel bir bağlantı vardır ancak nadiren paragraflar ve cümleler arasında kopmalar mevcuttur.	Metinde bütünlük vardır ancak sık sık paragraflar arasında kopukluklar ve cümle bağlantılarında zayıflıklar vardır.	Metinde, paragraflar ve cümleler arasında bütünlük yoktur.	
		Metinde anlam tekrarına sebep olan cümleler kullanmamıştır.	Metinde anlam tekrarına sebep olan cümleler kullanmamıştır ancak aynı sözcük veya söz gruplarını tekrar tekrar kullanmıştır.	Zaman zaman aynı içeriğe sahip cümleler kurmuş ve aynı sözcük veya söz gruplarını tekrar etmiştir.	Konu ile alakalı veya alakasız benzer cümleleri ve sözcükleri tekrar etmiştir.	
		Metnin türüne ve toplumdilbilimsel uygunluğa sahip üslupta yazmıştır, kendine özgü bir üslup geliştirebilmiştir.	Metnin türüne uygun yazsa da nadiren toplumdilbilimsel olarak uygun olmayan ifadeler yer vermiştir. Kendine özgü belirgin bir üslubu yoktur.	Her metin türünün kendine özgü üslubunun farkında olmasa da resmî ve resmî olmayan üslup geliştirebilmiştir.	Metnin türüne uygun üslup geliştirmemiş ve toplumdilbilimsel açıdan uygun olmayan ifadelere yer vermiştir.	
					<b>TOPLAM PUAN:</b>	

### Appendix A: B1 Level Writing Skill Analytic Rubric

Dimensions	Criteria	Criterion Descriptors				Score
		4. Points (Excellent)	3 Points (Good)	2 Points (Needs Improvement)	1 Points (insufficient)	
1. Grammar	1.1. Use of Modality	Made almost no semantic, structural, or functional mistakes when using modal suffixes (necessity, wish, and conditional moods).	Rarely made semantic, structural, or functional mistakes when using modal suffixes.	Frequently made semantic, structural, or functional mistakes when using modal suffixes.	Made excessive mistakes in the semantic, structural, or functional use of modal suffixes or did not use them at all when forming sentences.	
	1.2. Use of Voice	Sentences are consistent in terms of structure, meaning, and function. No subject-verb disagreement.	There are minor structural, functional, or semantic errors, but they do not affect the overall meaning of the text. No subject-verb disagreement.	Errors that may cause semantic confusion or change due to the use of voice; subject-verb disagreement is rare.	Frequent errors in the use of voice, affecting the meaning and function of the sentence; subject-verb disagreement is present.	
	1.3. Use of Verbals (Gerunds, Participles, Infinitives)	Made almost no semantic, structural, or functional mistakes when using verbals.	Rarely made semantic, structural, or functional mistakes when using verbals.	Frequently made semantic, structural, or functional mistakes when using verbals.	Made excessive mistakes in the semantic, structural, or functional use of verbals or did not use them at all.	

	1.4. Use of Complex Sentences	Formed long, linked, ordered, and complex sentences almost without mistakes.	Formed long sentences; mistakes in linked, ordered, and complex structures did not cause semantic confusion.	Rarely made mistakes that could cause semantic confusion when forming linked, ordered, and complex sentences.	Mostly used short sentences; made mistakes that could cause semantic confusion when trying to form longer sentences.	
	1.5. Syntax	Made almost no mistakes in forming grammatical sentences and arranging sentence elements in accordance with Turkish language rules (subject, object, complement, predicate, and phrases).	Could form grammatical sentences; rarely made mistakes when arranging elements in long sentences (e.g., misplaced complements).	Could form grammatical sentences but frequently made mistakes in the arrangement of elements (e.g., misplaced subjects, objects, or adjectives).	Had problems with sentence formation and arrangement of elements.	
	1.6. Level Appropriateness	Made almost no mistakes in A1 and A2 level grammar rules.	Rarely made mistakes in A1 and A2 level grammar rules.	Frequently made mistakes in A1 and A2 level grammar rules.	Made excessive mistakes in A1 and A2 level grammar rules, but the text remains understandable.	
2. Vocabulary	2.1. Content-Focused Word Choice	Selected words appropriate to the theme and used correct words in context and meaning.	Generally selected appropriate words for the theme but rarely made contextual or semantic mistakes.	Word choice is somewhat consistent with the theme, but frequent semantic or contextual mistakes are present.	Words chosen for the theme are limited to a basic level; frequent semantic and contextual mistakes are present.	
		He/she did not choose the wrong word or use the word in the wrong sense, he/she was able to choose from words with similar meanings according to their appropriateness of use. (black-black, heart-heart-heart, etc.)	He/she rarely used incorrect words without distorting the meaning and was able to distinguish between the different places where words with similar meanings were used.	There are occasional errors in word choice (wrong verb, wrong adjective, wrong pronoun choice, wrong auxiliary verb, etc.) and difficulty in distinguishing the differences in the use of similar-meaning words, but the text is understandable.	Wrongly chosen words made the text difficult to understand.	
		The vocabulary has been diversified and enriched with formulaic words, idioms, etc.	Although his/her vocabulary was diverse, he/she rarely repeated the same words.	There is not enough variety in his/her vocabulary, he/she sometimes repeats the same words.	Vocabulary is very limited, with frequent repetition of sentences and words.	

	2.2. Functional Word Choice	Used literal and figurative meanings of words correctly and appropriately in context.	Rarely made semantic and contextual mistakes when using literal and figurative meanings of words.	Knows the literal and figurative meanings of words but made mistakes in contextual use or rarely used them.	Made excessive semantic and contextual mistakes in using literal and figurative meanings or did not use them at all.	
		Used basic conjunctions/prepositions with correct meaning and function and varied them rather than repeating them (but, consequently, for this reason, therefore, therefore, thus, etc.).	He/she used basic conjunctions/prepositions in the correct meaning and function, but the variety of conjunctions is not very wide.	He/she generally used the same conjunctions or prepositions, and occasionally made context and function errors.	The use of conjunctions and prepositions is very limited, and incorrect uses are quite common.	
3. Spelling and Punctuation	3.1. Spelling	Made almost no spelling mistakes throughout the text.	Rarely made spelling mistakes, which did not affect the meaning.	Frequently made spelling mistakes, but the text remained understandable.	Made excessive spelling mistakes, making the text difficult to understand.	
		He/she almost never made mistakes in using uppercase and lowercase letters.	He/she made rare mistakes in using uppercase and lowercase letters, which required knowledge of the spelling rules in detail.	He/she applied the most basic rules in using uppercase and lowercase letters (proper nouns, beginning of sentences, etc.). However, he/she made frequent mistakes in rules that required detail.	He/she has no knowledge of uppercase and lowercase letters and has made many mistakes.	
	3.2. Punctuation	Used basic punctuation marks correctly and appropriately (period, comma, question mark, exclamation mark, quotation mark, semicolon, apostrophe, colon).	Rarely made mistakes in using punctuation, which did not cause semantic confusion.	Frequently made mistakes in basic punctuation but maintained overall understanding.	Frequently made mistakes in punctuation, making the text difficult to understand.	
4. Organization	4.1. Text Structure	The title is consistent with the content, engaging, and reflects the overall text.	The title is largely consistent with the content and reflects the general text.	The title is partially consistent with the content and reflects part of the text.	The title is missing or inconsistent with the content.	
		It provides a good introduction and definitions appropriate to the subject. The introduction section gives an idea about the general content of the text and is related to the next section.	He/she made a very general introduction to the subject and focused on a specific issue rather than the text as a whole.	The introduction to the topic is too short and inadequate.	The text has not been introduced in a way that is appropriate to the topic given in the instructions.	
		The text has been enriched with appropriate examples.	Although he/she supports what he/she wants to say with examples, it is not enough to prove or	There are examples in the text, but they are insufficient or	There are almost no examples in the text.	



			construct his/her thoughts.	irrelevant to the topic.		
		He/she brought the subject to the desired point and concluded by synthesizing the issues touched upon.	Although there were deficiencies in synthesizing supporting ideas, a general judgment about the content was made.	The conclusion section of the text is very short and insufficient.	He/she ended the text without drawing a conclusion.	
		All the format and content features required by the type of text are included in the text (letter, petition, greeting card, report, etc.).	He/she rarely made mistakes in applying the form and content specifications.	It has implemented only some of the format and content specifications.	He/she wrote without paying attention to the features of form and content.	
		He/she wrote in accordance with the number of words specified in the instructions.	He/she wrote approximately 20% less than the number of words specified in the instructions.	He/she wrote approximately 40% less than the number of words specified in the instructions.	Wrote less than half of the number of words specified in the instructions.	
	4.2. Content	The entire text is directly related to the topic specified in the prompt, handling the main idea and supporting ideas well without going off-topic.	Rarely went off-topic; the text is generally related to the topic, and supporting ideas are presented.	Occasionally went off-topic; managed to summarize thoughts in the conclusion but did not develop supporting ideas well.	The topic is scattered; lacks coherence, and the main idea and supporting ideas are not well-developed.	
		He/she used multiple ways of developing ideas according to the type of text and need (definition, exemplification, comparison, analogy).	He/she used only one of the ways of developing thought, or he did not use it in its proper place, but he handled the subject well.	He/she has made little use of the methods of developing thought and has not dealt with the subject sufficiently.	He/she did not take advantage of the ways of developing thought.	
	4.3. Coherence and Cohesion	The text is coherent and progresses logically; paragraphs and sentences are connected and consistent.	The text is generally coherent, but there are minor logical gaps between paragraphs or sentences.	The text is somewhat coherent but often jumps between ideas, with weak connections between paragraphs and sentences.	The text lacks coherence and logical order; paragraphs and sentences are disconnected.	
		The text is coherent and logically structured. Paragraphs and sentences are interconnected and interconnected.	Although the text appears generally coherent, there are occasional breakdowns in logical order. There is a general connection between paragraphs, but there are occasional breaks between sentences.	The text moves with a general coherence, but jumps frequently from one idea to another. There is unity in the text, but there are frequent gaps between paragraphs and weak sentence connections.	The text lacks coherence and lacks logical order. There is no unity between paragraphs and sentences.	

		Written in a style appropriate to the genre and/or sociolinguistic appropriateness of the text.	Although he/she wrote in accordance with the genre of the text, he/she rarely included expressions that were sociolinguistically inappropriate.	Although he/she wrote in a style appropriate to the type of text, he/she often included expressions that were sociolinguistically inappropriate.	He/she did not develop a style appropriate to the type of text and included expressions that were not sociolinguistically appropriate.	
						<b>TOTAL SCORE:</b>

### Appendix B: B2 Level Writing Skill Grading Rubric

Dimensions	Criteria	Criterion Descriptors				Score
		4. Points (Excellent)	3 Points (Good)	2 Points (Needs Improvement)	1 Points (insufficient)	
1. Grammar	1.1. Use of Compound Tenses and Voices	Constructed error-free sentences using compound voice and/or compound tenses with correct structure, meaning and function..	Rarely made mistakes when using compound voices and/or tenses.	He/she made frequent mistakes when using compound voices and/or tenses.	Made many mistakes when using compound voices and/or tenses, or did not use compound tenses even though needed.	
	1.2. Indirect Speech	He/she was able to make indirect speech without making any mistakes in using person, tense, verb (say-, de-) and accusative suffixes.	The person rarely made mistakes in the use of tenses, verbs, or accusatives; these mistakes usually concerned only one of these elements.	The person made frequent errors in the use of tenses, verbs, and accusatives; these errors usually concerned more than one of these elements and were very specific errors (such as using de- instead of say-).	He/she did not use indirect speech or made too many mistakes that made understanding difficult.	
	1.3. Verbal Usage	He/she made almost no errors in meaning, function, or structure when using verbal nouns.	He/she rarely made semantic, structural or functional errors when using verbal nouns.	He/she frequently made errors in the semantic, structural or functional use of verbal nouns.	He/she has made too many semantic, structural or functional errors in the use of verbal nouns or has not used verbal nouns at all.	
	1.4. Compound Sentence	By diversifying his/her sentences (simple, sequential, connected; elliptic, inverted sentence), he/she constructed long and complex sentence structures without making any mistakes.	He/she was able to construct long sentences and rarely made mistakes in constructing connected, sequential and compound sentences.	He/she made frequent mistakes while forming connected, sequential, and compound structured sentences.	He/she usually used short sentences and made many mistakes when he tried to use long sentences.	
	1.5. Syntax	He/she did not make any mistakes in the formation of regular and inverted sentences and the arrangement of the elements. (Subject, object,	He/she was able to form regular and inverted sentences, and rarely made mistakes in the arrangement of elements while forming long sentences (such as using complements in the wrong place).	He/she was able to construct regular and inverted sentences, and frequently made mistakes in the order of the elements (such as using the subject and object, adjectives,	There are problems in the formation of regular and inverted sentences and errors in the order of elements.	

		complement, predicate)		etc. in the wrong place).		
	1.6. Level Compliance	He/she made almost no mistakes in grammar rules at A1, A2 and B1 levels.	He/she rarely made mistakes in grammar rules at A1, A2 and B1 levels.	He/she made frequent mistakes in grammar rules at A1, A2 and B1 levels.	The number of errors in grammar rules at A1, A2 and B1 levels is quite high.	
2. Vocabulary	2.1. Content-Focused Word Selection	He/she was able to select, diversify and enrich words appropriate to the context and meaning required by the subject given in the instructions. He/she has the ability to use abstract words.	He/she was able to choose the words that make up the text according to the content, but he/she repeated the same words from time to time. He/she was able to establish abstract-concrete word relationships.	Although the vocabulary is compatible with the content, there is not enough variety. He/she does not have the vocabulary required for his/her level and uses concrete words throughout the text.	Vocabulary is very limited, sentences and words are repetitive. The words chosen are not compatible with the theme of the text.	
	2.2. Function-Oriented Word Selection	He/she was able to use the connotative, metaphorical and terminological meanings of words in his/her sentences correctly and in accordance with the context in accordance with the situation he/she wanted to explain (arm = organ, branch, door handle, profession, etc.).	He/she rarely made semantic and contextual errors when using connotative, metaphorical, and terminological meanings of words.	There are occasional errors in word choice (wrong verb, wrong adjective, wrong pronoun choice, wrong auxiliary verb, etc.). Even if the speaker knows the connotative, figurative and terminological meanings of the words, he/she makes mistakes in using them appropriately for the context or uses connotative, figurative and terminological meanings very rarely.	Wrongly chosen words made the text very difficult to understand. There were many errors in meaning and context when using connotative, metaphorical and terminological meanings of the words or they did not use connotative, metaphorical and terminological meanings.	
		He/she did not use the wrong word or word in the wrong meaning, he/she paid attention to the appropriateness of use.	There are rarely errors in word choice (wrong verb, wrong adjective, wrong pronoun choice, wrong auxiliary verb, etc.).	There are often errors in word choice (wrong verb, wrong adjective, wrong choice of pronoun, wrong auxiliary verb, etc.).	Errors resulting from using the wrong word or the wrong meaning of the word are numerous and impair the comprehensibility of the text.	
		He/she used conjunctions and prepositions in the correct meaning and function and was able to diversify them instead of repeating them. He/she was able to make long sentences and connect sentences with conjunctions.	He/she was able to use conjunctions and prepositions with the correct meaning and function, but there was not much variety in their use.	He/she generally used the same conjunctions or prepositions, and occasionally made context and function errors.	The use of conjunctions and prepositions is very limited and there are incorrect uses.	
3. Spelling and Punctuation	3.1. Spelling	He/she made almost no mistakes in spelling words throughout the text.	He/she rarely made mistakes in spelling words, these mistakes were usually spelling errors and did not distort the meaning.	The text is understandable, although there are frequent errors in spelling of words.	There are many mistakes in the spelling of words in the text and these mistakes make the text difficult to understand.	
	3.2. Punctuation	He/she did not make any mistakes	He/she never made mistakes in using basic	He/she rarely made mistakes in basic	He/she made frequent mistakes in the use of	

		even when using punctuation marks that required detail (for example, using quotation marks).	punctuation marks, and rarely made mistakes in using punctuation marks that required detail.	punctuation marks, but he made many mistakes in the use of punctuation marks that required detail.	basic punctuation marks, and these mistakes made the text difficult to understand.	
4.Organization	4.1. Text Plan	The title is compatible with the content, covers the entire text and is original.	The title is consistent with the content, covers the overall text, but is ordinary.	The title is partially consistent with the content and covers part of the text.	Title is missing/incompatible with content.	
		It provides a good introduction and definitions appropriate to the subject. The introduction section gives an idea about the general content of the text and is related to the next section.	He/she made a very general introduction to the subject and focused on a specific issue rather than the text as a whole..	The introduction to the topic is too short and inadequate.	The text has not been introduced in a way that is appropriate to the topic given in the instructions.	
		He/she enriched the text with examples, justified his/her views, supported them, produced solutions or discussed and criticized them.	He/she tried to support what he wanted to say with examples. However, detailing is not enough.	There are few examples in the text, and those examples are insufficient or irrelevant to the topic.	There are almost no examples in the text.	
		He/she was able to synthesize all supporting ideas and examples given in the text and conclude. He/she used concluding conjunctions (as a result, in summary, therefore, etc.).	Although there were deficiencies in synthesizing supporting ideas, a general judgment about the content was made.	The conclusion section of the text is very short and insufficient.	He/she ended the text without drawing a conclusion.	
		All the format and content features required by the type of text are included in the text (letter, petition, article, report, etc.).	He/she rarely made mistakes in applying the form and content specifications.	It has implemented only some of the format and content specifications.	He/she wrote without paying attention to the features of form and content.	
		He/she wrote in accordance with the number of words specified in the instructions.	He/she wrote approximately 20% less than the number of words specified in the instructions.	He/she wrote approximately 40% less than the number of words specified in the instructions.	Wrote less than half of the number of words specified in the instructions.	
	4.2. Content	The entire text is directly related to the subject specified in the instructions, the main idea and supporting ideas appropriate to the subject are well processed and developed, and the subject is not off-topic.	Although it rarely goes off topic, the text is generally related to the topic. It includes main ideas and supporting ideas.	Although he occasionally strays from the topic, he was able to collect his/her thoughts in the conclusion section; although he included the main idea of the text, he could not deal with the supporting ideas.	The topic is too scattered, has poor cohesion, does not adequately address the main idea, and does not include supporting ideas.	
		Each paragraph contains only one	Although paragraphs rarely contain more	Some paragraphs contain more than one	Most paragraphs contain more than one main idea	

		main idea, and all of these ideas are related to each other, to the main idea of the text, and to the message that is intended to be conveyed.	than one main idea, these ideas do not disrupt the flow of the text, are coherent within themselves, and are partially related to the main idea of the text and to each other.	main idea, and these ideas are distantly related to the main idea of the text and to each other.	and are independent of each other and the main idea of the text.	
		He/she used multiple ways of developing ideas according to the type of text and need (definition, exemplification, comparison, giving evidence and quotation, concretization, abstraction, use of numerical data, analogy).	He/she used only one of the ways of developing thought, or he did not use it in its proper place, but he handled the subject well.	He/she has made little use of the methods of developing thought and has not dealt with the subject sufficiently.	He/she did not take advantage of the ways of developing thought.	
4.3.Coherence and Cohesion		The language of the text is clear, understandable and fluent. It has managed to keep the reader's interest alive throughout the text.	The language of the text is generally clear and understandable, with little fluency.	The language of the text is partially clear and understandable, but there are expressions in the text that disrupt the flow of narration.	The language of the text is far from clarity and fluency, and there are many discontinuities in the text.	
		Both the text as a whole, paragraphs and sentences are interconnected and internally consistent.	There is an overall cohesion in the text, but there are occasional breaks between paragraphs and sentences.	The text is cohesive, but there are frequent gaps between paragraphs and weak sentence connections.	There is no unity between paragraphs and sentences in the text.	
		He/she did not use sentences that caused repetition of meaning in the text.	He/she did not use sentences that caused repetition of meaning in the text, but he used the same words or phrases over and over again.	From time to time, he/she made sentences with the same content and repeated the same words or phrases.	Repeated similar sentences and words, whether relevant or irrelevant to the topic.	
		He/she wrote in a style that was appropriate to the genre of the text and sociolinguistic appropriateness, and was able to develop a style unique to himself.	Although he/she writes in accordance with the genre of the text, he rarely uses expressions that are not sociolinguistically appropriate. He does not have a distinctive style of his/her own.	Although he/she was not aware of the unique style of each text type, he was able to develop formal and informal styles.	He/she did not develop a style appropriate to the type of text and included expressions that were not sociolinguistically appropriate.	
					<b>TOTAL SCORE:</b>	

### GENİŞLETİLMİŞ ÖZET

Öğretimde etkililiğin yansıtıcısı olan ölçme değerlendirme çalışmaları öğretimin iyileştirilmesi ve geliştirilmesi açısından büyük bir öneme sahiptir. Öğretim sürecinin

başlatılmasından sonlandırılmasına kadar tüm faaliyetler ölçme değerlendirme uygulamalarının kılavuzluğunda gerçekleştirilir. Eğitim öğretim uygulamalarının öznesi olan öğrencilerinin durumlarının tespiti için iyi yapılandırılmış ölçme uygulamalarıyla beraber, nesnel bir değerlendirmeye ihtiyaç duyulmaktadır. Yabancı/ikinci dil öğretiminde öğrencinin dilsel yeterliğinin en önemli göstergelerinden biri yazma becerisidir ve yazma becerisi de performansa göre değerlendirilir. Performans değerlendirmede nesnelliği sağlamak güç olduğundan değerlendirme için sınırlayıcı ve yönlendirici araçlara ihtiyaç duyulmaktadır. Bu araçlardan en önemli ve yaygın kullanıma sahip olanı dereceli puanlama anahtarlarıdır. Dil becerilerinin performansa dayalı değerlendirilmesinde bütüncül ve analitik olmak üzere iki tür dereceli puanlama anahtarı kullanılmaktadır. Katılımcı sayısı, amaç ve ihtiyaçlara göre bu türlerden biri tercih edilebilir.

Yazma; üsluptan anlatıma, sözcük bilgisinden metin planına birçok unsura sahip ve hedef dilde düşünebilmenin ve özellikle orta düzeyden itibaren üst bilişsel becerilerin aktif olduğu bir beceri olduğundan pek çok açıdan ele alınması gereken bir beceridir. Yazmanın bu çok boyutlu yapısı gereği analitik dereceli puanlama anahtarlarının farklı boyutları değerlendirebilme, öğrencinin zayıf/güçlü yönlerini tespit edebilme ve geri bildirim verebilme yönleriyle yazma becerisinin değerlendirilmesinde daha etkili olarak kullanılabileceği düşünülmektedir.

Çalışmada yabancı/ikinci dil olarak Türkçe öğrenen B1 ve B2 düzeyi öğrencilerin yazma performanslarının değerlendirilmesi için iki farklı analitik dereceli puanlama anahtarı geliştirilmesi amaçlanmıştır. Bu kapsamda dereceli puanlama anahtarı geliştirme aşamaları göz önünde bulundurularak çalışmada keşfedici sıralı desenli karma yöntem uygulanmıştır. Keşfedici sıralı desende önce nitel veriler toplanır ve nitel verilerden hareketle bir ölçek oluşturulur. Oluşturulan ölçek ile nicel veriler elde edilir ve son olarak nicel veriler yorumlanır. Puanlama anahtarı geliştirme kısmı olan nitel aşamada TÖMER’lerde kullanılan örnek puanlama anahtarları, DAOBM’de yer alan kazanımlar ve ders-arac gereçleri incelenerek oluşturulan madde havuzu dikkate alınarak boyutlar ve ölçütler belirlenmiştir. Ek olarak dereceli puanlama anahtarlarının değerlendiriciler tarafından tüm ölçütlerin doğru ve aynı anlaşılmasının ve puanlama anahtarlarının daha kolay ve etkili kullanımını sağlamak için değerlendirici kılavuzları da hazırlanmıştır. Geliştirilen puanlama anahtarlarının geçerlik çalışması için Davis (1992) tekniğiyle alanda en az 10 yıl deneyime sahip yedi öğreticiden uzman görüşü alınarak kapsam geçerlik indeksleri tespit edilmiştir. Araştırmanın nicel aşamasında ise Gazi Üniversitesi TÖMER’de öğrenim gören altı B1 ve



altı B2 düzeyinde olmak üzere on iki öğrencinin yazma performansları deneyimli ve deneyimsiz, yurt içinde ve yurt dışında Türkçe öğreten yirmi değerlendirici tarafından hazırlanan puanlama anahtarlarıyla değerlendirilmiş elde edilen verilerden hareketle Kendall's W uyum kat sayısı hesaplanmıştır. B1 ve B2 düzeyi puanlama anahtarları ile yapılan değerlendirmelerde yüksek uyum olduğu sonucuna ulaşılmıştır. Ayrıca çalışma kapsamında puanlama anahtarlarıyla yapılan puanlamalarda değerlendirici deneyiminin anlamlı fark oluşturup oluşturmadığı Kruskal Wallis tekniği ile analiz yapılmış, deneyimli ve deneyimsiz değerlendiricilerin puanlamaları arasında anlamlı bir fark bulunmamıştır. Ayrıca yabancı dil ve ikinci dil öğretenlerin puanlamaları arasında fark olup olmadığına ilişkin Mann Withney-U analizi yapılmış ve yurt dışında yabancı dil olarak Türkçe öğreten değerlendiricilerle yurt içinde ikinci dil olarak Türkçe öğreten değerlendiricilerin puanlamaları arasında anlamlı bir fark bulunamamıştır. Puanlama anahtarlarıyla gerçekleştirilen değerlendirmelerde deneyimin ve yabancı/ikinci dil öğretiminin değerlendirmeye herhangi bir etkisinin olmadığı sonucuna ulaşılmıştır. Çalışma kapsamında elde edilen sonuçlardan hareketle hazırlanan puanlama anahtarlarının alanda kullanılmasının yazma becerisinin tutarlı bir şekilde değerlendirilebilmesi, daha geçerli ve güvenilir sonuçlar elde edilebilmesi ve ölçme değerlendirme uygulamalarının iyileştirilmesi hususunda yararlı olacağı düşünülmektedir.