

A Study on Rubric Development to Evaluate Writing Skill in Teaching Turkish as a Foreign Language

Nevin KAHVECİ^a, Burcu ŞENTÜRK^{b*}

a Instructor, Çağrı Language School (<https://orcid.org/000-0002-8747-7412>)* nevinkahveci@gmail.com

b* Asst. Prof. Dr., Bartın University (<https://orcid.org/000-0001-8951-3256>)

Research Article

Received: 29.11.2022

Revised: 29.1.2022

Accepted: 29.1.2023

Abstract

This study aimed to develop a valid and reliable rubric for evaluating writing skill in teaching Turkish as a foreign language. The problems experienced in the scoring of the writing exams in institutions and the absence of a standard rubric in the field were the starting point of this study. The research was carried out on the teachers who were teaching at Bartın University Language Education Center. The data collection tool was developed by examining the relevant literature and similar rubric on the subject. In addition, the qualitative findings obtained as a result of the interview questions directed to the experts during the data collection process were used. Data were collected from teachers who are experts in their fields with a questionnaire consisting of 30 items. The obtained data were analyzed. The reliability of the rubric was made by examining the consistency of the scores given by the raters. Kendall W agreement coefficient method was used to determine the reliability. As a result of the analysis, it was found that the reliability of the rubrics developed were .700 for A1 level, .801 for A2 level, .828 for B1-B2 levels, and .807 for C1 level. In line with the results, the consistency coefficient values approaching 1 provides evidence for the reliability of the developed rubrics. As a result of the analysis, it can be said that rubrics developed in accordance with A1-A2-, B1-B2, C1 levels are valid and reliable measurement tools and can be used in the evaluation of writing exams.

Keywords: Writing, assessment and evaluation, scale, rubric

Yabancı Dil Olarak Türkçe Eğitiminde Yazma Becerisini Değerlendirmeye Yönelik Dereceli Puanlama Anahtarı Geliştirme Çalışması

Öz

Bu çalışmada, yabancı dil olarak Türkçe öğretiminde yazma becerisini değerlendirmede geçerli ve güvenilir bir ölçek geliştirmek amaçlanmıştır. Kurumlarda yazma sınavlarının puanlanmasında yaşanan sorunlar ve standart bir dereceli puanlama anahtarının alanda bulunmaması bu çalışmanın çıkış noktası olmuştur. Araştırma Bartın Üniversitesi Dil Eğitim Merkezinde eğitim vermekte olan öğretmenler üzerinden yürütülmüştür. Veri toplamak için hazırlanan uzman değerlendirme formları ilgili alanyazın ve konu ile ilgili benzer ölçekler incelenerek geliştirilmiştir. Ayrıca veri toplama sürecinde uzmanlara yöneltilen görüşme soruları sonucunda elde edilen nitel bulgulardan yararlanılmıştır. 30 maddeden oluşan anket form ile alanında uzman 30 öğretmenden veri toplanmıştır. Elde edilen veriler analize tabi tutulmuştur. Ölçeğin güvenilirliği, puanlayıcıların verdikleri puanlar arasındaki uyumun tutarlılığı incelenerek yapılmıştır. Güvenirlik tespiti için Kendall W uyum kat sayısı yöntemi kullanılmıştır. Analiz sonucunda geliştirilen ölçeklerin uyum katsayısı A1 seviyesi için .700, A2 seviyesi için .801, B1-B2 seviyeleri için .828, C1 seviyesi için .807 olarak bulunmuştur. Sonuçlar doğrultusunda tutarlılık katsayısı değerlerinin 1'e yaklaşması geliştirilen DPA'ların güvenilirliğine kanıt oluşturmaktadır. Yapılan analizler sonucunda A1, A2-, B1-B2, C1 seviyelerine uygun geliştirilen dereceli puanlama anahtarlarının geçerli ve güvenilir ölçme araçları olduğu yazma sınavlarını değerlendirmede kullanılabileceği söylenebilmektedir.

Anahtar kelimeler: Yazma, Ölçme ve Değerlendirme, Ölçek, Rubrik, DPA

To cite this article in APA Style:

Kahveci N. & Şentürk, B. (2023). A Study on rubric development to evaluate writing skill in teaching Turkish as a foreign language. *Bartın University Journal of Faculty of Education*, 12(3), 545-569. <https://doi.org/10.14686/buefad.1211840>

INTRODUCTION

Human beings, who are social creatures by nature, have tried and are trying to provide socialization through various communication ways in their relations with each other throughout the ages. Language, which is the oldest and undoubtedly the most important communication method, and culture, which is its inseparable part, are the main factors in the formation of past and present societies. Thus, language has become the main tool of communication. The language that has developed up to this time has become a bridge in many areas such as communication between other countries, economy, culture and education. This situation has paved the way for the formation and development of today's education programs by improving the teaching of Turkish as a foreign language. As a result of these developments, testing and evaluation were needed to evaluate the educational programs and materials in the field and to test the development of skill, as in mother tongue education.

In line with this need, testing and evaluation have become increasingly important in the education and training process from past to present. It is one of the main parts of education to determine the development of students in the education process and to determine whether the targeted gains are realized at the end of the process. Testing and evaluation, which have been very important in every field of education, also have great importance in language teaching. When going from general to specific, there is a problem of testing and evaluation, which is a known reality in teaching Turkish as a foreign language. Boylu (2019) states that in TÖMERs, "certification is given with arbitrary practices, healthy testing and evaluation practices cannot be made in textbooks. Testing and evaluation is result-oriented only with the course exams at the end of the process, many teachers working in the field prepare questions without considering any criteria. These are among the reasons of the problems in the testing and evaluation of the teaching Turkish as a foreign language field."

Incorrect testing and evaluations cause serious confusion and undesirable results in teaching Turkish as a foreign language. It is more difficult to eliminate the deficiencies of the students who learn Turkish as a foreign language during their education than those who receive education in their mother tongue. In order to minimize the confusion and undesirable results that may occur at the end of the process, accurate and qualified testing and evaluations should be made.

Teaching Turkish as a foreign language takes place around four skills: reading, writing, listening and speaking. "Because the Framework Text sets the standards that can be considered ideal in terms of teaching a language as a foreign language, a skill-based language education and training is carried out on the basis of the Framework Text in the teaching of Turkish as a foreign language (Yorgancı & Baş, 2021). In addition to four skills, it includes two main titles as receptive-comprehension: reading and listening comprehension and productive-telling: writing and speaking skills.

Writing skill, which is defined as a productive language skill, is one of the most difficult skill to acquire and develop during teaching. In the process of acquiring the mother tongue, individuals first acquire the listening skill and finally the writing skill. Although the order of acquisition of language skill in foreign language learning varies, it is known that writing skill are perceived as difficult and complex for individuals learning a new and different language (Azizoğlu, Tolaman, & Tulumcu, 2019).

Problems such as the student's having trouble in writing what he thinks cause a low level of motivation in the students, and the lack of correct determinations and feedback in the testing of the acquired skill, the failure of the student to receive feedback on where he/she made a mistake, his/her incompleteness affect the student's attitude towards this skill and difficulties during the writing education due to the negative attitude developed towards the skill. is happening.

The fact that the writing skill is a hard-to-develop and acquired productive skill also presents the definitions of writing in various ways. Students who encounter difficulties in writing, on the other hand, create a resistance against writing and do not like writing, and as a result, they can avoid writing (Yıldırım & Nurlu, 2016). "While learning a foreign language, individuals stay away from writing because of their prejudices towards writing, as well as being a difficult skill. In particular, writing activities for evaluation put individuals under stress and cause writing anxiety in individuals (Maden, Dincel, & Maden, 2015, p.754)"

The reason for these difficulties is also due to the fact that standardization in teaching Turkish to foreigners has not been achieved yet. Turkish has gained importance by keeping up with the global development in the process, but standardization is still not fully developed and continues as a problem. Deniz and Demir (2019) emphasize that the problem of standardization of Turkish should be solved immediately. The fact that these standardizations have not yet been formed has brought along the problem that the acquisitions of the skill cannot be determined correctly. In addition to the difficult development of the writing skill compared to other skill, the

fact that the standardization has not yet been achieved in the evaluation part brings up the problems of validity and reliability.

Studies on this subject show us that the motivation part of the writing skill in teaching Turkish to foreigners can be eliminated with the necessary methods, but there is a deficiency in the testing and evaluation part. This study is a recommendation for studies on determining the problem and the importance testing and evaluation in teaching Turkish as a foreign language, the necessity of a standard rubric in the evaluation of writing skill, testing and evaluation in general, and the evaluation of writing skill in particular.

“Although the problems experienced in assessment have various causes, the main ones of these problems are rater subjectivity, score compatibility between different raters, and assessment tool and its features. Shows observable features in evaluating written expression activities, also called analytical scoring guidelines. It is predicted that the use of rubric will reduce the problems experienced in evaluation” (Baş & Yorgancı, 2021, p.72).

LITERATURE REVIEW

Testing and Evaluation

The development of the quality of education is directly related to testing and evaluation. According to Göçer (2019), testing and evaluation guide the learning and teaching process. Giving information about teacher and student performances and proficiency and revealing the extent to which desired behaviors are gained make testing and assessment an indispensable part of education in all areas of education. Testing and evaluation are two concepts that are often used together. The reason for this is that the evaluation is made after the testing has taken place.

Özçelik (2016) explains testing and evaluation as trying to achieve desired changes in student behaviors by interacting students who participate in the process with certain entrance behaviors with a teaching-learning environment prepared for students. Thus, the answer to the question “whether the expected behavior is realized or not” is directly related to testing and evaluation. In order for education and training to be more functional, the studies carried out in the testing and evaluation processes should be qualified and in a way that determines the level. Testing and evaluation have important place in language teaching. Language is a tool that has its own rules and provides communication between people within the framework of these rules (Gurses,2006, p.131).

For this reason, it is necessary to test and evaluate the gains, inputs and outputs in language teaching well. In this context, the most important purpose of testing and assessment on language teaching is to accurately measure and improve the quality of the language teaching process, the teaching process and the result.

“Testing and evaluation play an important role in determining the extent to which students achieve their goals. Therefore, testing and evaluation should be learned and applied correctly” (Çakmak,2009, p.15). For this reason, of the most important issues that should be emphasized is that testing and evaluation tools to be used in language teaching should be reliable and valid. Although teaching Turkish to foreigners shows similarities with teaching Turkish in the mother tongue, there is a more complex structure in teaching Turkish as a foreign language. In addition to teaching the language, its involvement in the transfer of culture complicates the teaching of Turkish as a foreign language and the perspectives of foreign students on Turkish. Considering these reasons, the way of teaching and testing and evaluation should be shaped according to the process and its functionality should be carefully emphasized.

Testing

While Turgut (1997) defines testing as observing a quality subject to measurement and explaining the results of observation with numbers or symbols, Ercan and Kan (2004) define the concept of testing as making certain inferences based on the results of the evaluations by making evaluations about situations or objects. Based on these definitions, testing can be defined as making various evaluations about individuals or objects in general terms, measuring the attributes with appropriate tools and expressing the attributes with numbers or symbols according to the evaluation results.

Tekin (1996) states that the concept of difference is the basic concept for testing, and based on this, he emphasizes that the subject of testing is a changing feature, and expresses that testing emerges from differences.

In another definition, testing and evaluation is the effort to create desired changes in the behavior of students by combining the elements of a teaching-learning environment prepared for the students involved in the process (Özçelik, 2016). Thus, the answer to the question “whether the expected behavior is realized or not” is directly related to testing and evaluation.

The fact that learning levels of the students are different from each other and the determination of whether the targeted gains according to the students' levels in education are realized at the desired level are determined by testing and evaluation. In this context, it is one of the main elements of testing from past to present, due to some features that the student, who is the most important part of the education and training process, has. (Boylu, 2019).

Evaluation

The definition of evaluation is generally expressed by Turgut and Baykul (2014) as the process of reaching a value judgment by comparing measurement results with a criterion. Evaluation takes and examines the results obtained with measurement tools. After these examinations, a certain judgment is reached.

Testing and evaluation are concepts that affect each other. In order to make an accurate assessment, the instrument used in the testing must be reliable. Incorrect results from incorrect testing will result in an incorrect assessment. Evaluation should be based on a valid criterion and transactions should be carried out without errors (Turgut & Baykul, 2015, p.71). For this reason, if there is no testing, there will be no evaluation, and if the testing is done incorrectly, the evaluation will serve a wrong purpose. Evaluation is carried out in order to determine to what extent the targeted gains are gained and to what extent the studies carried out in the process are successful.

This important mission of testing and evaluation in education brings these two important concepts to the forefront in language teaching and becomes an important and necessary part of language teaching. This necessity has also shown itself in teaching Turkish as a foreign language, however, it has taken its necessary place and importance in teaching Turkish as a foreign language, which is a new field.

Validity & Reliability

Two of the concepts that are of great importance in testing and evaluation are validity and reliability. In order for us to make accurate and healthy measurements and evaluations, our measurement tool must have these two features.

Sönmez (2003, p. 418) expresses the concept of validity as “the degree to which a test really measures the quality it wants to measure”. Validity is related to how accurately the test measures the individual's desired feature without confusing it with other features (Büyükoztürk et al, 2019, p.121). In addition to these definitions, Göçer (2005, p. 53) expresses validity as the appropriateness of the scale to be used.

There are four types of validity that must be present in a measurement tool.

These are as follows;

Content validity: Whether the measurement tool covers the behaviors to be measured. Content validity is the ability to represent all subjects of the field to be measured by the measurement tool (Çakmak, 2009, p.19).

Büyükoztürk indicates that with the validity, the answer to the question “Do the test items adequately reflect the behavior that is intended to be measured?” is sought (2019, p.122).

Construct validity: Facts existing in the theoretical dimension of science are defined conceptually or functionally, and the relationships between phenomena are tried to be explained by using measurement. These hypothetical phenomena that cannot be proven but can be noticed by measurement are called “structures” (Büyükoztürk, 2019, p.123).

Tekin defines construct validity as determining whether a test or measurement process measures a theoretical construct (1991, p.52).

Face validity: Face validity is not about what a test actually measures, but what it seems to measure (Tekin, 1991, p.53).

Predictive validity: In the prediction night, the relationship between the test score and the behavior to be measured in the future is examined, and it is investigated to what extent the test scores predict the future behavior (Büyükoztürk, 2019, p.123).

Another important concept for testing and evaluation is reliability. Öncü (1994) defines the concept of reliability as an indicator of the consistency of the values obtained in repeated measurements under the same conditions. In addition to this definition, reliability is not only related to the measurement tool, but also a feature related to the measurement tool and the results of the measurement made with this tool (Ercan & Kan, 2004).

The more a measuring instrument is free from random errors, the more reliable it will be. In other words, if the measurement tool gives the same result every time it is applied, it is reliable (Çakmak, 2009, p.21). Measurements made with unreliable scales are invalid and useless. The concepts of validity and reliability, which have been defined above, are important requirements to be considered in education. These two concepts, which

are as important in foreign language education as they are in mother tongue education, are the two basic elements that make up measurement and evaluation.

Testing and Evaluation in Teaching Turkish to Foreigners

Testing and evaluation studies are the basic building blocks of education as they have great place in ensuring that the measurements made are suitable for the purpose, determining the acquisition status of the students aimed to gain, and guiding the studies to be done later (Göçer, 2007).

“At the last point of modern language education approaches, multiple choice tests, true-false tests, matching tests, complementary tests and exams, written exams based on subjective criteria, which do not allow students to present high-level mental skills at a sufficient level, have been abandoned and written expression skills of students in the process have been abandoned. It is adopted that these testing and evaluation studies are carried out with tools that include criteria and features that are accepted by everyone” (Şengül, 2011, p.6).

Testing and evaluation are indispensable parts of the education process. For this reason, it has an important place in teaching Turkish to foreigners. Mistakes in language teaching are inevitable parts of the process. Testing and evaluation made to minimize these errors are of great importance in language teaching. Since learning a language requires a process, it causes a waste of time when the teaching process is not activated.

Teaching Turkish to foreigners is based on 4 skills. These four basic skills are reading, writing, listening and speaking. The targeted gains were specified in the Common European Framework for Reference, and as a result of the training and education made for this purpose, the necessity of testing and evaluation has emerged in order to determine the conformity of the achievements to the framework. In this context, subjects such as students' attainment of the targeted gains, their use in daily life, and their competencies can be determined and regulated by testing and evaluation.

The main purpose of the educational process is to achieve success. For this, Turkish teachers working in this field should carry out the process correctly and in a planned way. The goal is to ensure the acquisition of language skills. Since the determination of whether these skills are acquired by students correctly or not is made with testing and evaluation. It is an area that requires more attention and devotion than it is thought in the acquisition of language skills and in the field of teaching Turkish to foreigners.

“Language skills can be summarized as individuals' ability to fully and accurately understand what they listen to, see, and read, and then explain them to others fully and accurately. Gaining these skills is based on four main activities such as listening, reading, speaking and writing. These activities are not separate from each other, but are in a complementary relationship (Maltepe, 2006, p.16)”.

The Importance of Writing Skill in Teaching Turkish as a Foreign Language

According to Özaslan (2018, p.19), “writing is considered as a visual form of speech, which is another narrative skill. Therefore, it is sometimes thought that it is sufficient and easier to teach students to speak a foreign language. However, writing language is not an extension of learning to speak. We learn to speak our mother tongue at home without systematic training, but many of us need systematic training in school to learn to write.”

Gündüz (2003, p.15) describes writing as exploring life and trying to know other people's universe. Based on this definition, we can say that the writing skill helps to get to know other cultures in socio-cultural terms.

Learning a foreign language means meeting a language and culture other than one's mother tongue. While learning a foreign language, every person first builds a psychological wall for himself and from time to time thinks that he cannot learn this new language. Therefore, all kinds of boredom and difficulties should be avoided while teaching a foreign language. Here, the task falls to the instructor. No single book is sufficient for language teaching (Barın,2004, p.20).

Writing education has been used at every stage of the education and training process. One of the skills involved in teaching Turkish as a Foreign Language, which is based on the acquisition of language skills and the transfer of acquired skills, is writing skill. It is extremely important to emphasize and develop the writing skill, among other skills. Because it is a known fact that the writing skill of students who learn Turkish as a foreign language greatly affect their success, especially when they receive university education (Büyükkiz, 2012).

"Writing skill in second language teaching have very different meanings especially for university students. It is seen that the writing skills of these students are more important for their academic success than their speaking skill" (Tok, 2013, p.251).

When evaluated from this perspective, the fact that the lack of writing skill in written language proficiency exams causes failure in other areas as well (Yıldırım & Şimşek, 2011, p.256) is an important issue that necessitates the development of writing skills in the field of teaching Turkish to foreigners. Özbay (2011) stated that the writing

skill has a positive contribution to the assessment and evaluation process and that students control and organize their knowledge; He stated that it is a very important skill because it allows him to expand his thoughts and express them in a planned way. Turkey is a country with a potential for foreign students due to its social and geographical structure. For this reason, the need for foreign language teaching is increasing day by day. This need is important for foreigners in Turkey or for foreigners who want to learn Turkish outside of Turkey. For this reason, the subject of teaching Turkish to foreigners is an important issue that must be taken into account. For this, significant researches and studies should be carried out that will contribute to teaching Turkish as a foreign language (Göçer & Mongol 2011).

Evaluation of Writing Skill in Teaching Turkish to Foreigners

Mankind has needed to communicate in every period of life. This need has revealed the necessity of using writing skill as well as speaking skill. Barın (2009, p. 21) states that the reason why the writing skill is more complex than other skills is that “the writer must use which words to express the message, how to arrange the sentence and how to write it in accordance with the spelling rules so that the message he gives in writing is understood correctly and completely by the reader. stressed the need to know These requirements complicate writing skill and teaching writing skill.

Temizkan (2003) stated in his study that the evaluation of written expressions is done haphazardly in the current system and it is almost of no benefit to the student. In addition to measuring and evaluating the writing skill, it is also important to score and evaluate the exams correctly. Even if the validity and reliability of the prepared exam is high, errors in the evaluation will reduce the validity and reliability of the exam.

Boylu (2019) multiple choice, true-false, yes-no etc. He stated that while the tests are difficult to prepare and easy to score, it is very difficult to score the writing and speaking exams easily. The reason for this is that there are no scoring keys that specify clear criteria in the scoring of the writing and speaking exams. Scoring for writing exams varies according to institutions, schools and courses. Some of these ratings are tried to be explained below.

The scoring format that the teacher scores according to their experience and experience is the scores given by the teachers working in this field without using general criteria. This scoring affects the correct evaluation of the exam. In this type of scoring, the necessity of evaluating at least two people who work in the paper field has been stated in the studies conducted in the field.

Özçelik (1992) stated that a teacher gave a different score when he scored a paper again at another time, or a different score was given when a different teacher scored the same paper, and there was no consistency between these scores. As proof of this problem, a study conducted in Boylu's (2019) doctoral thesis examined the consistency of teachers in scoring. In the study, “60 teachers with at least 2 years of experience in the related field were sent a written expression sheet of a student and the teachers were asked to score the paper over 25 points without using any scoring key and to determine the level of the student.” As a result, 60 teachers gave points from 22 to 8 points and determined their levels in the range of A1-C1. The result of this study is evidence of the inconsistency in the ratings.

Using a rubric is another way to score writing skill. This method is the most reliable form of scoring written expression papers. Because in this method, the rater knows according to which criteria to score the student's paper, and this increases the reliability of scoring. Because it was determined in advance according to which criteria the points to be given to the students' writing papers would be given (Boylu, 2019, p.127). Moskal and Leydens (2000) underlined that with rubrics, it is possible to determine in more detail what deficiencies the students will have at which level.

Using Rubrics

As there are different teaching techniques and methods in the education system, different measurement methods are used to measure how and at what level the targeted teaching is learned at the end of the process. Effective and reliable measurement tools are needed for the safe evaluation of students' written performance-based studies. In this context, various checklists, rating scales and rubrics are used to measure the quality to be measured (Aktaş and Alici, 2018).

There are measurement types that are frequently used in education to measure student achievement. These exams deal with the results of the exams, not the processes of the students. With the changing understanding of education, besides the product, alternative evaluation methods which evaluate the students in the process are presented to us.

Büyüköztürk (2007) described one of these evaluations as performance evaluation and expresses performance evaluation as an evaluation made with tools with high reliability and validity. Another point that creates a disadvantage for students in traditional assessment and evaluation approaches is that students do not have enough information about what is expected of them (Korkmaz, 2009, p.4). Rubrics are one of the most common measurement tools used in assessments.

In their article, Parlak and Doğan (2014) stated that the rubrics' structure consisting of criteria and performance levels minimizes the biases that may occur during scoring, thus providing students with more realistic and detailed feedback about their performance. In a broad sense, rubrics are detailed scoring guides used to measure student performance with predetermined criteria (Korkmaz, 2009). Rubrics, also known as rubrics, are powerful tools for measurement and evaluation.

Rubric types are divided into two as analytical rubrics and holistic (holistic) rubrics.

- Analytical Rubric: “It is a scoring tool that gives information about the success levels in various dimensions of student achievement. The degrees of each dimension are defined in detail” (Öztürk and Güdek, 2011, p.8). The student is given detailed information.

- Holistic Rubric: Through this scoring, the teacher evaluates the work as a whole (Öztürk & Güdek, 2016).

Benefits of using rubrics

Rubrics allow students to self-assess. “While the rubrics guide the students about the work to be done, they will enable them to realize their weaknesses and show them concrete ways to improve their deficiencies” (Korkmaz, 2009, p.23).

Rubrics are of great benefit to teachers and the assessment and evaluation process as well as to students. The most important of these benefits is that the instructors behave more objectively when evaluating the student's work, and have knowledge about the progress of the students and the areas that need improvement (Korkmaz, 2009). In addition to these, the use of rubrics reduces the time in the evaluation process and provides standardized evaluations.

Common European Framework of Reference

In teaching Turkish as a foreign language, the Common European Framework of Reference (CEFR), which is used as a resource in order to reveal the functions and achievements of Turkish, is used as a resource because it determines the language levels and the equivalents of these levels.

“Common European Framework of Reference is used in European countries to prepare curricula for teaching target languages, curriculum guidelines, exams and textbooks. It comprehensively describes what language learners need to do and learn to meet their communicative needs in this language, and what knowledge and competences they need to develop in order to be communicatively successful. These definitions also include the cultural dimension of that language. The CEFR also defines the proficiency levels required to measure the language learner's achievements throughout life and at each stage of the learning process” (TELC, 2013, p.11). In teaching Turkish as a foreign language, the curricula, exams, textbooks, materials are carried out taking into account the criteria in the CEFR. It is important in terms of catching certain and valid standards in language teaching. “Preparing a comprehensive, transparent and coherent Framework for language learning and teaching does not mean accepting a uniform order for all. On the contrary, the CEFR should be open and flexible so that it can be used in any situation and make necessary adaptations” (TELC, 2013, p.16). The CEFR defines itself as a measurement and evaluation tool.

Acquisitions of writing skill

The CEFR has provided specific schedules according to skills and levels to guide and facilitate language learners or trainers for specific purposes. There are sets of language proficiency levels in the CEFR. They divide these clusters into 3 according to their functions;

Sets of digits for users

Level descriptors for evaluators

Level clusters according to exam authors (TELC, 2013, p.42,43)

Studies on Testing and Evaluation in Teaching Turkish to Foreigners

Çakır (2010) emphasized the inclusion of student expectations in the writing process in his study, studied the students' thoughts about writing skills and how to gain effective writing skills. identified and made suggestions to make the lessons more productive.

In his study, Tiryaki (2013) stated that the teaching process of writing skill consists of certain stages and steps, he gave some focal points for the development of writing skill, and he explained each stage in his work.

In his article titled "Writing Teaching", Ünsal (2008) emphasized the necessity of emphasizing the writing skill, that it is in the background compared to other skills, but that written communication is as important as oral communication and conveyed the activities that can be done before, after and during the writing teaching process.

Arslan and Klicic (2015) put forward the problems encountered in the development of writing skills with the example of Bosnia and Herzegovina. In the study, controlled, guided and free writing activities were applied to the target group. As a result, it was determined that the students had problems due to the fact that Turkish is an agglutinative language, and it was suggested that Turkish teachers should pay attention to the writing of affixes while developing their writing studies.

Boylu and Şen (2017) developed a writing anxiety scale for those who learn Turkish as a foreign language in their study. A scale form was created and opinions were received by experts in the field for this scale form. The study was applied to 280 students learning Turkish at Aydın TÖMER and Yunus Emre Turkish Cultural Centers. As a result, it was concluded that this scale was applicable and the results of confirmatory factor analysis were acceptable.

Büyükciz (2012), selected as the study group in his article titled "Developing a writing skill self-efficacy scale for Turkish learners of a foreign language: validity and reliability study", developed the scale to measure writing self-efficacy for 156 students learning Turkish at Ankara TÖMER and Gazi University TÖMER. has applied. As a result, it has been seen that the scale is applicable and distinguishes students with high self-efficacy perceptions of writing skills from students with low self-efficacy perceptions. This scale is a study that can be useful for researchers who want to study on self-efficacy perceptions of writing skills.

While Selvikavak (2006) contributed to the development of writing skills with the paragraph writing program in his master's thesis, Şimşek (2017) examined whether the students' mother tongue and the language families they belong to had an effect on their writing skills, and conveyed the positive and negative results of these effects to us in his research. In addition to these studies, there are studies examining the contributions of some educational approaches to writing skills. Takil (2014) developed a language teaching model adapted to the communicative approach to measure its functionality in teaching Turkish as a foreign language. As a result of the study, it was concluded that the students to whom the language teaching model was applied had higher writing skill levels than the students to whom the traditional teaching was applied.

Gedik (2017), in his study titled "Testing and Evaluation in Teaching Turkish to Foreigners", stated that the points to be considered in the evaluation of writing skills, the necessity of having at least two raters, and the scoring of the third rater would be a more accurate assessment in cases where there is a difference of more than 5 points. it does. The most important point is that it emphasizes the importance of teachers using a scoring ruler when scoring their writing skills. The results of this research prove the necessity of analytical rubrics (Rubrik) in order to make accurate and healthy measurements when evaluating writing skills in teaching Turkish to foreigners. There is a need for analytical rubrics for writing skills with validity and reliability (Boylu, 2019, p.332).

Boylu (2019) stated that a valid rubric does not exist in the field of writing skills and included the points to be considered while scoring. Among the results of the study, it was suggested that the consistency of the teachers teaching Turkish to foreigners in measuring and evaluating their writing skills was not at the desired level, and on the basis of this, it was suggested to develop analytical rubrics whose validity and reliability were determined to measure and evaluate writing skills in teaching Turkish to foreigners.

Karatay (2011) stated that the main problem in writing skills is that having a reliable, valid and consistent measurement tool to follow the development process of written expression skills is extremely important to determine learning qualities, and that the most important problem of written expression/writing education in our country is easy and applicable evaluation studies. stated as absent.

Yorgancı and Baş (2021), in their article titled "B1 level writing skill rubric development study for learners of Turkish as a foreign language", developed a B1-level rubric and the developed DPA was used in the exam papers of 30 students. As a result of the study, it has been proven that the DPAs developed are usable.

As a result, studies on writing skills in teaching Turkish as a foreign language have focused on general topics such as gaining writing skills, the process of writing skills, and composition writing. No studies on development were found.

Research Questions

In line with the stated purposes of the research, “How should a rubric be used to evaluate writing skill in teaching Turkish as a foreign language?” addressed within the framework of its main problem. The sub-problems of the research are:

1. Are the rubrics which are developed in this study valid and reliable?
2. What are the views and practices of the instructors on using these rubrics?

METHOD

The model of this study, which aims to provide a standard and objective evaluation in the evaluation of writing exams, is the "mixed method" model. In this study, quantitative and qualitative methods were used together. “The basic assumption of mixed methods research is that combining statistical values (quantitative data) with narratives and personal experiences (qualitative data) will be more advantageous than using any of the research methods alone to understand the research problem in detail (Creswell, 2017, p. 2).” There are different hash methods used in hash method implementation. There are different types of approaches to the application of mixed methods. Here, quantitative and qualitative techniques can be combined in various ways. “While quantitative methods and techniques are sometimes brought to the fore in the research design, sometimes qualitative methods and techniques can be in the foreground. In addition, in some mixed method research, both methods can be used together with equal importance (Üst Can, 2020).

After the determination of the needs for the assessment of writing skill in teaching Turkish to foreigners, which is the main purpose of the research, it was concluded that the most appropriate method for developing rubric to evaluate writing skill was the exploratory sequential design.

The explanatory sequential pattern consists of two stages. In this design, the researcher begins his work with a quantitative research. Develops a measurement tool within the framework of the results obtained from quantitative research. And then, with this developed design, the qualitative data of the research is collected (Bulaç & Kurt, 2019). In this design, firstly, quantitative data is collected and then the qualitative data stage is started on the results obtained. Then, the qualitative data obtained are explained in a way that helps to explain the quantitative data. The data obtained in this research were transferred to the SPSS-24.0 statistical package program and analyzed.

In this context, rubric should be developed to evaluate writing skill in teaching Turkish to foreigners in order to achieve the main purpose of the research. For this reason, qualitative studies were carried out first and then the rubric development stage was started.

Quantitative Research Phase

In this section, information on, rubric developed to evaluate writing skill in teaching Turkish to foreigners is given. In the study, firstly, rubrics were developed by considering the qualitative data obtained in the first stage of the research. Information about this is presented under this heading.

Rubric Design

Under this title, regarding the rubric development phase, one of the quantitative parts of the research; Data collection tool, study group, preparation of rubric, preparation of items and criteria, getting expert opinions, validity and reliability analyzes of, rubrics are presented.

Data Collection Tools

Multiple data collection tools were used in this study. In the process of creating the rubric, expert evaluation questionnaire and interview questions were used in line with the purpose of the research. This questionnaire and interview questions created by the researcher were prepared in 4 stages: "defining the problem", item (writing a question), "receiving expert opinions" and making a preliminary application by scanning the field in line with the purpose of the research and taking the questionnaires and questions prepared before as an example. The aforementioned expert evaluation interview and questionnaire forms are presented in Appendix 1.

Sampling

The data of this study were obtained from 12 lecturers working at Bartın University Language Center, after the course exams, by scoring the written exams with the rubrics developed and then by the interviews and expert evaluation forms. Again, the rubric was sent to lecturers from different universities who are experts in the field, and their opinions were obtained with expert evaluation forms.

Rubric Preparation

Goodrich (1997) describes rubrics as documents listing the criteria of a study and describing these criteria in detail from good to bad. There are two types of rubrics: holistic and analytical. Holistic rubric when the performance or product of the students is considered and evaluated as a whole; When the performance of the students or the components related to the product are evaluated separately, an analytical rubric is used (Atılğan, Kan, & Doğan, 2007). In this study, it was decided to create an analytical rubric, taking into account the detailed scoring of the student's proficiency and level at each stage of the scoring.

In the process of creating rubrics, the steps in Andrade's (1997) scoring rubric creation process were used. These steps are as follows:

- a. Literature review on writing skill in teaching Turkish as a foreign language
- b. Defining criteria, levels and scores
- c. Preparation of the draft rubric
- D. Using the draft rubric
- to. Receiving feedback
- f. Reviewing and editing the draft rubric
- g. Conducting validity and reliability studies
- h. Finalizing the rubric

Preparation of Items and Criteria

After starting the process of creating rubrics suitable for the level, the process of preparing the items and criteria of the rubrics was started.

The research method used in the first step of the process of creating the items and criteria determined in the preparation of rubrics, which aims to score the writing skill in a more objective and standard way in teaching Turkish to foreigners, is document analysis, one of the qualitative research methods. Document analysis is a qualitative research method used to analyze the content of written documents meticulously and systematically (Wach, 2013). At the beginning of the process, information was collected by making a literature review.

While creating the items and criteria of the developed, rubrics, the competencies required to be possessed at the A1/A2, B1/B2, C1/C2 levels in teaching Turkish to foreigners were determined by using the CEFR descriptors.

The second stage is the division of the criteria in the rubrics into levels. The classification of definitions into levels was prepared as “very good, good, moderate, insufficient”, taking into account the studies in the literature. Later, during the pilot study, using the ratings of the institutions, it was prepared as 10-15-20-25 points, and adjustments were made as “good, moderate, insufficient” at the levels.

While creating the, rubric draft, the books used in the field from A1 to C1 levels were examined and their level contents were scanned. Success levels were determined for the criteria and definitions were made. A framework was created by determining the common contents appropriate for the level and the achievements in accordance with the CEFR. A draft, rubric was then prepared.

Getting Expert Opinions

Content validity is whether the measurement tool covers the behaviors to be measured. Content validity is the ability to represent all subjects of the field to be measured by the measurement tool (Çakmak, 2009, p.19). Content validity is explained by Karasar (2003) as deciding whether the measurement tool represents the area to be measured by consulting expert opinions on whether it is suitable for the purpose for which it will be used.

Rubric developed by considering the definitions were presented to 7 faculty members who are experts in their fields and evaluated in terms of content, structure and criteria. Questionnaires prepared by the researcher were given to the experts and submitted for their evaluation. Thematic analyzes of the data obtained from experts were made. Re-arrangements were made in line with the recommendations and opinions of the experts. The validity of the developed, rubrics was ensured in this way.

Rubric prepared at A1, A2- B1, B2, C1 levels were used in the evaluation of the writing exams of a group of 8 students at the end of the A1- A2-B1-B2 level in the institution where the researcher works. The finalized, rubrics were used at different levels in the writing sections of the final exams at TÖMER, Bartın University. After the exam evaluations, the opinions of the lecturers using the, rubrics were taken with the questionnaire form and interview questions created by the researcher. The analysis of the expert opinions received is included in the findings section.

Evaluation of Writing Skill According to Rubrics

The purpose of the research before the evaluation was explained to the teachers working at TÖMER at Bartın University and the rubrics to be used in the evaluation were introduced. Students were asked to evaluate the writing skill section with these rubrics in the final exams. 7 teachers who are experts in their fields were asked to score a total of 60 writing exam papers from A1 level to C1 level. After the evaluations were made, the harmony between the teachers was checked. This information showing the reliability of the prepared rubric is presented in the findings section of the study.

Validity and reliability studies of the research

The reliability and validity studies of the research were carried out in the specified ways. Sönmez (2003, p. 418) expresses the concept of validity as the degree to which a test really measures the quality it wants to measure. "Validity is about how accurately the test measures the individual's desired feature without confusing it with other features" (Büyüköztürk et al, 2019, p.121). Öncü (1994) defines the concept of reliability as an indicator of the consistency of the values obtained in repeated measurements under the same conditions. In addition to this definition, reliability is not only a feature of the measurement tool, it is a feature of the measurement tool and the results of the tool (Ercan & Kan, 2004, p.212).

The validity of the rubrics developed in the research was made around the structure, content and criterion. For this reason, "Expert Evaluation Form for the Development of Rubrics for Evaluating Writing Skill in Teaching Turkish as a Foreign Language" and "Expert Interview Questions" were prepared by the researcher in order to obtain expert opinions in order to determine the validity of the rubrics. With the expert evaluation forms prepared by the researcher, the validity of the rubrics developed by 3 lecturers who are experts in their fields were checked. The validity and reliability studies of the rubrics developed as a result of the pilot studies were carried out. After the prepared interview questions and rubrics were used in the exams, interviews were conducted with the instructors who administered the exam. These studies are included in the findings section. In this study, scores were made by independent evaluators to determine the reliability of rubrics and the compatibility between raters was checked. Kendall's W coefficient was calculated for inter-rater reliability and the percentage of agreement was calculated for intra-rater consistency.

Data Analysis

Expert evaluation questionnaire forms created by the researcher and expert interview questions were sent to experts in their fields before and after the exams. In this process, the rubrics, which were made in draft form during the design phase of the rubrics, were sent to the teachers in PDF format on the computer with the aim of getting expert opinion, and they were asked to examine the rubrics in a one-week period and then fill out the expert evaluation questionnaires on the computer and deliver them to us.

At the end of a period of approximately one week, the experts, whose examinations were completed, sent their opinions on rubrics as specified, by filling out questionnaires. After receiving the form, the researcher determined a date with the experts and prepared the expert interview questions as a 15-minute interview over zoom, interviews were held with the parties and the records of the interviews were kept. By analyzing the questionnaire forms, the frequency distribution was examined and a table was created. The percentages of the answers given by the experts and the answers given to the open-ended questions were noted on the created table, and the thematic analysis of the interview records was made by the researcher, and the changes deemed necessary to be changed or added to the rubrics were made by the researcher on the draft rubric.

Rubrics, which constitute the second stage of the study, were planned to be used in writing exams, and rubrics were introduced via Zoom to teachers working at Bartın University TÖMER and who would evaluate their writing skill at the end-of-term level exams. Afterwards, they were asked to share their opinions about the rubrics they used after the evaluation with the researcher on the dates determined via Zoom in a 1-week period. After the study was done, the questionnaire form was sent to different teachers who are experts in their fields in order to reach more expert opinions and thoughts, and their valuable opinions were taken. In total, 30 experts in their fields evaluated these rubrics through a questionnaire. The data obtained as a result of the research were transferred to the SPSS-24.0 statistical package program and their frequency distributions were examined and analyzed. At the end of the examinations made in line with this questionnaire and the opinions, the rubrics were given their final form. Changes and additions made in rubrics in line with expert opinions are included in the findings section.

The comments and analyzes made by the experts during the interviews were examined by the researcher and presented in the findings section.

As the last step, the reliability analyzes of the rubrics were made, and they were scored by the researcher at the level of A1-A2-B1-B2-C1 by 6 instructors, who are experts in the field of 60 writing exam papers, 12 each. It is aimed to increase reliability and minimize bias. For this reason, the students were selected not from the language education centers where the experts were working, but from the students who were studying at the language education center of another university, and the same papers were evaluated by different evaluators. The increase in the number of raters allows performance studies to be evaluated in many aspects. At the same time, determining whether the feature to be measured exists in the individual based on the opinion of a single rater will reduce the reliability of the assessment (Mancar, 2019, p.20). After these scores, the exam papers were analyzed by using Kendall's coefficient of agreement method by looking at the agreement between the raters and the consistency between the raters was examined. The concordance numbers that emerged as a result of the analysis were interpreted in the findings section. Analysis and interpretation of research data is extremely important for research. For this reason, it is desired to mention the methods used in the data analysis process.

The Kendall W agreement coefficient determines the extent to which there is agreement between the scores given by multiple referees. The scores given by the referees or observers are in the nature of ordinal rubric data. Kendall's concordance coefficient was calculated. Kendall's W coefficient, which can take values between 0 and 1, shows that the value approaches 1, indicating that there is agreement between the raters (Howell, 2002).

Qualitative Research

In the first stage, a literature review was conducted. As a result of the literature review, various TÖMERs were contacted and information was obtained about how the writing skill was evaluated, and various data were collected. Research was carried out on how rubrics should be by determining the needs, some determinations were made in the light of the information obtained, and then the rubric development process was started. In the qualitative data collection process, data were collected by presenting interview questions developed by the researcher to experts in the field whose opinions were taken at the institutions. After deciding on the sample of the research, the target was determined about how many participants would be reached. Participants took place with 12 participants selected from different TÖMERs. Content analyzes were made by the researchers on the opinions obtained from the teachers.

Research Ethics

The ethics committee of the institution approved the study, and then the researcher informed the participants.

FINDINGS

In this section, the analysis of the findings obtained with the questionnaire form and interview forms applied to the instructors participating in the research, the thematic examinations of the teachers' opinions, the compatibility analyzes between the scores given by the evaluators were examined and analyzed with the SPSS-21 package program, and the validity and reliability analyzes were arranged and interpreted according to the sub-problems of the research.

What is the Validity and Reliability of the Rubrics?

Validity

Validity is one of the most basic and important concepts in testing and evaluation. This concept refers to the extent to which the measurement tool is consistent with the content of the field it is related to, and the degree to which it measures the desired feature and does this job without reflecting the effects of other features on the criteria (Ülper, 2008). Expert opinions were sought to determine the consistency of the rubric with each other and the validity of the rubrics developed. 30 participants who are experts in their fields were given questionnaires and their opinions were sought.

Expert opinions

Karasar (2002) explains content validity as making a decision about whether the area to be measured is represented or not, by consulting expert opinions on whether the measurement tool is suitable for the purpose for which it will be used. Opinions were received from 30 participants working in the field of teaching Turkish to foreigners in order to examine the adequacy of the rubric prepared in this study to develop a Rubric for Evaluating Writing Skill in Teaching Turkish as a Foreign Language. Participants evaluated the rubric through a questionnaire consisting of 13 rating questions consisting of 4 options (Totally Agree, Agree, Disagree, Strongly Disagree) and 3 open-ended questions aimed at learning their opinions, suggestions and criticisms.

The answers to the rating questions are summarized in Table 1.

Table 1. Frequency Distribution of Participants' Opinions (N=30)

Questions	Totally Disagree	Disagree	Agree	Totally Agree
	f (%)	F (%)	f (%)	F (%)
1. The criteria in the rubric are sufficient for evaluation.	0 (0,00)	3 (10,00)	11 (36,66)	16 (53,33)
2. The criteria measure writing skill.	0 (0,00)	2 (6,66)	16 (53,33)	12 (40,00)
3. The criteria are prepared to reflect the difference between the levels.	0 (0,00)	3 (10,00)	16 (53,33)	11 (36,66)
4. The criteria in the rubric are clear and understandable.	3 (10,00)	0 (0,00)	17 (56,66)	10 (33,33)
5. The explanations regarding the criteria definitions have been prepared in a way that accurately reflects the degrees.	0 (0,00)	2 (6,66)	16 (53,33)	12 (40,00)
6. No meaning deficiencies in items.	1 (3,33)	4 (13,33)	18 (60,00)	7 (23,33)
7. The content of each criterion is limited to its own purpose, not overlapping with other criteria.	0 (0,00)	4 (13,33)	13 (43,33)	13 (43,33)
8. The goals are mutually supportive. They do not contradict each other.	0 (0,00)	3 (10,00)	17 (56,66)	10 (33,33)
9. The rubrics are useful.	0 (0,00)	4 (13,33)	11 (36,66)	15 (50,00)
10. The criteria given in the rubrics are useful.	0 (0,00)	5 (16,66)	14 (46,66)	11 (36,66)
11. Criterion explanations are far from relative narration. It offers objectivity.	2 (6,66)	0 (0,00)	15 (50,00)	13 (43,33)
12. Criteria provides feedback.	0 (0,00)	2 (6,66)	13 (43,33)	15 (50,00)
13. There is harmony between the achievements to be measured and rubric.	0 (0,00)	1 (3,33)	17 (56,66)	12 (40,00)
Average (SS)	0,46 (0,93)	2,54 (1,50)	14,92 (2,24)	12,08 (2,34)
Percent Average (SS)	1,54 (3,10)	8,46 (4,99)	49,75 (7,45)	40,25 (7,79)

As seen in the table, 49.75% of the participants agreed with the evaluation questions, 40.25% stated that they completely agreed, 8.46% did not agree, and 1.54% disagreed at all. The fact that 90% of the participants agreed with the opinion that rubrics are an adequate, useful and objective assessment tool and that a high percentage of the participants also agreed with the items suggests that the developed rubrics are an adequate scoring key.

In line with the suggestions given by the participants in the open-ended questions where their opinions and suggestions were received:

In the A1 level rubrics, the content was edited, the genre information was reduced, 2 items were corrected for coherence and consistency, an "introductory part" was added to the content part, and expression corrections were made in the grammar part;

In the A2 level rubric, the expression "long and connected texts" has been removed, and "handwriting beauty" has been removed from the criteria;

In the B1-B2 level rubric, subjective and objective judgments were removed from the grammar part, an introduction and closing sentence were added, and

In the C1 level rubric, changes were made in the word part, the expression phrase was changed and the "figurative expression" in the grammar part was moved to the word part and the rubrics were given their final form.

Reliability

In order for a measurement tool to be qualified as reliable, different raters must have given similar scores in the measurements made. Whether there is a concordance between the scores given by different evaluators is revealed by concordance analysis.

In order to test the reliability of the scores obtained from the rubric, the consistency between the raters was also examined. It is recommended to look at the reliability coefficients between the raters in the analysis of the reliability of the scores to be obtained from such measurement tools in which multiple scoring will be made (Moskal & Leydens, 2000).

Table 2. Coefficients of Concordance Between Evaluators

Level	Kendall's W Coefficient
A1	.700
A2	.801
B1-B2	.828
C1	.807

Different methods can be applied in different situations when performing compliance reviews. In this study, Kendall's concordance coefficient was calculated to measure the concordance between raters. Kendall's W coefficient, which can take values between 0 and 1, shows that the value approaches 1, indicating that there is agreement between the raters (Howell, 2002).

If the calculated value approaches zero, it indicates the inconsistency between the raters. In order to test whether there is a harmony between the scores given using the Rubric, a total of 60 exam papers, 12 exam papers from A1, A2, B1, B2 and C1 levels, were scored by 6 different evaluators considering rubric. The Kendall W analysis results, which were carried out to measure the concordance between the raters, are given in Table 2, and as can be seen, the concordance values at all levels are above the acceptable value.

According to the table, the lowest fit is determined at the A1 level, while the highest fit is seen at the B1-B2 level. Kendall W fit coefficient takes values between 0 and 1. If Kendall's W coefficient is close to 1, it indicates that there is agreement between the raters, while the value approaching 0 indicates that the agreement between the raters is low (Howell, 2002).

Şencan (2005) states that the agreement between raters should be at least 0.80. The consistency coefficients obtained in this study approach 1 as shown in the table, and the findings obtained as a result of the analysis provide evidence for the reliability of the developed rubrics.

What are the Views and Practices of the Instructors on Using Rubrics?

The results of the data collected and analyzed from the interviews with the teachers after the use of the rubrics developed by the researcher in the exams are presented in this section. By analyzing the data obtained from the interviews with the teachers, it was tried to determine to what extent the rubrics created by the researcher supported the scoring of the writing exams and what the perceptions of the teachers who were experts in their fields were. The interviews with the teachers were analyzed qualitatively by categorizing them around the research questions. Categories are large units of information containing many codes of common ideas. These can be thought of as basic patterns, findings, or abstractions that emerge in response to research questions. (Çelik et al, 2020) The findings obtained as a result of the thematic analyzes of the teacher interviews are presented in the table below.

Table 3. Teachers' views on the rubric

Themes	f
Providing the standard in measurement	7
Ease of scoring	5
Identifying the source of the error	5

Providing Standards in Measurement

Creating a standard is easier and more useful in terms of testing and evaluation. In order to make the assessment and evaluation process more functional and to obtain more accurate feedback, measurement tools with high validity and reliability are needed.

The participants of the research stated that the writing success with the developed rubrics has an effect on more objective and objective evaluation. Some of these views are presented below:

“Since the scoring rubric is standard and distinctive, I think that I made a more objective assessment compared to other assessments. (T7)

Another teacher said:

“I think rubrics provide standardized assessments, otherwise I feel uncomfortable and think that there should be at least two assessors when assessing rubrics as well as writing exams. For this reason, this study is more objective than the traditional evaluations of teachers. I applied it in the exam and saw that it was true.” (T3)

Another teacher said:

“I used it in the exam, it was very clear and clear, I think I will use this rubric in my next exams, one of the aspects I liked the most is that it is not complicated, some criteria in the rubric we used before were open to interpretation, but this rubric was very clear, I think I made a very objective assessment after this exam” (S5)

The findings obtained from the teachers' opinions are that the rubrics developed do not reveal clear differences, and that the scores given by the teachers who made the ratings make them think that the scores they give are more objective. Considering these findings, it can be concluded that although the developed do not fully provide objectivity, they provide the standard to a large extent compared to traditional evaluation methods.

Ease of Scoring

One of the things that teachers give the most importance when scoring exams is the concept of time. It is very important for teachers that the scoring rubric used are functional and facilitate scoring. For this reason, scoring tables should provide convenience.

The participants stated that the rubric, which was developed based on the fact that the objectives were appropriate for the level and the criteria were defined in detail, provided ease of use. Some opinions on the subject are presented below:

“The achievements are appropriate for the level, the criteria are detailed and beautiful, the common use of the B1-B2 level has made the scoring keys easier in terms of usability” (T2)

“In general, it is good that it is measured in terms of criteria and scoring, that each criterion is consistent within itself and corresponds to a score, and that maximum detail is provided, and that the achievements are compatible with the criteria has made the scoring rubric more functional” (T1)

“It was very practical, I scored quickly, it helped me a lot.” (T3)

“One of the most important things for a teacher is that the assessment tool to be used is practical and useful, I think it is a study that provides this, it is clear that it has been studied, the distinctions are very clear, I think this is a positive thing, when everything is clear, scoring is done faster” (T4)

“There are a lot of deficiencies in the field of Turkish, especially the standard rubrics are almost non-existent, I thought it was necessary to make rubrics for other skill along with this study, because I personally found it very practical for me. Having everything in its place made it easy to find and write the score and calculate it when I searched.” (T2)

Identifying the Source of the Error

The developed rubrics should be useful for both the teacher and the student. Rubrics is a tool that both teachers and students can use. The use of rubrics by students allows students to gain awareness and evaluate themselves.

Participants in the research stated that scoring with rubric allows to see the sources of error clearly. Some opinions on the subject are presented below:

“Before, I used to score the exams myself without being tied to a ruler, sometimes we have rulers, but most of the time I find it superficial, but this study is successful in enabling the student to see clearly the mistake he made.

"Having a scoring rubric was good not only for us but also for the students, I saw better where the student was missing and helped the student to see it more clearly after the evaluation, so I support such studies "(T7)

It was concluded that the scoring keys developed based on the interviews with the teachers after the exam had content validity, that the rubrics were prepared in accordance with the levels, and that they measured the desired achievements..

CONCLUSION

In this study, in order to evaluate the written expression skill of students in teaching Turkish as a foreign language, an analytical rubric from A1 level to C1 level based on writing skill was developed and the validity and reliability proofs of the rubrics were revealed. In this research, after the drafts of the scoring keys were prepared, they were presented to the expert opinion and necessary arrangements were made in line with the opinions. Within the scope of the validity study of the scoring key, the content validity was examined through expert evaluation questionnaires and the frequency distribution percentages of the answers given by the experts were examined. Developed rubrics include 5 criteria to measure written expression skill in accordance with all levels. For each rubric, these criteria include level achievements and criteria that should be in written expression. The percentages of agreement determined by the experts of this study were determined by the validity frequency of 90% for each rubric. Having a high percentage of participants stating that they agree with all items proves that the developed rubrics are an adequate scoring key. It has been proven in the light of expert opinion that the criteria and the explanations regarding the criteria are sufficient and appropriate. In the reliability coefficient calculations, it was determined that the inter-rater reliability numbers of the rubrics used in scoring the written expression skill were A1-.700, A2-.801, B1-B2-.828, C1-.807.

Validity Findings and Validity Discussion

- The validity analyzes of the research were determined by looking at the frequency ranges of the survey results of 30 participants who are experts in the field. 49.75% of the participants stated that they agreed with the evaluation questions, 40.25% stated that they completely agreed, 8.46% did not agree, and 1.54% did not agree at all. He stated that 90% of the participants agreed with the opinion that rubrics are an adequate, useful and objective assessment tool.

- Consistency between expert opinions constitutes proof of content validity of relevant rubrics. Yorgancı and Baş (2021) developed a rubric for B1 level writing skill. The validity proofs of the developed rubrics were provided by expert opinions on criteria and sub-descriptors. The validity analyzes of this research were also provided by expert evaluation questionnaires and expert interview forms. In the results of these two rubric development studies, the suitability and adequacy of the criteria and definitions were approved by the experts. Again, in the validity analyzes of the scoring key developed by Demir (2020) for the evaluation of academic writing skill, the agreement percentages were determined between 75% and 100% and it was concluded that the rubrics developed were valid. In another rubric development study to evaluate written expression skill, the validity findings were obtained by taking the opinions of the experts in terms of relevance for the purpose, and it was determined that the percentage of agreement between the opinions of the experts varied between 87% and 100%. These calculated compliance percentages were accepted as proof of the content validity of the developed rubric (Demir & Yıldırım, 2019).

Reliability Findings and Reliability Discussion

- Inter-rater reliability coefficient was calculated using Kendall' W. In the Kendall W fit test, the fit between the measurement values of the data is checked. When the value approaches 1, it is an indication that the harmony between the evaluators or the referees has increased and they have reached a common opinion (Karagöz, 2019). In the results of the reliability analysis of this study, it was determined that the inter-rater reliability numbers of rubrics were A1-.700, A2-.801, B1-B2-.828, C1-.807. The high reliability numbers for each rubric proves that it is a reliable measurement tool.

- Yorgancı and Baş (2021), developed a rubric for B1 level writing skill for learners of Turkish as a foreign language, and they calculated the reliability analyzes of the study with correlation and confidence intervals between raters. The inter-rater consistency coefficient was calculated as 0.98. The developed rubric is considered reliable. The reliability coefficients for each rubric of this study are above the coefficient determined in Yorgancı and Baş's study.

- In the study of Erman Aslanoğlu and Kutlu (2003) on the use of rubrics in the evaluation of presentation skill, teachers and audience students were given rubrics for the evaluation of the presentations and scored. Kendall

W test was applied to check the consistency between the scores. As a result, the concordance value of the scores given by the teachers was .83, and the concordance between the scores given by the audience students was .42. The consistency between the scores given by the teachers was considered high. It was concluded that they are compatible with each other. However, it was concluded that the audience students were less in agreement with each other.

- The rubrics developed according to all these findings prove to be a tool that can measure the writing skill of students learning Turkish as a foreign language. Polat Demir (2020) analyzed the reliability of the analytical rubric he developed for the evaluation of academic writing skill with the same method and reached similar results.

- Among the results of the study, it can be interpreted based on the findings that analytical rubrics do not completely eliminate the inconsistency between the raters in the evaluations but increase the consistency between the raters' scores and make the evaluations more objective. These results are similar to the study of Hızarcıoğlu (2013), which supports the opinion that there is a harmony, although not high, between the scores made by teachers without using rubric in his master's thesis, but this agreement is lower than the use of rubric, and that more objective evaluation is made with rubric. arrives. Studies conducted on this subject (Anıl, 2015; Bıkmaz Bilgen & Doğan, 2017) show that evaluations made with rubrics increase reliability, and analytical rubrics, which provide more standard and more objective results among raters, provide more consistent scoring, thus increasing reliability.

Statements of Publication Ethics

This study is based on the master's thesis of the first author. Therefore, ethical approval was obtained.

Conflict of Interest

This study is based on the master's thesis of the first author. The authors contributed equally.

REFERENCES

- Aktaş, M., & Buyer, D. (2018). Analytical Rubric Development for Evaluating the Written Story: Validity and Reliability Study. *Journal of Mersin University Faculty of Education*, 14 (2), 597-610.
- Ari, G. (2010). Evaluation of Storytelling Texts Written by Sixth and Seventh Grade Students. *Turkology Studies*, 43-75.
- Arslan, M., & Klicic, E. (2015). Problems Encountered in the Development of Writing Skills in Teaching Turkish as a Foreign Language: The Case of Bosnia and Herzegovina. *Kırıkkale University Journal of Social Sciences*, 5 (2), 169-182.
- Atılğan, H., Kan, A., & Doğan, N. (2007). *Measurement and Evaluation in Education*. Ankara: Memoir Publishing.
- Europe, K. (2013). *European common recommendations framework for languages: learning, teaching and assessment*. Germany: TELC GmbH.
- Azizoğlu, N. İ., Demirtaş Tolaman, T., & İdi Orange, F. (2019). Academic Writing Skills in Teaching Turkish as a Foreign Language; Encountered Problems and Solution Suggestions. *International Journal of Turkish Teaching as a Foreign Language*, 2(1),7-22.
- Barin, E. (2004). Principles in Teaching Turkish to Foreigners. *Hacettepe University Turkic Studies (HUTAD)*, (1), 19-30.
- Barin, E. (2009). The importance of dictation and written expression in teaching Turkish to Turkish children abroad. *TDAY Belleten*, 1.21-32.
- Bıkmaz Bilgen, Ö., & Doğan, N. (2017). Comparison of Inter-rater Confidence Determination Techniques. *Journal of Measurement and Evaluation in Education and Psychology*, 8 (1) , 63-78.
- Boylu, E. (2015). Levels of Use of Language Learning Strategies by Learners of Turkish as a Foreign Language. Master Thesis. Ankara: Gazi University Institute of Educational Sciences.
- Boylu, E. (2019). Assessment and evaluation practices and setting standards in teaching Turkish to foreigners. Doctoral Thesis. Çanakkale: Çanakkale Onsekiz Mart University, Institute of Educational Sciences, Department of Turkish and Social Sciences Education.
- Büyükkiz, K. (2012). Development of the Writing Self-Efficacy Scale for Foreigners Learning Turkish as a Second Language: A Study of Validity and Reliability. *Journal of Mustafa Kemal University Institute of Social Sciences*, 9(18), 69-80.
- Büyükoztürk, Çakmak, Akgün, Karadeniz, & Demirel. (2019). *Scientific Research Methods*. Ankara: Pegem Academy.
- Buyukozturk, S. (2007). *Manual of data analysis for social sciences*. Ankara: Pegem Publishing.

- Cakir, I. (2010). Why Is It Difficult to Acquire Writing Skills in Foreign Language Teaching? *Erciyes University Journal of Social Sciences Institute*, 1 (28), 1-12.
- Cakmak Akata, A. (2009). A Study on the Functionality of the Assessment and Evaluation Process of the Turkish Curriculum (Example of Tekirdağ Province). Bolu: Abant İzzet Baysal University, (Unpublished, Master's Thesis).
- Çelik, H., Baykal, N. B., & Kılıç Memur, H. N. (2020). Qualitative Data Analysis and Its Basic Principles. *Journal of Qualitative Research in Education – Journal of Qualitative Research in Education*, 8(1), 379-406.
- Cetin, B. (2002). The Relationship of Scores Estimated from the Formal Features of Composition in Composition Type Exams with the Scores Obtained from Scoring with Key and General Impression. Ankara: Hacettepe University Institute of Social Sciences, (Unpublished) Master's Thesis.
- Demir, B. P. (2020). Development of Analytical Rubric for Evaluation of Academic Writing Skills. *Amasya University Journal of Education Faculty*, 9 (2), 318-327.
- Deniz, H., & Demir, S. (2019). Development of the Writing Anxiety Scale: Reliability and Validity Study. *International Journal of Turkish Literature, Culture and Education*, 8(2), 1034-1051.
- Deniz, H., & Demir, S. (2019). Development of the Writing Anxiety Scale: Reliability and Validity Study. *International Journal of Turkish Literature, Culture and Education (TEKE)*, 8 (2).
- Eran, İ., & Kan, İ. (2004). Reliability and validity in scales. *Journal of Uludag University Faculty of Medicine*, 30(3), 211-216.
- Gedik, D. (2009). Teaching Turkish to Foreigners: The Case of Ankara University TÖMER and Gazi University TÖMER. Bolu: Abant İzzet Baysal University Social Sciences Institute, Unpublished Master's Thesis.
- Gedik, E. (2017). Assessment and Evaluation in Teaching Turkish as a Foreign Language. Istanbul: Istanbul Arel University, Institute of Social Sciences, Master Thesis.
- Goodrich, H. G. (1997). Understanding Rubrics. *Educational Leadership*, 54(4).
- Gocer, A. (2007). Investigation of Textbooks Used in Teaching Turkish as a Foreign Language in Terms of Assessment and Evaluation. *Language Journal*, 137.30-46.
- Gocer, A. (2011). Evaluation of students' written expression studies by Turkish teachers. *Ondokuz Mayıs University Journal of the Faculty of Education*, 30(2), 71- 97.
- Göçer, A. (2019). The Use of Feedback as an Important Practice Technique Affecting the Functionality of Assessment and Evaluation in Turkish Education. *Kırıkkale University Journal of Social Sciences*, 9 (1), 111-126.
- Göçer, A., & Mongolian, S. (2011). Related to Teaching Turkish as a Foreign Language. *International Periodical for The Languages, Literature and History of Turkish or Turkic*, 6(3), 797-810.
- Güdek, B., & Öztürk, D. (2016). The Effect of Self-Assessment Practices in Cello Teaching on Students' Performance and Attitudes. *Turkish Studies*, 11(3), 1149-1162.
- Gunay, V. D. (2007). *Text Information*. Istanbul: Multilingual Publications.
- Gunduz, S. (2003). *The Art of Writing Stories and Novels*. Istanbul: Toroslu Library.
- Güneş, F. (2007). Turkish Teaching and Mental Structuring. Ankara: Nobel Publishing House. Gurses, R. (2006). The Place of Assessment and Evaluation in Language Teaching. *Turkish Language Studies Yearbook-Belleten*, 54, 131-154.
- Hizarcioglu, B. O. (2013). Investigation of the Consistency of Using a Rater in the Problem-Solving Process. Bolu: Abant İzzet Baysal University, Institute of Educational Sciences, Master Thesis.
- Howell, D. C. (2002). *Statistical methods for psychology (5th ed.)*. Pacific Grove CA: Duxbury.
- Isisak, K. U., & Demirel, O. (2010). Using the Common European Framework of Reference for Languages in the Development of Speaking Skills. *Education and Science*, 35, 156.
- Kahveci, N., & Şentürk, B. (2021). A Case Study on the Evaluation of Writing Skill in Teaching Turkish as a Foreign Language. *International Journal of Education, Technology and Science*, 1(4), 170–183.
- Kalfa, M. (2014). Developing Basic Level Students' Writing Skills with Educational Games in Teaching Turkish to Foreigners. *Hacettepe University Turkic Studies*, 11(85).
- Karagöz, Y. (2019). *SPSS - AMOS - META Applied Statistical Analysis*. Ankara: I Nobel Academic Publishing Education Consultancy Tic. Ltd. Sti.
- Karasar, N. (2003). *Scientific Research Method*. Ankara: Nobel Publication Distribution.

- Korkmaz, Y. (2009). The Effect of Rubric Use Education in Science Teaching on Teachers' Views and Practices Regarding Measurement and Evaluation. Konya: Selcuk University, Institute of Science, Department of Primary Education, Science Education Program, Master Thesis.
- Kürşat Yorgancı, O., & Baş, B. (2021). A Rubric Development Study for B1 Level Writing Skills for Learners of Turkish as a Foreign Language. *Literacy Education Studies*, 9 (1), 67-80. DOI: 10.35233/oyea.934684.
- Maden, S., Dincel, Ö., & Maden, A. (2015). Learning Turkish as a Foreign Language Learners' Writing Anxiety. *International Journal of Turkish Literature, Culture and Education (TEKE)*, 4 (2), DOI: 10.7884/teke.488.
- Maltepe, S. (2006). Evaluation of Writing Processes and Products in Turkish Lessons in terms of Creative Writing Approach. Ankara: Ankara University/Institute of Educational Sciences, Doctoral Thesis (Unpublished).
- Mancar, S. A. (2019). Comparison of Inter-rater Reliability Techniques in Performance Based Case Determination. Ankara: Ankara University, Institute of Educational Sciences, Master Thesis.
- Melanlıoğlu, D. (2016). Metacognitive Rubric for Listening Skills for Foreign Learners. *Erzincan University Journal of Education Faculty*, 18(2),1206-1229.
- Moskal, B. M., & Leydens, J. A. (2000). Scoring Rubric Development: Validity and Reliability. *Practical Assessment, Research & Evaluation*, 7(10). 71-81.
- Öncü, H. (1994). *Measurement and Evaluation in Education*. Matser Basım San. and Tic. Ltd. Sti.
- Ozbay, M. (2009). *Listening Education as a Language Skill*. Pegem Academy Publishing.
- Ozbay, M. (2011). *Writing Training*. Pegem Academy.
- Özbay, M., & Melanlıoğlu, D. (2009). Evaluation of Idioms in Turkish Education in terms of Teaching and Learning Process. *National Education*, 38(181), 8 - 19.
- Ozcelik, D. (1992). *Quantification and consideration*. OSYM.
- Ozcelik, D. A. (2016). *Quantification and consideration*. Pegem Academy.
- Ozenc, M. (2013). Determination of Alternative Assessment and Evaluation Knowledge Levels of Classroom Teachers. *Journal of Dicle University Ziya Gökalp Education Faculty*, 21,157-178.
- Ozkan, U. B. (2019). *Analysis Method for Educational Science Research*. Pegem Academy.
- Öztürk, D., & Güdek, B. (2016). Development of Rubric for Cello Performance Evaluation. *Journal of Academic Music Research*, 2(3),6.
- Parlak, B., & Dogan, N. (2014). From Rubric and Rubric Obtained. *Hacettepe University Faculty of Education Journal*, 29(2), 189-197.
- Razon, N. (1982). Reading Difficulties. *International Journal of Education and Science*, 19-29. Richard, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, 17 (2), 219-240.
- Selvikavak, E. (2006). An Application on Developing Paragraph Writing Skills of Advanced Students in Teaching Turkish as a Foreign Language. Ankara: Ankara University, Institute of Social Sciences, Linguistics Department, Master Thesis.
- Sever, S. (2015). *Turkish teaching and full learning*. Ankara: Memoir Publishing.
- Sonmez, V. (2003). *Introduction to the Teaching Profession*. Ankara: Memoir Publishing.
- Şen, Ü., & Boylu, E. (2017). Writing Anxiety Scale for Learners of Turkish as a Foreign Language. *International Journal of Turkish Literature, Culture and Education*, 6(2), 1122-1132.
- Şencan, H. (2005). *Reliability and Validity in Social Behavioral Measurements*. Ankara: Seçkin Publishing.
- Sengul, M. (2011). Primary Education II. Evaluation of the Effectiveness of Teaching and Assessment-Evaluation Approaches for Writing Skills in Secondary Turkish Teaching. Elazığ: Fırat University, Institute of Educational Sciences, Department of Turkish Education, Doctoral Thesis.
- Simsek, P. (2017). In Teaching Turkish as a Foreign Language, the Effects of Students' Mother Languages and Their Language Families on their Turkish Writing Skills. Ankara: Gazi University, Institute of Educational Sciences, (Unpublished) Doctoral Thesis.
- Takil, N. B. (2014). The Effect of Communicative Approach on Writing Skills in Teaching Turkish as a Foreign Language. Ankara: Gazi University, Institute of Educational Sciences, Doctoral Thesis.
- Tekin, H. (2003). *Measurement and Evaluation in Education*. Ankara: Judicial Publications: 2003.

- Temizkan, M. (2003). An Evaluation of Turkish Teachers' Studies in the Framework of Written Expression Activity. Hatay: Mustafa Kemal University Institute of Social Sciences, Unpublished Master's Thesis.
- Tiryaki, E. N. (2013). Writing Education in Teaching Turkish as a Foreign Language. *Native language Education Journal*, 1(1), 38-44.
- Tok, M. (2013). The Need for Academic Writing in Teaching Turkish as a Foreign Language. *Journal of Mustafa Kemal University Institute of Social Sciences*, 10(23), 1-25.
- Tok, M., & Potur, O. (2015). Trends of academic studies in the field of writing education (2010 - 2014) . *Journal of Mother Tongue Education*, 3 (4), 1-25.
- Turgut, F., & Baykul, Y. (2010). *Measurement and Evaluation in Education*. Ankara: Pegem Academy.
- Ulper, H. (2008). A Comparative Study on the Validity of Different Assessment (Grading) Approaches Used in Measuring Writing Skills (Grading) in the Context of the Purposes of Writing Essays. Ankara: 1st National Congress of Measurement and Evaluation in Education and Psychology.
- Unsal, G. (2008). Writing Teaching. *Language Journal*, (142), 46-60. DOI: 10.1501/Dilder_0000000102.
- Wach, E., & Ward, R. (2013). Learning about Qualitative Document Analysis. Institute of Development Studies.
- Yıldırım, A., & Şimşek, H. (2011). *Qualitative research methods in the social sciences*. Ankara: Seçkin Publishing.
- Yorgancı, O. K., & Baş, B. (2021). A Rubric Development Study for B1 Level Writing Skills for Learners of Turkish as a Foreign Language. *Literacy Education Research*, 9 (1).
- Zorbaz, K. Z. (2013). Scoring of written expression. *Abant İzzet Baysal University Journal of Education Faculty*, 13(1), 179 - 193.

APPENDIX

A1 SEVİYESİ YAZMA BECERİSİ DERECELİ PUANLAMA ANAHTARI

	Puan	Ölçüt
İçerik ve tür	6	ÇOK İYİ: Fikirler verilen konuya göre geliştirilmiş, konuya hâkim, metin giriş cümlesine sahip
	4	İYİ: Bazı detaylar atlanmasına rağmen konuyla ilgili büyük oranda bilgi aktarılmış, fakat yeterli ayrıntıya yer verilmemiş, giriş cümlesi olabilecek bir cümleye sahip
	2	ORTA: Verilen konuyla ilgili sınırlı bilgi verilmiş, ana fikir kısıtlı olarak geliştirilmiş
	1	YETERSİZ: Metin oluşturmak için yetersiz
Metin planı- cümleler Arası düzen	3	ÇOK İYİ: - paragraf planı uygulanmış, paragraf düzeni çok iyi-Düzeğe uygun bağlaçlar kullanılmış-başlık konuyla ilgili etkili biçimde oluşturulmuş
	2	İYİ: zaman ifadeleri az tekrarlar ile kullanılmış- paragraf düzeni var, başlık ve paragraf başları belirtilmiş, başlık kullanılmış, seviyeye uygun bağlaç kullanımı yapılmış- (“ile ve” vb. Basit düzey bağlaçlar.)
	1	ORTA: - Cümleler arası yoğun tekrarlar yapılmış fakat sınırlı sayıda bağlaç kullanılmış- (“ile ve” vb. Basit düzey bağlaçlar.), başlık kullanılmış
	0,5	YETERSİZ: paragraf bütünlüğü yok-bağlaçlar metinde kullanılmamış, paragraf düzeni (paragraf başı -sonu) oluşturulmamış, metin oluşturmak için yetersiz
Tutarlılık ve bağdaşıklık	5	ÇOK İYİ: İstenilen türe uygun tutarlı içerik oluşturulmuş, biçimsel olarak seviyeye uygun zamansal değişim ve gelişim bağlaçları etkili şekilde kullanılmış-sıralama bağlantıları etkin kullanılmış (ve ile, âmâ)-anlam bütünlüğü tam
	3	İYİ: Metin yeni bilgiler ve destekleyici açıklamalarla geliştirilmiş-anlam bütünlüğü sağlanmış-konu dışına çıkılmamış- tekrarlar az yapılmış-
	2	ORTA: Verilen konuya göre basit şekilde zamana bağlı gelişim ve değişim ifadeleri kullanılmış (önce, sonra, daha sonra)- Metnin türüne uygun mantıksal öğeler kullanılmış fakat çelişkili ifadeler ve tekrarlar sıklıkla yapılmış
	1	YETERSİZ: Metin oluşturmak için gerekli biçimsel ve mantıksal bağlantılar kurmakta yetersiz.
Dil bilgisi yapılarını kullanma	7	ÇOK İYİ: Basit yapı cümleler metinde etkin şekilde kullanılmış-fiil çekimleri doğru yapılmış-ses olayları doğru yazılmış- edatlar ve zamirler etkili ve hatasız kullanılmış-özne ve yüklem uyumludur- dil bilgisi yapıları etkin şekilde kullanılmış
	5	İYİ: Basit yapı cümleler kurulmuş- zamanlar seviyeye uygun şekilde çekimlenmiş, A1 seviyesine uygun edatlar ve zamirler bazı hatalarla beraber kullanılmış, özne ve yüklem uyumludur- seviyeye uygun zamanlarla, zaman ifadeleri uyumlu kullanılmış (yarın, dün, şimdi)
	3	ORTA: Özne -yüklem uyumlarında yoğun hatalar yapılmış. Cümle yapısında belirgin hatalar yapılmış-fiil çekimleri doğru yapılmış-cümle dizilişinde basit hatalar yapılmış-dil bilgisi unsurları genellikle yanlış yazılmış
	1	YETERSİZ: Dil bilgisi ve cümle yapısı metin oluşturmak için yetersiz- metinde anlatılmak istenen anlaşılmamaktadır- cümle ve paragraf oluşturmak için gerekli hâkimiyet kurulmamış
Söz varlığı ile Yazım ve Noktalama	4	ÇOK İYİ: Yazım ve imla kuralları iyi derecede kullanılmış-hatasız-paragraf başları ve büyük-küçük harf kullanımında hata yok- zıt anlamlı kelimeler metinde etkili kullanılmış-tekrarlara düşülmemiş, bağlama uygun kelimeler etkili kullanılmış, beklenen sayıda ve üstünde sözcük kullanılmış
	3	İYİ: Yazım ve noktalamada ufak hatalar yapılmış fakat metin oluşturulmuş, paragraf başları belirgin-büyük, küçük harf yazımlarında az hata yapılmış- Seviyeye uygun kelimeler metinlerde kullanılmış- kelime bilgisi paragraflar için yeterli, beklenen sayıya yakın sözcük kullanılmış
	2	ORTA: “nokta, virgül, ünlem, soru işareti” gibi temel noktalama işaretlerinde hatalar yapılmış- büyük, küçük harf yazımına dikkat edilmemiş-paragraf başları belirtilmemiş- kelime ve ifadelerin seçimi ve kullanımında sık sık hatalar yapılmış, beklenen sayının yarısı kadar sözcük kullanılmış
	1	YETERSİZ: Yazım ve imla kuralları kullanılmamış- büyük/küçük harf, paragraf başı kullanımında yoğun hatalar yapılmış-değerlendirmek için yeterli değil- Seviyeye uygun yeterli kelime bilgisine sahip değil-sözcük sayısı yetersiz
GEÇERSİZ	1	* Konu ve/veya türden bağımsız: verilen soruyu anlamakta ve/veya tamamlamakta başarısız
NOT Yönergede verilen kelime limitine göre	0	-%20
	1	+%20

A2 SEVİYESİ YAZMA BECERİSİ DERECELİ PUANLAMA ANAHTARI

	Puan	Ölçüt
İçerik ve tür	6	ÇOK İYİ: Fikirler verilen konuya ve türe göre genişletilmiş, metin giriş ve kapanış cümlesine sahip
	4	İYİ: Bazı detaylar atlanmasına rağmen fikir verilen konu ve tür göre genişletilmiş, kapsamlı, metin giriş cümlesine sahip
	2	ORTA: Verilen konuyla ilgili sınırlı bilgi verilmiş, tür ve konu desteklenmemiş, genişletilememiş
	1	YETERSİZ: Verilen konuyla ilgili içerik ve tür gelişmemiş, sınırlı bilgi var değerlendirmeye uygun değil
Cümleler arası plan- düzen	3	ÇOK İYİ: Cümleler arası geçişler etkili yapılmış, giriş-gelişme-sonuç gibi paragraf planı etkili şekilde uygulanmış, bağlaçlar etkili şekilde kullanılmış-paragraf bütünlüğü ve düzeni yapılmış, başlık oluşturulmuş
	2	İYİ: sebep sonuç bağlaçları kullanılmış-zaman ifadeleri az tekrarlar ile kullanılmış-cümle geçişleri ve paragraf düzeni etkin şekilde oluşturulmuş- başlık ve paragraf başları belirtilmiş, başlık oluşturulmuş
	1	ORTA: Paragraflar arası geçişler bazı hatalarla beraber yapılmış- cümleler arası yoğun tekrarlar yapılmış fakat sınırlı sayıda bağlaç kullanılmış- (“ile ve” vb. Basit düzey bağlaçlar.) başlık kullanılmamış
	0,5	YETERSİZ: paragraf bütünlüğü kurulmamış- bağlama bağlaçları kullanılmış, paragraf düzeni (paragraf başı -sonu) oluşturulmamış, seviyeye uygun metin oluşturmak için yetersiz
Tutarlılık ve bağdaşıklık	5	ÇOK İYİ: İstenilen türe uygun tutarlı ve uyumlu içerik oluşturulmuş, biçimsel olarak dilsel bütünlük sağlanmış- cümleler arası anlamsal ve biçimsel bağlantı oluşturulmuş, verilen içerikte kronolojik sıra mantıksal çerçevede oluşturulmuş, karşılaştırma bağlaçları hatasız ve etkili şekilde kullanılmış, anlam bütünlüğü tam
	3	İYİ: içerik ve türe uyumlu ifadelerle metin desteklenmiş- neden-sonuç bağlantıları etkin kullanılmış (bu yüzden, bu sebeple vb.)
	2	ORTA: Cümleler arası biçimsel bağlantı tam sağlanmamış -metnin türüne uygun mantıksal öğeler kullanılmış fakat çelişkili ifadeler yer verilmiş ve tekrarlar sıklıkla kullanılmış-neden-sonuç bağlantıları anlaşılır değil
	1	YETERSİZ: Metin oluşturmuş fakat seviyeye uygun gerekli biçimsel ve mantıksal bağlantılar kurmakta yetersiz.
Dil bilgisi yapılarını kullanma	7	ÇOK İYİ: Seviyeye uygun bağlaçlar ve edatlar (gibi, kadar)cümlede etkin şekilde kullanılmış, fiil çekimlerinde hata yapılmamış,birleşik yapı cümleler kullanılmış (fiilimsi kullanımları yapılmış), seviyeye uygun zaman çekimleri doğru kullanılmış,
	5	İYİ: fiilimsi kullanımı ile birleşik yapı cümleler yazmaya çalışılmış, zaman zarfları etkin ve doğru kullanılmış, edatlar ve bağlaçlar (gibi, kadar, vb.) az hata ile sıklıkla kullanılmış, cümleler seviyeye uygun zamanlar ile kurulmuş
	3	ORTA: Az sayıda birleşik cümle oluşturulmuş, fiilimsiler yanlış kullanılmış, edat ve bağlaç kullanımı seviyeye göre az kullanılmış, Hatalar metni anlamayı zorlaştırmış
	1	YETERSİZ: Dil bilgisi kullanımı cümle oluşturmak için yeterli değil, yoğun hatalar yapılmış, seviyeye uygun zaman kullanımı yapılmamış, metin oluşturulmuş fakat seviyeye uygun dil bilgisi yapıları kullanılmamış
Söz varlığı ile Yazım ve noktalama	4	ÇOK İYİ: Yazım ve imla kuralları etkin kullanılmış -paragraf başları ve büyük-küçük harf kullanımı hatasız oluşturulmuş - kelime tekrarlarına düşülmemiş, kelimelerin farklı anlamları metinde kullanılmış, beklenen sayıda ve üstünde sözcük kullanılmış
	3	İYİ: Yazım ve noktalamada bazı hatalar yapılmış, -paragraf başları belirgin şekilde belirtilmiş-büyük, küçük harf yazımlarında bazı hatalar yapılmış-A2 seviyesinde duruma ve zamana uygun kalıplar kullanılmış -kelime bilgisi düzeye göre yeterli, -az hata ile uzun metin yazılmış, beklenen sayıya yakın sözcük kullanılmış
	2	ORTA: “nokta, virgül, ünlem, soru işareti” gibi temel noktalama işaretlerinde hatalar yapılmış- büyük, küçük harf yazımında yoğun hatalar yapılmış- Seviyeye uygun temel kelimeler kullanılmış-sınırlı kelime kullanılmış- fiil tekrarlarına düşülmüş-kelime bilgisi uzun paragraflar için yeterli değil, beklenen sayının yarısı kadar sözcük kullanılmış
	1	YETERSİZ: Yazım ve imla kuralları kullanılmamış- büyük/küçük harf, paragraf başı kullanımında yoğun hatalar yapılmış- değerlendirmek için yeterli değil, sözcük sayısı yetersiz
GEÇERSİZ	1	* Konu ve/veya türden bağımsız: verilen soruyu anlamakta ve/veya tamamlamakta başarısız
NOT Yönergede verilen kelime limitine göre	0	-%20
	1	+%20

B1-B2 SEVİYESİ YAZMA BECERİSİ DERECELİ PUANLAMA ANAHTARI

	Puan	Ölçüt
İçerik ve tür	6	ÇOK İYİ: Verilen konu bağlama ve türe uygun eksiksiz yazılmış, düşünceyi geliştirme teknikleri metinlerde kullanılmış, yeterli ayrıntıya yer verilmiş, düşünce yeterince desteklenmiş, Metinde giriş ve kapanış cümlesi kullanılmış
	4	İYİ: Fikir geliştirilirken bazı hatalar yapılmış, verilen düşünce içerik ve tür etrafında etkili planlanıp, düzenlenmiş, Metninde giriş ve kapanış cümlesi olabilecek cümlelere yer verilmiş
	2	ORTA: Verilen konuyla ilgili fikir desteklenmemiş, tür ve içerik uygun fakat geliştirilmemiş
	1	YETERSİZ: Verilen konuyla ilgili seviyeye uygun destekleme ve geliştirme yapılmamış, sınırlı bilgi verilmiş, değerlendirmek için yetersiz
Cümleler arası plan-düzen	3	ÇOK İYİ: Giriş-gelişme-sonuç gibi paragraf planı etkili şekilde uygulanmış, bağlaçlar etkili şekilde kullanılmış-paragraf bütünlüğü ve düzeni yapılmış, bağlama uygun başlık oluşturulmuş.
	2	İYİ: geçişler ve paragraf düzeni etkili şekilde yapılmış, -başlık ve paragraf başları belirtilmiş, bağlama uygun başlık oluşturulmuş.
	1	ORTA: Paragraflar arası geçişlerde bazı hatalar var- cümleler arası geçişlerde yoğun hatalar var, başlık ve paragraf başları belirtilmiş.
	0,5	YETERSİZ: paragraf bütünlüğünde hatalar var-başlık ve paragraf başları belirtilmemiş, seviyeye uygun plan-düzen oluşturulmamış, değerlendirmek için yeterli değil.
Tutarlılık ve bağdaşıklık	5	ÇOK İYİ: İstenilen türe uygun tutarlı ve çelişkili olmayan içerik oluşturulmuş, biçimsel olarak dilsel bütünlük sağlanmış-, karşılaştırma, pekiştirme bağlaçları hatasız ve etkili şekilde kullanılmış- cümleler arası ve paragraflar arası biçimsel ve anlamsal bağlantı var, anlam bütünlüğü tam.
	3	İYİ: içerikle çelişkili olmayan ifadeler kullanılmış, metin türe uygun yazılmış, düşünceyi geliştirme yolları metinde kullanılmış, cümleler arası biçimsel olarak dilsel bütünlükte bazı hatalar var, karşılaştırma, pekiştirme bağlaçları az hata ile metinde kullanılmış, bağlama uygun kelime seçimleri yapılmış.
	2	ORTA: Biçimsel olarak seviyeye uygunluk tam sağlanamamış-metin türüne uygun mantıksal öğeler kullanılmış fakat çelişkili ifadeler ve tekrarlar sıklıkla yapılmış-bağlaçlar yoğun hatalar ile kullanılmış
	1	YETERSİZ: Metin oluşturmuş fakat seviyeye uygun gerekli biçimsel ve mantıksal bağlantılar kurmakta yetersiz.
Dil bilgisi yapılarını kullanma	7	ÇOK İYİ: Seviyeye uygun bağlaçlar (karşıtlık ve pekiştirme) etkili şekilde kullanılmış, birleşik zamanlı fiil çekimleri kullanılmış, birleşik yapı cümleler hatasız kullanılmış, dolaylı anlatım ve edilgen yapı cümleler etkili kullanılmış.
	5	İYİ: Birleşik yapı cümleler az hatalı şekilde kullanılmış, dolaylı anlatım ve edilgen yapı cümleler kullanılmış, bağlaçlar sınırlı sayıda kullanılmış, fiil çekimleri seviyeye uygun, birleşik zamanlı filler kullanılmış.
	3	ORTA: Birleşik yapı cümle çekimlerinde yoğun hatalar yapılmış bağlaç kullanımı var fakat seviyeye uygun değil, metin yapısı istenilen şekilde oluşturulmamış, anlamda karışıklık var, fiil çekimlerinde bazı hatalar var, birleşik zamanlı yapılar kullanılmamış.
	1	YETERSİZ: Dil bilgisi kullanımı seviyeye uygun değil, yoğun hatalar var, seviyeye uygun zaman kullanımı yapılmamış, metin oluşturulmuş fakat yeterli dil bilgisi yapıları kullanılmamış, değerlendirmek için yetersiz.
Söz varlığı ile Yazım ve noktalama	4	ÇOK İYİ: Noktalama işaretleri işlevine uygun kullanılmış, İmla ve yazım kuralları iyi derecede kullanılmış:- kelime hazinesi yeterli, tekrarlara düşülmemiş -bağlama uygun mecazlı ifadeler kullanılmış, kelimeler etkili kullanılmış, beklenen sayıda ve üstünde sözcük kullanılmış.
	3	İYİ: Yazım ve noktalamada ufak hatalar var, paragraf başları belirgin-büyük, küçük harf yazımlarında hata yapılmamış, Seviye uygun kalıplar kullanılmış, -kelime bilgisi yeterli, -az hata ile iyi kelime kullanımına sahip, beklenen sayının yarısı kadar sözcük kullanılmış.
	2	ORTA: Büyük, küçük harf yazımına dikkat edilmemiş-paragraf başları belirli değil, noktalama işaretleri kullanılmış, Seviye uygun temel kelimeler kullanılmış-farklı kelime kullanımları yapılmamış (mecazlar ve deyimler)- kelime tekrarlarına düşülmüş-kelime bilgisi yeterli değil, beklenen sayının yarısı kadar sözcük kullanılmış.
	1	YETERSİZ: Yazım ve imla kuralları kullanılmamış- büyük/küçük harf, paragraf başı kullanımında yoğun hatalar yapılmış- Sınırlı kelime ile metin yazılmış, değerlendirmek için yeterli değil, sözcük sayısı yetersiz
GEÇERSİZ	1	* Konu ve/veya türden bağımsız: verilen soruyu anlamakta ve/veya tamamlamakta başarısız
NOT	0	-%20
Yönergede verilen limite göre	1	+%20

C1 SEVİYESİ YAZMA BECERİSİ DERECELİ PUANLAMA ANAHTARI

	Puan	Ölçüt
İçerik ve tür	6	ÇOK İYİ: Verilen konu bağlama ve türe uygun eksiksiz geliştirilmiş, fikirler hatasız şekilde (atasözleri, deyimler ile) verilen konu desteklenmiş, Bağlama uygun örtülü ve göndermeli metinler yazılmış, geliştirilmiş, Giriş ve kapanış cümlesi oluşturulmuş.
	4	İYİ: Fikirler içerik ve tür etrafında etkili kullanılmış, çoğu doğru ve düşünceyi destekler nitelikte, giriş ve kapanış cümleleri oluşturulmuş.
	2	ORTA: Verilen konu bağlam ve türle çok az bağlantılı ve fikirler seviyeye uygun şekilde geliştirilmemiş, giriş cümlesi olabilecek bir cümleye sahip.
	1	YETERSİZ: Verilen konu bağlama ve türe göre geliştirilmemiş, hatalı, İçerik ve tür, seviyeye uygun değil, değerlendirmek için yetersiz.
Cümleler arası plan- düzen	3	ÇOK İYİ: Giriş-gelişme-sonuç gibi paragraf planı etkili şekilde uygulanmış, bağlaçlar etkili şekilde kullanılmış-paragraf bütünlüğü ve düzeni etkili şekilde yapılmış, başlık içerik ve türe uygun oluşturulmuş
	2	İYİ: geçişler ve paragraf düzeni etkili şekilde yapılmış, - başlık ve paragraf başları belirtilmiş, başlık oluşturulmuş.
	1	ORTA: Paragraflar arası geçişlerde bazı hatalar var- cümleler arası geçişlerde yoğun hatalar var, başlık ve paragraf başları belirtilmiş, başlık oluşturulmuş.
	0,5	YETERSİZ: Paragraf bütünlüğünde hatalar var-başlık ve paragraf başları belirtilmemiş, seviyeye uygun plan-düzen oluşturulmamış, değerlendirmek için yeterli değil.
Tutarlılık ve bağdaşıklık	5	ÇOK İYİ: Cümleler ve paragraflar arası anlamsal ve biçimsel bağlantı etkili şekilde sağlanmış, biçimsel bağlayıcılar (sözcük, söz öbekleri) etkin şekilde kullanılmış, seviyeye uygun bağlaçlar zenginleştirilerek kullanılmış (açıklama, varsayım, koşul bağlaçları) bağlama uygun kelime seçimi var, çelişkili ifadeler yok.
	3	İYİ: Cümleler ve paragraflar arası biçimsel bağlantıların çok azında hata var, anlamsal bağlantı sağlanmış seviyeye uygun bağlayıcılar kullanılmış, çelişkili ifadeler yok.
	2	ORTA: Anlamsal ve biçimsel bağlantı tam sağlanamamış -bağlayıcı kullanımı var fakat yoğun hatalar yapılmış, bağlama uygun kelime seçimi sağlanamamış, metnin türüne uygun mantıksal öğeler kullanılmış fakat çelişkili ifadeler var.
	1	YETERSİZ: Metin oluşturmuş fakat seviyeye uygun cümleler ve paragraflar arası gerekli biçimsel ve mantıksal bağlantılar kurmakta yetersiz.
Dil bilgisi yapılarını kullanma	7	ÇOK İYİ: Seviyeye uygun bağlaçlar (varsayım, açıklama, koşul) etkili şekilde kullanılmış, birleşik zamanlı fiiller metni zenginleştirilecek şekilde kullanılmış, etken ve edilgen yapı cümleler ile metin zenginleştirilmiş,
	5	İYİ: Birleşik zamanlı cümleler az hatalı şekilde kullanılmış, dolaylı anlatım ve edilgen yapı cümleler kullanılmış, Seviyeye uygun bağlaçlar (varsayım, açıklama, koşul) tekrarlar olmakla beraber kullanılmış.
	3	ORTA: Birleşik zamanlı fiil çekimlerinde yoğun hatalar yapılmış bağlaç kullanımı var fakat seviyeye uygun değil, anlamda karışıklık var, cümlelerin çok azı dil bilgisel açıdan doğrudur
	1	YETERSİZ: Cümlelerin neredeyse tümü dil bilgisel yönden hatalıdır. Değerlendirmek için yetersiz.
Kelime kullanımı İle Yazım ve noktalama	4	ÇOK İYİ: Noktalama işaretleri işlevine uygun kullanılmış, İmla ve yazım kuralları iyi derecede kullanılmış, Mecazlı ifadeler etkili şekilde kullanılmış, ikilemeler ile metin desteklenmiş, konu alanına uygun, mecaz, deyim teri vb. söz varlıkları etkili şekilde kullanılmış, kelime bilgisi uzun ve karmaşık metinler için yeterli, , beklenen sayıda ve üstünde sözcük kullanılmış
	3	İYİ: Yazım ve noktalamada ufak hatalar var, paragraf başları belirgin-büyük, küçük harf yazımlarında hata yapılmamış, Seviyeye uygun kalıplar kullanılmış, az hata ile iyi kelime kullanımına sahip, bazı tekrarlar mevcut, konuya uygun mecaz, deyim, terim vb. söz varlığı ile metin zenginleştirilmiş, beklenen sayıya yakın sözcük kullanılmış
	2	ORTA: Büyük, küçük harf yazımına dikkat edilmemiş-paragraf başları belirli değil, noktalama işaretleri kullanılmış, Kelime kullanımı sınırlı, konu alanına uygun sözcükler kullanılmamış, kelime bilgisi uzun paragraflar için yeterli değil, sınırlı kelime ile paragraf oluşturulmuş, yoğun tekrarlar mevcut, beklenen sayının yarısı kadar sözcük kullanılmış
	1	YETERSİZ: Yazım ve imla kuralları kullanılmamış- büyük/küçük harf, paragraf başı kullanımında yoğun hatalar yapılmış- Kelime bilgisi seviyeye göre yeterli değil, değerlendirmek için yetersiz, sözcük sayısı yetersiz
GEÇERSİZ	1	* Konu ve/veya türden bağımsız: verilen soruyu anlamakta ve/veya tamamlamakta başarısız
NOT Yönergede verilen kelime limitine göre	0	-%20
	1	+%20