Khezerlou, E. (2023). A proposed rubric for assessing student performance in sight translation. *Journal of Sustainable Educational Studies (JSES)*, (Ö2), 108-122.

JSES

Journal of Sustainable Educational Studies

e-ISSN: 2757-5284

Geliş/Received: 06.03.2023 Kabul/Accepted: 29.03.2023

Makale Türü (Article Type): Araştırma Makalesi/Research Article

A Proposed Rubric for Assessing Student Performance in Sight Translation[1]

Ebrahim KHEZERLOU[2]

Abstract

Rubrics are scoring tools used to assess student performance based upon a specific set of criteria. They enable instructors to provide a more accurate, unbiased and consistent scoring, and they can, on the other hand, give students a clear sense of the expectations in a given assignment or task. It has been observed that the assessment of students' abilities in the sight translation course offered in the English Language Translation and Interpreting departments of Turkish universities is not enough efficient and consistent. So, the goal of the present study is to establish a simple, reliable and effective rubric for assessing the undergraduate students' nonverbal and verbal communicative skills when they read a written text in English and then translate it orally into their mother tongue (Turkish). While using the validity model of Lawshe (1975), the content validity of the proposed rubric in a triple period was tested by 5 experts in the first trial, 8 ones in the second trial and 10 ones in the third trial. The inter-rater reliability of the tool was also substantiated in the assessment of the recorded performance of eighteen students by three ratters. The results disclosed the highest content validity ratio for the nonverbal criterion of Eye Contact and the lowest for the verbal criterion of Enthusiasm.

Keywords: Sight translation; rubric; assessment; performance

Andaş Çeviride Öğrenci Edim Değerlendirmesi İçin Önerilen Bir Rubrik

Özet

Rubrikler bir tür değerlendirme aracıdır. Bu araçlar belirli bir dizi kritere dayalı olarak öğrencinin edim değerlendirmesi için kullanılır. Onlar öğretmenlerin daha doğru, tarafsız ve tutarlı bir puanlama yapmalarını sağlar. Öte yandan, öğrencilere belirli bir ödev veya görevdeki beklentiler hakkında net bir fikir verebilirler. Türkiye üniversitelerinde İngilizce Mütercim Tercümanlık bölümlerinde sunulan andaş çeviri dersinin öğrenci yeteneklerinin değerlendirilmesi yeterince verimli ve tutarlı olmadığı kanıtlanmıştır. Bu nedenle şu çalışmanın amacı, lisans öğrencilerinin İngilizce yazılı bir metni okuduktan sonra sözlü olarak onu ana dillerine (Türkçe) çevirdiklerinde iletişim becerilerini ölçmek için basit, güvenilir ve etkili bir rubrik oluşturmaktır. Lawshe'nin (1975) modeli kullanılırken bu rubriğin kapsam geçerliği ilk denemede 5, ikinci denemede 8 ve üçüncü denemede 10 uzman tarafından test edilmiştir. Ayrıca, bu aracın değerlendiriciler arası güvenilirliği, üç değerlendirici tarafından on sekiz öğrencinin kayıtlı performansının değerlendirilmesinde doğrulanmıştır. Araştırmanın nicel bulguları sözel olmayan Göz Teması kriteri için en yüksek kapsam geçerlilik oranını ve sözel Tutku kriteri için en düşük kapsam geçerlilik oranını göstermiştir.

[1] This study is an extended version of the oral presentation presented at the FSMVU Eğitimde Mükemmeliyet Araştırmaları Kongresi (EMAK-2022) on December 16, 2022.

[2] Assist. Prof., Cappadocia University, Faculty of Humanities, Department of English Translation and Interpreting, Nevşehir-Turkey, e.khezerlou@gmail.com, ORCID: 0000-0002-6723-3760

Anahtar Sözcükler: Andaş çevirisi; rubrik; değerlendirme; edim

1. INTRODUCTION

Sight translation/interpreting (ST) simply involves reading a text silently in the source language, and then translating it orally in the target language. Mikkelson (1994, p. 381) denotes it as the "oral translation of a written text". Similarly, Pöckhacker (2016, p. 20) defines it as the oral "rendition of a written text 'at sight'" from one language into another. Likewise, Chen (2015, p. 144) describes it as a type of "rendition of a written message in one language into oral form in another language". Finally, Agrifoglio (2004) characterizes ST as "a hybrid of interpreting (oral output) and written translation (written input)" (Cited in Krapivkina, 2018, p. 696). As implied from these definitions, a sight translator needs to read ahead a written text in the source language before rendering it orally in another one, that is, they have to develop the ability of reading skills in one language and the ability of speech production skills in another language to successfully accomplish the task.

This has truly been claimed by Chen (2015, p. 149) that a successful sight translator's "performance depends on two factors: comprehension of the source text and the ability to produce clear" or fluent delivery of the same message segments through the target text. These skills are central to ST, but they entail various abilities, "including public speaking, reading ahead, analysis of the source text, parsing and chunking of information, sentence completion, paraphrasing skills, the ability to expand and condense, register manipulation, producing the target language version quickly, domain knowledge, and understanding cultural nuances" (Chen, 2015, p. 149). Čeňková (2010) also highlights the significance of chunking skills in her skill set for sight interpretation, namely "fast retrieval of verbs, key words, and numbers in the source text, numbering various components when restructuring a long sentence, filtering out secondary information, speedy transfer of meaning, parsing long and complex sentences, fluent delivery, and avoiding repetition and unnecessary shift" (Cited in Chen, 2015, p. 150). Moreover, Agrifoglio (2004) describes features of ST as: Continuous access to information in the text, attention sharing between visual input and oral production, coordination of reading and production effort, monitoring production while reading, progressive access to new information (no previous access) or prior access to information (previous reading) and extreme risk of source text interference. In this regard, Lee (2012) also lists a number of capabilities required from fruitful sight interpreters, such as reading comprehension, the ability to differentiate between main and subsidiary ideas, avoiding the source language interferences, speed reading, quick response, avoiding redundant translation, developing chunking strategies and condensing source information. Lastly, Ersozlu (2005) identifies six major skills that are essential for ensuring an accurate, coherent and fluent sight interpretation at undergraduate level, specifically fast reading and comprehension, domain knowledge, detailed reading, dealing with unknown words, chunking skills and meaning retention.

To enable students to act successfully in a real world setting of ST, "many translation and interpretation programs around the world include ST as part of their curricula, with the emphasis ranging from a few classes at the beginning of a translation and/or interpretation course to a full course lasting for the entire semester" (Chen, 2015, p. 146). In line with this global pedagogical goal, the translation and interpreting programs in Turkey offers a semester-lasting ST course coded and titled as 'IMT305 Sight Translation' (in Turkish, IMT305 Yazılı Metinden Sözlü Çeviri) where students have to read various types of informative texts in English, and then render the contents orally in their mother tongue (i.e., the Turkish language). The course specifically aims at familiarizing the students with the fundamental ST skills, such as reading comprehension, scanning for the main and secondary ideas, fast reading, accurate and fluent delivery of the content, proper paraphrasing and summarising, filtering out unnecessary information, domain knowledge, memory expansion, organising an oral speech, using chunking skills and strategies, subject interest, vocabulary enrichment and practice of reading aloud techniques (e.g., breathing, pauses and enunciation). At the same time, it targets to provide the students with the background knowledge on consecutive and simultaneous interpretation courses as well.

As mentioned above, the reception and production processes of ST require a variety of competences (e.g., linguistic, sociolinguistic and pragmatic). However, the problem is that these vital skills are hardly identified and evaluated properly because of the multi-dimensionality and complexity of the phenomenon, which "may explain why there have been few attempts to validly and reliably measure translation competence/ability" (Angelelli, 2009, p. 13). Accordingly, the assessment of an interpreter's competence and performance usually seems to be conducted in a subjective manner. Indeed, when I asked my colleagues how they assess their students' performance in the ST course, they could not describe it accurately. In addition, little has been published on the

subject in Turkey. As Angelelli and Jacobson (2009, p. 4) reasonably claimed "there is a lack of empirical research on both translator and interpreter competence and performance, and on assessing processes and products". Therefore, the general purpose of the study is to define a comprehensive, transparent and coherent framework for assessing the students' ST competence and performance.

One of the valid pedagogical tools that may perfectly help teachers systematize the measuring of different levels of ST competency is a rubric. Andrade (2005, p. 27) defines it as "an assessment tool that lists the criteria for a piece of work or what counts" and "articulates gradations of quality for each criterion, from excellent to poor". According to Suskie (2009), rubrics enable teachers to (a) get away from vague, fuzzy pedagogical activities, (b) to set up a simple way for grading tasks or assignments through an explicit set of criteria, (c) provide a more accurate, unbiased, consistent and objective scoring, (d) improve feedback to the faculty and students and (e) decrease grading time and student complaint. They can also give students a clear sense of what the expectations are for a given task and how they can be met (Brookhart, 2018). So, the study aims to create and validate a simple, reliable and effective rubric for assessing students' non-verbal and verbal communication skills when they read and analyse a written foreign language text (i.e., English), and then translate them orally into their own native language (i.e., Turkish). We are aware that "distinguishing between verbal and nonverbal communication is not as conceptually straightforward as it might at first seem" (Hargie, 2011, p. 43). "Neither is it useful to think of the two as being operationally discrete" since they are operating side by side- as part of the same system (Hargie, 2011, p. 43-44). Yet, these categories of communication differ from each other.

1.1. Nonverbal Communication

Nonverbal communication, often done subconsciously, is the process of sending and receiving messages without using spoken or written words (DeVito, 2005; Hargie, 2011; Hogg & Vaughan, 2018). Vaughan and Hogg (1998) define it as the transfer of meaningful information from one person to another by means of a linguistic medium other than written or spoken language. Hargie (2011, p. 43) also argues that nonverbal communication concerns "with those forms and functions of face-to-face interaction that do not rely primarily upon the content of what we say" rather, "the focus is upon how we communicate through, for example, a glance, gesture, postural shift or facial expression".

Nonverbal signals such as facial expressions, gestures, posture, eye contact, touch and the use of space communicate a person's intentions and emotions that words won't always speak. They can also be used to reinforce or replace verbal communication. Moreover, they can be used to convey messages that would be considered inappropriate if conveyed verbally. In this regard, Dickson and Hargie (2003, p. 50) suggest that nonverbal communication is used to (a) replace verbal communication in situations where it may be impossible or inappropriate to talk, (b) supplement the verbal communication, thereby enhancing the overall message, (c) modify the spoken word, (d) contradict, either intentionally or unintentionally, what is said, (e) regulate conversation by helping to mark speech turns, (f) express emotions and interpersonal attitudes, (g) negotiate relationships in respect of, for instance, dominance, control and liking, (h) convey personal and social identity through such features as dress and adornments and (i) contextualise interaction by creating a particular social setting.

There are many different types of nonverbal communication criteria that can be taken into account in ST. Accordingly, the study concerns with those face-to-face interaction ones that may be used in ST classes between teachers and students. Four non-verbal criteria were chosen for this purpose. Basic definitions of each criterion are given below. It should be noted that these definitions deserve further considerations, which was beyond the scope of the paper.

- *Eye contact*: It is a form of nonverbal communication that humans use to communicate many forms of emotions that words won't always speak. It is a fundamental factor in building interpersonal relationships, an important indicator of attentiveness and interest in what is being said and a crucial signifier of control or power over a situation. While delivering a speech, a speaker should make eye contact with the whole members of a speech event and avoid fixing eyes on a particular individual or looking at a particular section of the audience (Hargie, 2011; Hogg & Vaughan, 2018).
- *Gestures*: A gesture is a form of non-verbal communication in which visible bodily actions communicate particular messages, either in place of, or in conjunction with, speech. Gestures include movement of the hands, face, or other parts of the body. They differ from physical non-verbal communication that does not

communicate specific messages, such as purely expressive displays, proxemics, or displays of joint attention (Kendon, 2004). When presenting a speech, the gestures must seem fluid and help her/him to visualize the content. The most common type of gestures used in the presentation of a speech are hand movements, which are often used to express our emotions, tell a story, or comfort ourselves. The hands must not be hidden from others during a presentation. Hands in pockets, hands behind back, or closed fists can all act as barriers against the ability of an individual to communicate effectively (Hargie, 2011; Hogg & Vaughan, 2018).

- *Proxemics*: It is a type of nonverbal communication involving how we deal with the environment around us. People perceive and use physical space around them to achieve their communication goals (Hogg & Vaughan, 2018; Vaughan & Hogg, 2002).
- *Paralanguage*: Paralanguage, another form of nonverbal communication, refers to the vocal elements of communication that are not covered by the literal meaning of what is being said. In one word paralanguage is "not what you say but how you say it" (Baxodirjonova, 2020, p. 322). It includes our tone, pitch, volume, rhythm and other vocal elements. The ability to interpret this kind of human communication correctly is considered an important competency both in personal and professional settings. We can say how confident or anxious one feels simply by listening to their voice (Hogg & Vaughan, 2018; Vaughan & Hogg, 2002).

1.2. Verbal Communication

Verbal communication is the use of spoken or written words to share information with other people (Ali, 2018). It encompasses both how one receives and delivers messages. There are a large number of verbal communication skills, which range from the obvious one of speaking or listening to someone to the more subtle ones of reflecting and clarifying. However, the study focuses on those verbal communicative skills that are necessary for the ST: Recreating the original messages of the written texts while using a combination of the written materials, memory and general knowledge (Albl-Mikasa, 2008; Gillies 2017). Taking into account this definition and the strategies of an effective communication, the following six verbal criteria were selected for the study. A very brief definition of each was operationally provided as:

- *Organization*: The text structure refers to how the information within a text is organized (Armbruster, 2004, Ghorbani Shemshadsara, Ahour & Hadidi Tamjid, 2019). Grabe (2002, p. 10) states that "discourse, or text, structures can be understood as knowledge structures or basic rhetorical patterns in texts." The organization of the presentation in ST concerns with the ability of a student in providing an informative, logical, interesting, complete and effective presentation with clear, focussed and organised introduction, body and conclusion, where the topic is developed logically and supported interestingly and informatively. Skilled students with the knowledge of text structure are capable of constructing mental models of the main ideas, as well as learning and remembering the information presented in the text (Meyer & Rice, 1982).
- *Memory span*: The memory refers to the ability of both preserving and recovering information. The memory span denotes the skilfulness of a student in processing, retaining and retrieving of a large amount of the content information in her/his presentation. The memory is differentiated as the *working memory* which stores the processed information temporarily, *short-term memory* which keeps the information for a few seconds and the *long-term memory* which saves the information for a very long time (Baddeley, 2012). "Since memory is accepted as a component of intelligence or cognitive operations, it can be an important component in consecutive and simultaneous interpretations" (Vural, 2021, p. 138).
- *Background knowledge*: Background knowledge or prior knowledge refers to the information or experiences that influence the learning and memory abilities of the individuals. Smith, Snow, Serry, and Hammond (2021, p. 216) characterises background knowledge as "all of the world knowledge that the reader brings to the task of reading." When a student reads, the way they try to make sense of what the text says is by reference to their pre-existing background knowledge on that topic (prior knowledge). So, the more a student knows about a topic, the better their reading is. Smith, Snow, Serry, and Hammond (2021, p. 226) suggest that "readers who have a strong knowledge of a particular topic, both in terms of quantity and quality of knowledge, are more able to comprehend a text than a similarly cohesive text for which they lack background knowledge." Here, background knowledge refers to how a student accurately bridges between her/his background knowledge and the content information of the text materials in order to fully present the topic.

- *Content knowledge*: The content knowledge refers to the body of knowledge and information that teachers teach and that students are expected to learn in a given subject, such as English language arts, mathematics, science, or social studies. It generally refers to the facts, concepts, theories, and principles that are taught and learned in specific academic courses (Cabell & Hwang, 2020). The criterion in the study refers to the ability of a student in understanding the contents of the written materials and delivering their messages. That is, how well a student addresses the content information during the presentation by referring to the main and supporting ideas.
- *Expressiveness*: In functional linguistics expressiveness fulfils "the communication purposes and aims at attaining the highest degree of communication effectiveness" (Apresyan, 2018, p. 8). Expressive language simply refers to the way a speaker expresses him/herself. ST requires the application of a battery of speaking strategies to put forward an active or passionate speech through overcoming the translation problems. This construct is measured by how well the student use these commutative strategies, such as emphasizing key words, rephrasing, interpreting the meaning of words contextually, appealing for help, etc. during their presentation.
- *Enthusiasm*: Enthusiasm refers to the ability to show a keen interest in a subject or an activity, as well as a readiness to get involved (Harackiewicz, Smith & Priniski, 2016). Enthusiastic students have a strong feeling of eagerness to do the given task. Until they see that the task fulfilled, they never want to give up and are not easily demotivated by the temporal crisis. Enthusiastic students work with passion and possess an intrinsic motivation that spurs them on. Whether the students demonstrate strong positive feelings to the subject of the text material in their presentations and fulfil their task successfully.

However, the study tries to answer to the following research questions:

RQ1. How valid is the content of the proposed rubrics?

RQ2. How reliable is the proposed rubrics?

2. METHOD

2.1. Research Design

The research design employed in the study was a quantitative method. Quantitative research is a type of research that collects and analyses numerical data to test hypotheses and answer research questions (Cohen, 1988). This research typically involves a large sample size and uses statistical analysis to make inferences about a population based on the data collected. It often involves the use of surveys, experiments, or other structured data collection methods to gather quantitative data. Quantitative Research Methods include 'descriptive', 'correlational' 'quasi-experimental' and 'survey' research designs. This study was a descriptive research in a sense that descriptive statistics were used to compute the content validity ratios of the criteria. It was also a quasi-experimental research in a sense that inferential statistics (i.e., the Fleiss' Kappa Test) were used to investigate the cause-and-effect relationships between the variables (Cohen, 1988).

2.2. Participants

The participants were 18 third-year English Translation and Interpreting students at a Turkish private university in the fall semester of 2022-2023 academic year. None of the students had received the course before.

2.3. Data Collection

The data for the present study comes from two midterms and one final exams: the first midterm exam was held after 6th session, the second one after 12th session and the final exam after the 14th session. The performances of the 18 students were recorded in each exam, but 75% of the participants (n=14) were randomly selected for the study. The materials used in the exams were a total of 25 informative texts, each including a word length of 1000-1500 words. The students had practiced them in their classes before, however, they had to select randomly one text in the exam and had 5-10 minutes right to refresh their minds before presenting the selected written material orally in their mother tongue.

2.4. Analysis Procedures

In an attempt to produce an acceptable rubric for assessing the students' ST communicative competence at the undergraduate level, the study determined to test the validity and reliability of the proposed rubric at three trials. These are two fundamental elements in the evaluation of any measuring tool. Validity is the extent to which an

instrument measures what it is intended to measure. There are different kinds of validity (i.e., construct, criterion, content and face). While all of these validity types are required in a test, content validity is perhaps the most relevant one for an excellent assessment of a tool. So, we focus on this type of validity in the study. The term refers to how well a survey or test measures the content knowledge that it sets out to measure. The content validity of the proposed rubric was tested by using the validity model of Lawshe (1975). The model measures content validity ratio (CVR) of each item based on the established critical values of a panel of judges and content validity index (CVI) of the overall items through calculating the average score. When an item passes the threshold value, it is considered as acceptable. Moreover, as the CVI of the entire test gets closer to 1, it is considered as valid. The rubric was validated by 5 experts in the first trial, 8 ones in the second trial and 10 ones in the third trial.

Reliability is the overall consistency of an instrument (Tavakol & Dennick, 2011). A measure is said to be reliable if it produces similar results under consistent conditions. Highly reliable scores are characterized to be precise, reproducible and consistent from one testing occasion to another. Various kinds of reliability coefficients have been used to indicate the amount of effectiveness of a measure: The inter-rater reliability assesses the degree of agreement between two or more ratters in their judgements; the test-retest reliability assesses the degree to which test scores are consistent from one test administration to the next; the inter-method reliability assesses the degree to which test scores are consistent when there is a variation in the methods or instruments used; finally, the internal consistency reliability assesses the consistency of results across items within a test (Haradhan, 2017; Taherdoost, 2016). However, for the purpose of the study we used the inter-rater reliability, which is considered the most effective method of reliability.

As mentioned in the previous paragraph, the inter-rater reliability test shows different ratings with almost similar values despite the independent scoring process undertaken. The inter-rater reliability of the proposed rubric was measured using the Fleiss' κ coefficient, which is a statistical metric for measuring the reliability of agreement between a fixed number of ratters in assigning the categorical ratings to items. Landis and Kock (1977) proposed the following Kappa benchmark values for interpretation of a test item:

Kappa Statistic	Interpretation
<0.00	Poor Agreement
0.00-0.20	Slight Agreement
0.21-0.40	Fair Agreement
0.41-0.60	Moderate Agreement
0.61-0.80	Substantial Agreement
0.81-1.00	Almost Perfect Agreement

Figure 1. Kappa benchmark values

Landis and Koch (1977, p. 165)

The more the Kappa value of an item gets closer to 1, the more it is considered as reliable and acceptable. Kappa values between (0.61 - 0.80) and (0.81 to 1.00) indicate a 'substantial' and 'almost perfect' agreement levels, respectively, and they are considered as acceptable indexes in judging about the reliability of an item. The inter-rater reliability of the proposed rubric based on this method is obtained from a total of three ratters, who rated the non-verbal and verbal skills of the students in a 1-4 Likert-type scale (High level/excellent=4, moderate level/good=2, low level/acceptable=3 and does not meet the criterion/poor=1) in three trials of a midterm exam, a quiz and a final exam.

2.5. Scientific Research and Publication Ethics

In the study, all the rules specified to be followed by the *"Higher Education Institutions Scientific Research and Publication Ethics Directive*" were complied with. None of the actions specified in the 2nd part of the Directive, titled "*Actions Contrary to Scientific Research and Publication Ethics*", were carried out.

2.5.1. Ethics Committee Approval

There is no ethics committee approval of this research.

3. RESULTS

3.1. Validity of the Proposed Rubric

In the first trial, we had a panel of five judges whose critical value index based on the Lawshe's (1975) model was 0.99. As shown in table 1, the highest CVR value was observed for Eye Contact (CVR= 1) and the lowest CVR value was for Enthusiasm (CVR= .42). The results revealed that most of the items failed to pass the threshold value of .99 in the first trial (See table 1). Accordingly, the CVI value of the entire test was quite low (CVI= .646). This indicates that the test items do not likely measure well the construct that they set out to measure. Therefore, we removed and modified the items that had low CVR values to improve the overall content validity of the test. Additionally, the number of panellists were increased to 8 to lower the critical value of five judges.

In the second trial, we had a panel of eight judges whose critical value index based on the Lawshe's (1975) model was 0.75. The highest CVR score was recorded again for Eye Contact (CVR= .96) and the lowest coefficient for Expressiveness (CVR= .52). As shown in table 1, the content validity ratio of five of the items fell below the critical value of .75 in the second trial: Paralanguage (CVR= .61), Memory span (CVR= .67), Background knowledge (CVR= .72), Expressiveness (CVR= .52) and Enthusiasm (CVR= .58). This means that these items still failed to measure the construct of interest as well as they should. The CVI value of the entire test was .736. Accordingly, the items were modified again to improve their CVR values and develop the overall content validity of the test. The number of panellists were also increased to 10 to lower the critical value of eight judges.

In the third trial, we had a panel of ten judges whose critical value index based on the Lawshe's (1975) model was 0.62. The highest score was observed for Eye Contact (CVR= .93) and the lowest scores for Enthusiasm (CVR= .64). As shown in table 1, all of the items passed the threshold value of .62 in the third trial. The average coefficient of the test was .782, which got closer to 1, that is, the content validity increased. Thus, the content validity of the proposed rubric was confirmed through its CVR and CVI values by using the Lawshe's (1975) model (See table 1).

Table 1. Content Validity Ratios and Indexes of the Items in the First, Second and Third Trails

Skill	Criteria	CVR of 5 Experts in Trail 1	CVR of 8 Experts in Trail 2	CVR of 10 Experts in Trail 3
Non-Verbal	Eye contact	1	.96	.93
	Gestures	.74	.79	.82
	Proxemics*	.56	.83	.78
	Paralanguage*	-	.61	.74
Verbal	Organization	.68	.82	.88
	Memory span	.53	.67	.73
	Background knowledge	.64	.72	.78
	Content knowledge	.78	.86	.85
	Expressiveness	.47	.52	.67
	Enthusiasm	.42	.58	.64
	CVI	5.82/9=.646	7.36/10=.736	7.82/10=.782

* Indicating that the construct was added, removed, renamed or replaced in the first trial.
-Lawshe's (1975) Critical Values: For a panel of 5 judges= .99, For a panel of 5 judges= .75, For a panel of 10 judges= .62.

3.2. Reliability of the Proposed Rubric

In the first trial, the Fleiss' Kappa Test results for inter-rater agreement revealed two different categories for the non-verbal items and four categories for the verbal ones, according to Landis and Koch's (1977) Kappa benchmarks. The non-verbal skills of Eye Contact (z=3.715, k=.367, p=0.000) had 'fair agreement' and both Gestures (z= 1.867, k=.183, p=.062) and Poise (z=.606, k= .063, p=.545) had 'slight agreement' (See table 2).

Whereas, the verbal skills of Content Knowledge (z=4.344, k=.481, p=.000) had 'moderate agreement', Organization (z=2.120, k=.249, p=.034) had 'fair agreement', Memory Span (z=.330, k=.037, p=.741), Background Knowledge (z=1.364, k=.129, p=.172), Expressiveness (z=1.568, k=.153, p=.117) and Enthusiasm (z=1.309, k=.143, p=.190) had 'slight agreement' and Elocution (z= -.876, k=-.084, p=.381) had 'poor agreement' (See table 2).

Table 2. Inter-rater Reliability of the Non-verbal and Verbal skills in the First Trail

		Fleiss' Kappa Test				
Skills	Criteria	Overall Kappa	Standard Error	Z	P-Value	Agreement
Non-Verbal	Eye contact	.367	.099	3.715	.000	Fair
	Gestures	.183	.098	1.867	.062	Slight
	Poise*	.063	.105	.606	.545	Slight
Verbal	Organization	.249	.117	2.120	.034	Fair
	Memory span	.037	.111	.330	.741	Slight
	Background knowledge	.129	.095	1.364	.172	Slight
	Content knowledge	.481	.111	4.344	.000	Moderate
	Elocution*	-.084	.096	-.876	.381	Poor
	Expressiveness	.153	.098	1.568	.117	Slight
	Enthusiasm	.143	.109	1.309	.190	Slight

* Indicating that the construct was added, removed, renamed or replaced in the first trial.
-Landis and Koch's (1977) Kappa benchmarks: < 0= Poor, 0.01- 0.20= Slight, 0.21-0.40= Fair, 0.41- 0.60= Moderate, 0.61- 0.80= Substantial and 0.81-1.00= Almost perfect.

This result indicated that most of the test items failed to measure consistently. Therefore, we had to run the second trial after modifying the items that had low CVR values.

In the second trial, the Fleiss' Kappa Test results for inter-rater agreement revealed four different categories for both the non-verbal and verbal items. The non-verbal skills of Eye Contact (z=6.849, k=.853, p=0.000) had 'almost perfect agreement', Gestures (z=5.472, k=.759, p=.000) had 'substantial agreement', Proxemics (z= 4.852, k= .479, p=.005) had 'moderate agreement' and Paralanguage (z= 1.920, k=.157, p=.074) had 'slight agreement' (See table 3).

While, the verbal skills of Content Knowledge (z=5.386, k=.627, p=.000) had 'substantial agreement', Organization (z=2.954, k=.439, p=.034) had 'moderate agreement', Background Knowledge (z=2.428, k=.318, p=.042) had 'fair agreement' and Memory Span (z=1.927, k=.187, p=.076), Expressiveness (z=1.568, k=.147, p=.068) and Enthusiasm (z=1.704, k=.171, p=.094) had 'slight agreement' (See table 3).

Table 3. Inter-rater Reliability of the Non-verbal and Verbal skills in the Second Trail

		Fleiss' Kappa Test				
Skills	Criteria	Overall Kappa	Standard Error	Z	P-Value	Agreement
Non-Verbal	Eye contact	.853	.167	6.849	.000	Almost perfect
	Gestures	.759	.204	5.472	.000	Substantial
	Proxemics	.479	.142	4.852	.005	Moderate
	Paralanguage	.157	.118	1.920	.074	Slight
Verbal	Organization	.439	.172	2.954	.034	Moderate
	Memory span	.187	.184	1.927	.076	Slight
	Background knowledge	.318	.167	2.428	.042	Fair
	Content knowledge	.627	.147	5.386	.000	Substantial
	Expressiveness	.147	.099	1.568	.068	Slight

Enthusiasm	.171	.104	1.704	.094	Slight

-Landis and Koch's (1977) Kappa benchmarks: < 0= Poor, 0.01- 0.20= Slight, 0.21-0.40= Fair, 0.41- 0.60= Moderate, 0.61- 0.80= Substantial and 0.81-1.00= Almost perfect.

The above results showed that the reliability of the test items have been improved, but still some items were unsuccessful to measure consistently. Therefore, we had to run the third trial after modifying the items that had low CVR values.

The Fleiss' Kappa coefficient results in the third trial revealed two different ratter-agreement categories for the non-verbal and verbal items. The non-verbal skills of Eye Contact (z=7.253, k=.847, p=0.000), Gestures (z= 6.438, k=.831, p=.000) and Proxemics (z=5.284, k=.824, p=.000) had 'almost perfect agreement', while Paralanguage (z= 4.567, k=.682, p=.01) had 'substantial agreement' (See table 4).

Where, the verbal skills of Organization (z=7.713, k=.864, p=.000), Background Knowledge (z=6.248, k=.872, p=.000) and Content Knowledge (z=6.146, k=.847, p=.000) had 'almost perfect agreement', while Memory Span (z=3.452, k=.743, p=.02), Expressiveness (z=3.855, k=.683, p=.014) and Enthusiasm (z=2.797, k=.649, p=.024) had 'substantial agreement' (See table 4).

Table 4. Inter-rater Reliability of the Non-verbal and Verbal skills in the third Trail

		Fleiss' Kappa Test				
Skills	Criteria	Overall Kappa	Standard Error	Z	P-Value	Agreement
Non-Verbal	Eye contact	.847	.254	7.253	.000	Almost perfect
	Gestures	.831	.257	6.438	.000	Almost perfect
	Proxemics	.824	.271	5.284	.000	Almost perfect
	Paralanguage	.682	.234	4.567	.001	Substantial
Verbal	Organization	.864	.252	7.713	.000	Almost perfect
	Memory span	.743	.2847	3.452	.002	Substantial
	Background knowledge	.872	.291	6.248	.000	Almost perfect
	Content knowledge	.847	.237	6.146	.000	Almost perfect
	Expressiveness	.683	.243	3.855	.014	Substantial
	Enthusiasm	.649	.189	2.797	.024	Substantial

-Landis and Koch's (1977) Kappa benchmarks: < 0= Poor, 0.01- 0.20= Slight, 0.21-0.40= Fair, 0.41- 0.60= Moderate, 0.61- 0.80= Substantial and 0.81-1.00= Almost perfect.

The Fleiss' Kappa coefficient values of the proposed rubric ranged between (0.61 - 0.80) for a 'substantial agreement' and between (0.81 to 1.00) for a 'almost perfect agreement'. This indicates there exists a high inter-rater reliability across ratters in the third trial. Thus, it can be argued that the proposed rubric (see appendix one) now measures consistently than before.

4. DISCUSSION

The study examined the assessment of ST performance of Turkish students through a proposed rubric, whose validity and reliability were statistically proved by the raters' feedbacks on the scale. The results revealed that the critical value index of each 'nonverbal' and 'verbal' item was beyond the threshold value of 0.62 in the third trial (See table 1). This implies that the proposed rubric is statically a valid tool. The Fleiss' Kappa coefficient values of the proposed rubric indicated that it is a reliable tool because there existed 'substantial' and 'almost perfect' agreements among the items in the third trial (See table 1). The results confirm that the students' 'nonverbal' and 'verbal' abilities have been improved from the mid-term to final examinations. This can certainly be attributed to the wording improvement of the scale in the following trials which increased the scale's measuring accuracy. Moreover, it can also be argued that the students might be uncertain about what it was expected from them in the first and second mid-term examinations, however, they appeared to have become more aware of completing their tasks in the final examination. This finding supports the assumption of Yamada (2020, p. 348) that "ST may be a trainable activity and that the more ST training students receive, the better the skills they will acquire".

The results also disclosed the highest and the lowest content validity ratio for the nonverbal criteria of Eye Contact (CVR= .93) and Paralanguage (CVR= .74), respectively (See table 1). In the case of eye contact, this implies that the visual contact with the source written text and the audience at the same time did not disturb the students' attention. This finding indirectly supports the view of Agrifoglio (2004, p. 47) who noted that "the main difficulty of ST lies not in the written nature of the source text, but in the smooth coordination of the Reading, Memory and Production Efforts, while struggling against increased visual interference from the source language." However, it is not in line with the findings of Mikkelson, Willis and Alvarez (1995) who argued that the "constant shifting of the translator's visual contact with the source text and the presence of the written text makes it much more difficult for the interpreter to perform ST" (Cited in Yamada, 2020, p. 344).

As regards to paralanguage, the students failed to put forward a good performance. According to Mehrabian's (1980, p. 196) communication model, "only 7% of communication takes place through the words. 38% is made up by tone and voice and 55% of communication takes place through the body language we use" (Cited in Baxodirjonova, 2020, p. 322). This result does not advocate Mehrabian's claim that "it is extremely easy with paralinguistic features" in a conversation (Baxodirjonova, 2020, p. 323) although we very often use them unintentionally in our daily speech. Regarding the issue, Girsang, Sumbayak and Yusuf (2021, p. 14) argued that "using the right paralanguage features especially pitch will bring harmonization in our speaking and the communication between the speaker and the listener will turn to successful communication." It is implied from the quotation that the tone of the students' voice might be an important predictor of their low performance.

Moreover, the results of verbal skill released the highest and the lowest content validity ratio for the verbal criteria of Organization (CVR= .88) and Enthusiasm (CVR= .64), respectively (See table 1). In the matter of organization, it means that the students were successful in arranging their speech's information in a logical way from the mid-term to final examinations. The organization of the text is the arrangement of ideas and the relationships among them (Armbruster, 2004). When planning a speech, it is important to arrange the information logically because it impacts the way the audience understands the speech. An organized speech gives credibility to the speaker as well. The speaker will appear more professional or competent in their role as a speaker. The students' efficient performance, however, is probably assigned to the fact that the informative text materials were already well-organized by the researcher. The students often followed the organizational patterns of the texts in their presentations. The well-organized texts are easier to understand and remember and are usually more enjoyable as well. Carrell (1989) claims that the amount of information that students recall depends largely on the kind of structure of a text. Baxodirjonova (2020, p. 323) also argues that paralanguage means are "as organizers of communication", and "colorful and well-organized speech is usually full of paralinguistic features". This assumption, however, is not verified by the findings of this study because the students' performance was very low in this regard.

Regarding to enthusiasm, the students' performance demonstrated that the subject of their text materials was not interesting. Interest is distinguished as 'situational' and 'individual' (Hidi, & Renninger, 2006). The former, characterized as a psychological state, concerns with "an individual's momentary experience of being captivated by an object" while the latter denotes "more lasting feelings that the object is enjoyable and worth further exploration" (Harackiewicz, Smith, & Priniski, 2016, p. 2). It is argued that "experiencing situational interest can directly promote learning by increasing attention and engagement (Harackiewicz, Smith, & Priniski, 2016, p. 2). If this type of interest develops into an individual interest, the student will more likely reengage with the material overtime and explore the topic further (Harackiewicz, Durik, Barron, Linnenbrink-Garcia, & Tauer, 2008). Based on this interest theory, it may be argued that low student enthusiasm is originated from their situational interest, that is, stressful situations such as examinations may have affected their ST processes in different ways, and thus resulted in negative feelings towards the subject text materials among the students.

5. CONCLUSION AND RECOMMENDATION

The purpose of the study was to design a valid and reliable measure for assessing the undergraduate students' communicative skills in a ST course titled as 'IMT305 Sight Translation' in the context of Turkey, where they had to read a written text in English, and then translate it orally into their own native language. The proposed rubric including the 'nonverbal' and 'verbal' domains was validated by 5, 8 and 10 experts in three trials of a midterm exam, a quiz and a final exam. Likewise, the inter-rater reliability of the rubric was also confirmed by three ratters, who rated the non-verbal and verbal skills of the students in three trials. Through the statistically proven results it was concluded that the proposed rubric in the last trial was a valid and reliable measure.

Nevertheless, these results should be interpreted with caution since there were very few subjects with only Turkish background in the study. For this reason, there is need to test the tool further with students and ratters of different language backgrounds to be virtually assured of its validity and reliability further.

6. DISCLOSURE STATEMENT

Scientific Research and Publication Ethics: In the study, all the rules specified to be followed by the *"Higher Education Institutions Scientific Research and Publication Ethics Directive*" were complied with. None of the actions specified in the 2nd part of the Directive, titled "*Actions Contrary to Scientific Research and Publication Ethics*", were carried out.

Ethics Committee Approval: There is no ethics committee approval of this research.

Statement of Researchers' Contribution: This article was created by a single author.

Conflict of Interest: The authors declare no potential conflict of interest.

Financial Conflicts: There are no financial conflicts of interest to disclose.

7. REFERENCES

Agrifoglio, M. (2004). Sight translation and interpreting: A comparative analysis of constraints and failures. *Interpreting, 6*(1), 43-67. https://doi.org/10.1075/intp.6.1.05agr

Albl-Mikasa, M. (2008). (Non-)Sense in note-taking for consecutive interpreting. *Interpreting, 10*(2), 197–231. https://doi.org/10.1075/intp.10.2.03alb

Ali, M. (2018). Communication skills 3: Non-verbal communication. *Nursing Times, 114*(2), 41-42. www.nursingtimes.net

Andrade, G. H. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching, 53*(1), 27-31. https://doi.org/10.3200/CTCH.53.1.27-31

Angelelli, V. C. (2009). Using a rubric to assess translation ability: Defining the construct. In C. V. Angelelli and H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies: A call for dialogue between research and practice* (pp. 13-48). Amsterdam: John Benjamins Publishing Company.

Angelelli, V. C., & Jacobson, E. H. (2009). *Testing and assessment in translation and interpreting studies*. John Benjamins Publishing Company.

Apresyan, M. (2018). On the concept of "expressiveness" in modern linguistics. *Annals of Language and Literature, 2*(4), 8-12. https://www.sryahwapublications.com/annals-of-language-and-literature/pdf/v2-i4/2.pdf

Armbruster, B. B. (2004). Considerate texts. In D. Lapp, J. Flood and N. Farnan (Eds.), *Content area reading and learning: Instructional strategies* (2nd ed.), (pp. 47-58). Mahwah, NJ: Erlbaum.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baxodirjonova, X. (2020). The role of paralinguistics in English communications. *Science and Education Scientific Journal, 1*(9), 321-324. www.openscience.uz

Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education, 3,* 1-12. https://doi.org/10.3389/feduc.2018.00022

Cabell, S. Q., & Hwang, H. J. (2020). Building content knowledge to boost comprehension in the primary grades. *Reading Research Quarterly, 55*(1), 99-107. https://doi.org/10.1002/rrq.338

Carrell, P. L. (1989). Metacognitive awareness and second language reading. *The Modern Language Journal, 73*(2), 121–134. https://doi.org/10.1111/modl.1989.73

Chen, W. (2015). Sight translation. In H. Mikkelson and R. Jourdenais (Eds.), *Routledge Handbooks in Applied Linguistics* (pp. 144-153). London and New York: Routledge.

Čeňková, I. (2010). (2010). Sight translation. In Y. Gambier and L. Van Doorslaer (Eds.), *Handbook of translation studies, Vol. 1* (pp. 320-323). John Benjamins.

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

DeVito, J. (2005). *Essentials of human communication* (5th ed.). Boston, MA: Pearson Education.

Dickson, D., & Hargie, O. (2003). *Skilled interpersonal communication: Research, theory and practice*. London: Routledge.

Ersozlu, E. (2005). Training of Interpreters: Some suggestions on sight translation teaching. *Translation Journal, 9* (4). http://translationjournal.net/journal/34sighttrans.htm

Gillies, A. (2017). *Note-taking for consecutive interpreting: A short course* (2nd ed.). London: Routledge.

Girsang, I. M., Sumbayak, M. D., & Yusuf, M. (2021). Paralinguistic features in students' speaking performance. *LingPoet: Journal of Linguistics and Literary Research, 2*(2), 1-16. https://doi.org/10.32734/lingpoet.v2i2.4452

Ghorbani Shemshadsara, Z., Ahour, T., & Hadidi Tamjid, N (2019). Raising text structure awareness: A strategy of improving EFL undergraduate students' reading comprehension ability. *Cogent Education, 6*, 1644704. https://doi.org/10.1080/2331186X.2019.1644704

Grabe, W. (2002). Dilemmas for the development of second language reading activity. In J. C. Richards and W. A. Renandya (Eds.), *Methodology in language teaching an anthology of current practice* (pp. 276-286). Cambridge: Cambridge University Press.

Harackiewicz, J. M., Durik, A. M., Barron, K.E., Linnenbrink-Garcia, L., & Tauer, J. M. (2008). The role of achievement goals in the development of interest: Reciprocal relations between achievement goals, interest, and performance. *Journal of Educational Psychology, 100*, 105-122. https://doi.org/10.1037/0022-0663.100.1.105

Harackiewicz, M. J., Smith, L. J., & Priniski, J. S. (2016). Interest matters: The importance of promoting interest in education. *Policy Insights Behav Brain Sci., 3*(2), 220–227. https://doi.org/10.1177/2372732216655542

Haradhan, M. J. (2017). Two criteria for good measurements in research: Validity and reliability. *Annals of Spiru Haret University*, *17*, 59-82. https://doi.org/10.26458/1746

Hargie, O. (2011). *Skilled interpersonal communication: Research, theory and practice* (5th ed.). London and New York: Routledge.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*, 111-127. https://doi.org/10.1207/s15326985ep4102_4

Hogg, A. M., & Vaughan, M. G. (2018). *Social psychology* (8th ed.). Pearson Education Limited.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Krapivkina, A. O. (2018). Sight translation and its status in training of interpreters and translators. *Indonesian Journal of Applied Linguistics, 7*(3), 695-704.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174. http://dx.doi.org/10.2307/2529310

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*, 563-575.

Lee, J. **(**2012**)**. What skills do student interpreters need to learn in sight translation training? *Meta: Translators' Journal* [Meta: Journal des traducteurs], *57*(3), 694**-**714. https://doi.org/10.7202/1017087ar

Mehrabian, A. (1980). *Silent messages: Implicit communication of emotions and attitudes* (2nd ed.). Belmont, CA: Wadsworth.

Meyer, B. J. F., & Rice, G. E. (1982). The interaction of reader strategies and the organization of text. *Text-Interdisciplinary Journal for the Study of Discourse, 2*(1-3), 155-192. https://doi.org/10.1515/text.1.1982.2.1-3.155

Mikkelson, H. (1994). Text analysis exercises for sight translation. In W. Peter (Ed.), *Vistas: Proceedings of the 31st Annual Conference of the American Translators Association* (pp. 381–390). Learned Information.

Mikkelson, H., Willis, J., & Alvarez, N. (1995). *The interpreter's edge*. Spreckels, CA: ACEBO.

Pöckhacker, F. (2016). *Introducing interpreting studies* (2nd ed.). London: Routledge.

Smith, R., Snow, P., Serry, T., & Hammond, L. (2021). The role of background knowledge in reading comprehension: A critical review. *Reading Psychology, 42*(3), 214-240. https://doi.org/10.1080/02702711.2021.1888348

Suskie, L. (2009). *Assessing student learning: A common sense guide* (2nd ed.). San Francisco, CA: John Wiley & Sons, Inc.

Taherdoost, H. (2016). Sampling methods in research methodology; How to choose a sampling technique for research. *International Journal of Academic Research in Management, 5*, 18-27. https://doi.org/10.2139/ssrn.3205035

Tavakol, M., & Dennick, R. (2011). Making sense of Chronbach's Alpha. *International Journal of Medical Education, 2*, 53-55. http://dx.doi.org/10.5116/ijme.4dfb.8dfd

Vaughan, G. M., & Hogg. A. M. (1998). *Introduction to social psychology* (2nd ed.) Sydney, Australia: Prentice Hall.

Vaughan, G. M., & Hogg, A. M. (2002). *Introduction to social psychology* (3th ed.). Australia: Pearson Education.

Vural, H. (2021). The probable reasons for mistakes in second versions in simultaneous interpretation. In M. Kahyaoğlu (Ed.), *Multidisciplinary perspectives in educational and social sciences iv* (pp. 133-152). Ankara: İksad Publishing House.

Yamada, H. (2020). Efficacy of sight translation in English-Japanese consecutive interpreting training in a university course. *Theory and Practice in Language Studies, 10*(4), 343-352. http://dx.doi.org/10.17507/tpls.1004.01

8. EXTENDED ABSTRACT

Sight translation simply involves reading a text silently in the source language, and then translating it orally in the target language. Pöckhacker (2016, p. 20) defines it as the oral "rendition of a written text 'at sight'" from one language into another. Likewise, Chen (2015, p. 144) describes it as a type of "rendition of a written message in one language into oral form in another language". A sight translator needs to read ahead a written text in the source language before rendering it orally in another one. Chen (2015, p. 149) argued that a successful sight translator's "performance depends on two factors: comprehension of the source text and the ability to produce clear" or fluent delivery of the same message segments through the target text, that is, they have to develop the ability of reading skills in one language and the ability of speech production skills in another language.

It has been observed that the ST skills of the students are hardly identified and evaluated properly in the ST course because of the complexity of the phenomenon. Therefore, the general purpose of the study is to frame a comprehensive, transparent and coherent tool to help teachers in the assessment of the undergraduate students' ST performance. Rubrics are among the valid tools used to assess the students' ST performance based upon a specific set of criteria. They help teachers teach relevant concepts and validate their instruction and assessment. Rubrics, on the other hand, can make things clear for students and help them how to accomplish their work.

The proposed rubric tries to assess the ST performance of the students when they read a written text in English, and then translate them orally into their own native language (Turkish). It consists of two main domains of 'nonverbal' and 'verbal' communication skills. Nonverbal communication, sometimes called behavioural communication, refers to gestures, facial expressions, tone of voice, eye contact, body language, posture and other ways that people use to communicate with others without using the language. It is often done subconsciously. Verbal communication, on the other hand, is the use of spoken or written words to communicate information with others.

To produce an acceptable rubric, the study determined to test its validity and reliability, which are two fundamental elements in the evaluation of any tools. Validity is the extent to which a tool measures what it is intended to measure. There are different kinds of validity (i.e., construct, criterion, content and face), but the content validity is perhaps the most relevant one in the assessment of a tool. The validity model of Lawshe (1975) was used to test the content validity of the proposed rubric. This model measures the content validity ratio (CVR) of each item based upon the established critical values of a panel of judges and the content validity index (CVI) of the overall items, which is obtained from calculating the average of the CVR scores. When an item passes the

threshold value of Lawshe (1975) or the CVI of the entire test gets closer to 1, it is considered as an acceptable item and at the same time a valid test. The rubric was validated by 5 experts in the first trial, 8 ones in the second trial and 10 ones in the third trial. The results of the third trial, however, showed that all of the items of the proposed rubric had a critical value of over .62, indicating that the proposed rubric is a valid tool.

Reliability refers to the overall consistency of an instrument. A measure is said to be reliable if it produces similar results under the consistent conditions (Haradhan, 2017; Taherdoost, 2016). There are different kinds of reliability used to indicate the effectiveness of a measure: The inter-rater reliability, the test-retest reliability, the inter-method reliability and the internal consistency reliability (Haradhan, 2017; Taherdoost, 2016). However, for the purpose of the study the inter-rater reliability was used. It assesses the degree of agreement between two or more ratters. Using the proposed rubric, three ratters assessed the non-verbal and verbal performance of eighteen English language translation and interpreting students through a 1-4 Likert-type scale (Excellent=4, Good=2, Acceptable=3 and Poor=1) in three trials of a midterm exam, a quiz and a final exam (see appendix 1). The results of the third trial proved that the Kappa values of the items ranged between (0.61 - 0.80) and (0.81 to 1.00) for a 'substantial' and 'almost perfect' agreement levels, respectively. Thus, it was confirmed that the proposed rubric is a reliable tool.

9. APPENDIX ONE: THE INTERPRETING RUBRIC

The latest version of the proposed ST rubric was given in the following.

The Interpretation Rubric

You are required to read the given written English (SL) material and provide an appropriate interpretation of it in your own language (Turkish/TL) while using your background knowledge, memory and the text information to showcase your interpreting ability.

Nonverbal Skills

Criteria	Excellent=4	Good=3	Fair=2	Poor=1
Eye contact	Holds the attention of the entire audience consistently with the use of direct eye contact, seldom looking at the written material.	Displayed enough direct eye contact with the audience, but still returns to the written material.	Displayed minimal eye contact with the audience, while reading mostly from the written material.	No eye contact with the audience, as the entire information is read from the written material.
Gestures	Movements or descriptive gestures seem fluid and help her/him to visualize the content.	Movements or descriptive gestures enhance her/his articulation enough.	Very little movements or descriptive gestures.	No movements or descriptive gestures (e.g., Hands in pockets, hands behind back, or closed fists).
Proxemics	Student proves a relaxed, self-confident nature about self in the location.	Student displays little tension in the location: Standing with hunched shoulders.	Student displays mild tension in the location: Looking around, standing with crossed ankles, or fidgeting with objects.	Tension or nervousness is obvious in the location: Looking and moving around unreasonably.
Paralanguage	Student speaks so confidently and uses properly various paralanguage devices (i.e., pauses, fillers, speech rate, tone, pitch, loudness, emphasis) to enhance the clarity and fluency of her/his message. The audience can hear the speech and finds it interesting.	Student speaks confidently, and her/his use of paralanguage devices are reasonable. The audience can hear the speech but shows some interest to it.	Student speaks less assuredly, and his voice is low and monotonous. There are a few long pauses and vocal hesitations (e.g., 'um' or 'uh', etc.) in her/his speech. The audience has some difficulty in hearing the speech and shows little interest to it.	Student mumbles and speaks too quietly (or quickly). There are a lot of long pauses and vocal hesitations (e.g., 'um' or 'uh', etc.) in her/his speech. The audience has most difficulty in hearing the speech and shows no interest to it.

Verbal Skills

Organization	The presentation is very informative, logical, interesting, complete, and effective. It has a clear and focussed introduction, body, and conclusion. The topic is clearly developed, and paragraphs are linked with smooth and effective transitions. The audience can easily follow the presented information.	The presentation is mostly clear and informative. It has introduction, body, and conclusion. The topic is generally developed, and paragraphs are linked with transitions enough. The audience can follow the presented information.	The presentation is present but somehow unclear. Introduction, body, and conclusion are attempted, but are jumbled. Transitions are used minimally. The sentences used to develop the topic are related but are not in the order that they should be. The audience has difficulty following presentation.	The presentation is not discernible. There is no indication of paragraphing: introduction, body, and conclusion are missing and lacking required transitions. The sentences used to develop the topic are unrelated. The audience cannot understand and follow the presentation.
Memory span	Student processes a large amount of content information and retains them temporarily for an effective presentation.	Student processes enough content information and retains them temporarily to lead her/his presentation.	Student processes just the main points of the content and retains them temporarily to lead her/his presentation.	Student completely fails to process the content information and retain them temporarily to run her/his presentation.
Background knowledge	Student accurately bridges between her/his background knowledge and the content information in order to fully present the topic.	Student somehow links between her/his background knowledge and the content information during the presentation.	Student does not have relevant background knowledge but present the content information truly.	Student does not have relevant background knowledge and reads the text during her/his presentation.
Content knowledge	The content information was fully addressed; all of the main ideas were explored in-depth with appropriate supporting details. Even, the relevant content was elaborated with personal interpretations and claims.	The content information was addressed adequately; the main ideas were explored, but relevant supporting details were used occasionally.	The content information was minimally addressed; Some main ideas were explored, but they were not developed with supporting details.	Student did not have grasp of the content information and provided irrelevant information.
Expressiveness	Student uses a battery of speaking strategies, such as emphasizing key words, rephrasing, interpreting the meaning of words contextually, appealing for help to provide a very active, emotional, or passionate speech.	Student offers an active, emotional, or passionate speech.	The speech of the student is somehow active, emotional, or passionate.	Student fails to present an active, emotional, or passionate speech.
Enthusiasm	Demonstrates a strong, positive feeling about the topic during the entire presentation.	Occasionally shows positive feelings about the topic.	Shows some negativity toward the presented topic.	Shows absolutely no interest in the topic presented