



(ISSN: 2587-0238)

Yıldız, A., Es, H. & Türkdöğän, A. (2021). Development of Rubric for the Evaluation of Semantic Feature Analysis Chart in Mathematics, *International Journal of Education Technology and Scientific Researches*, 6(14), 64-89.

DOI: <http://dx.doi.org/10.35826/ijetsar.297>

Article Type (Makale Türü): Research Article

DEVELOPMENT OF RUBRIC FOR THE EVALUATION OF SEMANTIC FEATURE ANALYSIS CHART IN MATHEMATICS*

Ahmet YILDIZ

Dr. Teacher, Sivas Science and Art Center, Sivas, Turkey, ahmetyildiz58@gmail.com

ORCID: 0000-0002-9149-5859

Hasan ES

Assistant Professor, Gazi University, Ankara, Turkey, hasanes@gazi.edu.tr

ORCID: 0000-0002-7732-8173

Ali TÜRKDOĞAN

Assistant Professor, Cumhuriyet University, Sivas, Turkey, aliturkdogan@hotmail.com

ORCID: 0000-0003-0216-5426

Received: 16.12.2020

Accepted: 21.03.2021

Published: 01.04.2021

ABSTRACT

Correctly prepared semantic feature analysis charts can be used to concretize abstract mathematical concepts. This study, it is aimed to develop a valid and reliable rubric that can be used to evaluate the accuracy of the semantic feature analysis charts created by mathematics teachers. The data of the research consists of 36 semantic feature analysis charts created by mathematics teachers who have been trained in preparing a semantic feature analysis chart. During the validity study, the opinions of the experts about the rubric prepared were examined individually in the terms of content, structure and criteria. The percentages of concordance of expert opinions on sub-dimensions vary between 83% and 91%. The expert opinion concordance rate for the graded scoring key as a whole is found to be 87%. These calculated concordance percentages show that the rubric developed is valid. During the reliability study, the consistency between the scores evaluated for the semantic feature analysis charts by two mathematicians was calculated. The consistency of the raters in the sub-criteria was calculated with "Cohens's Kappa". A perfect fit in the sub-dimensions of instruction writing and text writing, and an acceptable fit in the sub-dimension of table design were observed. The concordance between the total scores of the two scorers was calculated with the "Spearman Correlation Coefficient" and a positive, significant and high level of concordance was found. These results show that the rubric is a valid and reliable tool in evaluating the semantic feature analysis charts used in maths lessons. The developed rubric can be used in other fields of education by making necessary changes.

Keywords: Mathematics education, semantic feature analysis, performance evaluation, rubric.

* This study is derived from the first author's doctoral dissertation, which he conducted under the supervision of the second and third authors in the Department of Mathematics Education at Gazi University Institute of Educational Sciences.

INTRODUCTION

Mathematics has an abstract nature. The most important reason for this situation stems from the abundance of abstract concepts. Schemes formed as a result of the association of many concepts with each other make this abstract structure even more complex. Hence, it is necessary to use alternative approaches, technologies and materials that provide meaningful and permanent learning by revealing the relationships between abstract concepts better in mathematics teaching. One of these alternative materials is "Semantic feature analysis chart (SFAC)".

SFAC uses a table to help students discover how a set of concepts related to each other (Mantasiah et al., 2020). While there are concepts or objects in one row/column of the table, there are properties in the other column/row (Taskin, 2008). Ausubel (1968), states that it is important to explain the descriptive and distinctive properties of a concept and relevant sub-concepts in the realization of meaningful learning. In this context, SFAC can facilitate meaningful learning.

SFAC enables the students to examine and associate between relevant concepts and properties. It also encourages them to learn about concepts and properties by analyzing the completed table. This means that they can concretize the linkages, make predictions, and better understand important concepts. Furthermore, it contributes to learning the relationships between familiar and less familiar terms, while at the same time learning the meaning of an unknown term by creating analogies and examples (Olsen, 2016).

Moreover, SFAC encourages the students to compare key terms cognitively by associating them with examples, functions, properties or attributes. Even if there is some uncertainty in the concepts in SFAC, the presence of the table contributes to a deeper understanding of the subject and increased interest (Wickens & Parker, 2019).

The table used in SFAC activities provides a representation of the information organized by the brain. With the help of this table, the students who recognize the similarities and differences between the concepts together obtain the opportunity to analyze the relationships between them and discuss their properties (Buehl & Alexander, 2001). During the discussion, the students are encouraged to differentiate the similarities, differences and hierarchical structure between the concepts. Thus, they gain the competence to see the relationships between them and to learn important concepts (Stahl & Vancil, 1986).

Thanks to SFAC, new words are connected with previously learned words and concept learning is realized. It can also be used as an effective tool in learning the descriptive and distinctive properties of concepts. The use of SFAC both stimulates the students' prior knowledge about the concept; making them more interested in the subject they will learn (Anders & Boys, 1986).

The students expressing concepts by using SFAC have the opportunity to recognize the relationships between concepts and to construct their own knowledge. If there is a shortcoming, mistake or misconception in the

student's conceptual knowledge, it is thus revealed. As a result of the discussions, the mistake or misconception of the student is eliminated. It should be kept in mind that SFAC can never replace a teacher. In other words, it is not sufficient alone in teaching the concepts. SFAC, which is not suitable for the student level, does not serve the desired purpose. On the other hand, the frequent use of SFAC may cause students to get bored (Cetinkaya & Tas, 2011).

Various information can be collected by using measurement tools to generate an idea about students' progress. The property to be measured (knowledge, skill, attitude, perception, etc.), criteria such as the purpose of the measurement to be carried out and the characteristics of the students are effective in determining the measurement tool. Knowledge tests such as written, short answers or multiple choices can be used to measure the academic success of students in any subject. However, performing a task (for example, creating a semantic feature analysis chart), in other words, measuring a performance displayed can be difficult with knowledge tests. Because knowing how to do a job may not mean doing this job according to the determined criteria. For this reason, student performances in education can be measured indirectly with the help of performance-related assessment tools. This process is called performance evaluation (Tekin, 2009).

In performance evaluation, the students are expected to use their knowledge and skills together and create a product as output (Stiggins, 1994). In this way, it can be measured to what extent the information is learned, as well as the extent to which it can be transferred to new or real-life situations (Biemer, 1993).

One of the common measurement tools that can be used in performance evaluation is rubrics. A rubric is an indirect scoring tool that lists the criteria for each performance and specifies what should be met in performance (Popham, 1997). In a rubric, it is explained that the performance that the learner should fulfill and which behaviors are more prominent in evaluating this performance (Arter & McTighe, 2000). Rubrics are valid and reliable measurement tools because they make the measurement process independent from the evaluator and time (Moskal & Leydens, 2000). In this way, the use of a rubric in evaluating the performance of learners can make the measurement process more reliable by minimizing the measurement errors caused by the rater (Dunbar et al., 2006).

Both the process and the result are evaluated together in rubrics. During the process, learners can identify what they need to do using a rubric. At the end of the process, the teacher can evaluate the resulting performance according to the determined criteria (Mertler, 2001; Popham, 1997). Rubrics consist of three parts (Popham 1997);

- Assessment Criteria: They are used to distinguish students' acceptable performance from their unacceptable performance.
- Criteria Definitions: They are used to indicate qualitative and quantitative differences in students' performance.

- Scoring Strategy: An analytical rubric can be used when the performance to be evaluated can be decomposed into lower dimensions and when a more detailed scoring is desired. If performance cannot be divided into sub-dimensions, a holistic rubric can be used (Haladyna, 1997).

This study aims to develop a rubric for evaluating the SFAC creation skills of mathematics teachers considering the abovementioned properties of rubrics. No rubric was encountered in the literature review for the evaluation of SFACs. Therefore, it is thought that the rubric to be developed within the scope of the study will be an important tool in evaluating the SFACs to be prepared in mathematics education.

METHOD

This section contains detailed information about the research design, working group, data collection tools and data analysis.

Research Design

This research is of a descriptive study in which a rubric that can be used to evaluate the SFAC is developed. Descriptive research can also be used to determine the level of behavior or performance (McMillan & Schumacher, 2014). In this study, a descriptive research model was preferred due to the development of a rubric that could be used in other studies about mathematics in the future.

Working Group

The study group of the research was determined by taking into account the convenient sampling and easy accessibility, one of non-random sampling methods. In purposeful sampling, it is tried to reach the most suitable people for the study by considering the aims of the research (Plano-Clark & Creswell, 2015). The easily accessible sampling method, on the other hand, enables the researcher to select situations that are close and easy to reach and accelerates the research (Yildirim & Simsek, 2013). In this context, volunteer mathematics teachers working in secondary schools in Sivas city center and who received training on preparing signification analysis were included in the study.

Data Collection Tools

The SFACs drawn by the teachers and the rubric prepared to evaluate these SFACs were used as a data collection tool in research. The teachers, who were trained in preparing an SFAC, drew an SFAC without any class level or learning area restrictions. Teachers drew a total of 36 SFACs in the fall semester of the 2018-2019 academic years.

A rubric was created to evaluate the SFAC to obtain research data. While developing a rubric, the steps suggested by Goodrich (2000), Haladyna (1997) and Moskal (2000) were followed. Information on these stages is presented in detail.

1. Determining the purpose of use of the key according to the expected performance: The purpose of the rubric created is to evaluate the SFACs created by mathematics teachers.
2. Determination of the scoring key type (analytical or holistic): While determining which type of rubric to use, the purpose of the evaluation was taken into consideration first. Also, it can be decomposed into lower dimensions due to the fact that there are many elements in the SFAC. For this reason, an analytical rubric was used as it was desired to examine the SFAC in each sub-dimension and to make a more detailed scoring.
3. Determination of the features (criteria) to be scored according to expected performance: The criteria to be included in the rubric (SFAC_Rubric) used in the evaluation of the SFAC, by examining various studies (Mantasiah et al., 2020; Olsen, 2016; Taskin, 2008; Wickens & Parker, 2019) (i) Instruction writing, (ii) Table design, (iii) Subject-concept-feature relationship, and (iv) Text writing.
4. Determining and scoring success levels for criteria: Teacher performance levels and scores; It is arranged as a "Beginner performance" (1 point), "Performance to be improved" (2 points), "Successful performance" (3 points) and "Excellent performance" (4 points). If nothing was done on any criterion, zero points were awarded for this criterion.
5. Making performance definitions for each success level of the criteria: Performance definitions have been made for all success levels, considering the characteristics of the criteria. These definitions; Mantasiah et al. (2020), Olsen (2016), Taskin (2008) and Wickens and Parker (2019) have been prepared by taking into account the points to be considered in the preparation of the semantic feature analysis chart. Complete and exemplary performances are defined as excellent performance, while the entry level is defined as the performance with significant deficiencies.
6. Taking expert opinions: The prepared rubrics were first examined by a Turkish teacher in terms of Turkish language rules. Then, expert opinion was obtained from two mathematics educators.
7. Pilot study of the developed scoring key: The pilot application of the developed rubric was carried out by evaluating various SFAC in the mathematics education literature.
8. Validity and reliability studies of the scoring key: The validity and reliability studies of SFAC_Rubric were carried out using the data obtained as a result of the pilot study.

The validity of rubrics is discussed in the context of content, structure and criteria (Moskal & Leydens, 2000). For this reason, to test the validity of the rubric, considering the questions stated by Boston (2002), it was asked to expert academicians in their field. These questions;

- Content: Do the evaluation criteria have content other than the specified performance? Do the evaluation criteria cover all aspects of performance?
- Structure: Are all important elements of performance assessed by scoring criteria available? Are there any evaluation criteria that are not related to the relevant performance?

- Measure: What are the important elements of the relevant performance to be evaluated? How do the scoring criteria reflect the elements of the relevant performance? How do scoring metrics measure important elements of relevant performance?

Rubric's reliability study was determined by examining the harmony between raters in the context of both classical response theory and item response theory. It has been determined that the percentages of agreement between the opinions of the experts on the suitability and competence of these to SFAC_Rubric vary between 83% and 91%. The content validity index of SFAC_Rubric in all dimensions is greater than the threshold value of 0.80.

Data Collection Process

During the research, within the scope of the first author's doctoral thesis, participating teachers created SFACs in the fall semester of the 2018-2019 academic year. The SFACs drawn were evaluated by the raters in the fall semester of the 2018-2019 academic years. The scorers consist of two experts who have a master's degree in mathematics education and have received training in preparing SFAC. Publication ethics were followed in collecting data and conducting the study. It was stated that the participants can participate in the research voluntarily and leave the research whenever they want. It has been committed to keep the names and institutions of the participants confidential. All researchers contributed to the study.

Analysis of Data

The consistency between the evaluations from expert academicians was calculated using the reliable formula suggested by Miles and Huberman (1994).

$$\text{Percentage of Compliance} = \frac{\text{Number of compromises}}{\text{Number of compromises} + \text{Number of disagreements}}$$

It is considered acceptable if the percentage of compliance obtained as a result of this calculation is above 70% (Tavsancil & Aslan, 2001).

The reliability study was carried out with the participation of two raters, considering the scores given to each criterion and the total of SFAC_Rubric. The reliability study was conducted by considering both classical test theory and item response theory. Within the scope of classical test theory, the level of consistency between the two raters was determined by the "Spearman Correlation Coefficient" since the scores did not show normal distribution. On the other hand, the level of consistency in each sub-criterion was determined by "Krippendorff Alpha" and "Cohen's Kappa". In the light of item response theory, the multi-faceted Rasch measurement model was used. While the analyzes within the scope of classical response theory were done with SPSS, analyzes within the scope of item response theory were made with the FACET program.

If Spearman correlation (Rho) is greater than 0.70, it can be said that there is a high level of relationship. For Cohen's kappa value, insignificant agreement for 0.00 - 0.20; Moderate agreement for 0.21-0.40; Acceptable

agreement for 0.41-0.60; Significant agreement for 0.61-0.80; and 0.81 - 1.00 are expressed as perfect match (Buyukozturk, 2007). For the Krippendorf Alpha value, values less than 0.67 are considered to be weakly fit, between 0.64-0.80 to be moderate and 0.80 and above as high-level fit (Krippendorf, 1995).

FINDINGS

Findings Regarding Validity

To evaluate the validity of the developed SFAC_Rubric, the answers and suggestions from the experts were taken into consideration. Experts' views on SFAC_Rubric were examined one by one in the context of content, structure and criteria. It has been determined that the percentages of agreement between the opinions of the experts on the suitability and competence of these to SFAC_Rubric vary between 83% and 91%. The expert opinion compliance percentage regarding the whole SFAC_Rubrics 87%. These calculated compliance percentages prove that the developed SFAC_Rubrics valid both in terms of content, structure and criterion sub-dimensions as well as all. These calculated compliance percentages prove that the developed SFAC_Rubric is valid in terms of content, structure and criteria.

To obtain detailed information about validity, the content validity index of SFAC_Rubric is calculated and the results are given in Table 1.

Table 1. Rubric Scope Validity Indexes

Dimensions	Questions	Scope Validity Index
Content	Do the evaluation criteria cover all aspects of performance?	1.00
	Do the evaluation criteria have non-performance content?	1.00
Structure	Are all important elements of performance assessed by scoring criteria?	1.00
	Are there any evaluation criteria that are not related to the relevant performance?	1.00
Criteria	Are the important elements of the relevant performance to be evaluated included?	1.00
	Do the scoring criteria accurately reflect the elements of the relevant performance?	1.00

When Table 1 is examined, it is seen that the content validity index of SFAC_Rubric in all dimensions is greater than the threshold value of 0.80. This result proves that the developed rubric is valid for evaluating the SFACs.

Also, some necessary corrections have been made to increase validity. For example, if there is no drawing for the feature that should be included in any sub-dimension of the SFAC, SFAC_Rubric has been updated to give zero points to this sub-dimension performance.

Findings Related to Reliability

The SFACs created by the teachers were evaluated by two separate scorers. Reliability studies were carried out by calculating both the total scores obtained from SFAC_Rubric and the scores obtained from each sub-dimension. The results of the Spearman correlation coefficient analysis conducted to determine the consistency between the total scores given by two different raters to SFAC_Rubric are given in Table 2.

Table 2. Spearman Correlation Results Regarding SFAC_Rubric’s Total Scores

		Scores of the first evaluator	Scores of the second evaluator
Scores of the first evaluator	Spearman's Rho	1.000	,884*
	P		.000
	N	36	36
Scores of the second evaluator	Spearman's Rho	,884*	1.000
	P	.000	
	N	36	36

When Table 2 is examined, it is seen that there is a positive, significant and high (Rho = 0.884, p <0.01) relationship between the scores given by the two evaluators to the rubric, which will be used in the evaluation of SFAC. This result shows that the prepared SFAC_Rubric is reliable enough in evaluating the semantic feature analysis charts.

Also, while the consensus between the scores given by two different raters to the SFAC_Rubric sub-criteria was determined with the Krippendorf Alpha Coefficient, the reliability of this consensus was analyzed with Cohen's Kappa coefficient. The results obtained are given in Table 3.

Table 3. Compliance between Raters for the SFAC_Rubric Sub-Criteria

Sub-Criteria	Alpha Coefficient	Cohen’s Kappa
Instruction Writing	0.879	0.943
Table Design	0.612	0.557
Subject-Concept-Feature relationship	0.695	0.743
Text Writing	0.875	0.840

When the alpha coefficients in Table 3 are examined, it is seen that there is a medium level of harmony in the "subject-concept-feature relationship" sub-dimension, and a high level of concordance in "instruction writing" and "text writing" sub-dimensions. When the Cohen's Kappa results are examined, there is a perfect fit in the sub-dimensions of "Instruction writing" and "text writing", and a significant level of agreement in the "Subject-concept-feature relationship" sub-dimension. Considering the alpha coefficient, although the fit in the "table design" sub-dimension is weak, Cohen's Kappa values are at the acceptable level of agreement in the "table design" sub-dimension

The reliability index results for surfaces within the scope of the multi-faceted Rasch model within the scope of item response theory are given in Table 4.

Table 4. SFAC_Rubric Reliability Index of Surfaces

Surfaces	Confidence Index
Rater	0.920
Scored (SFAC)	0.787
Items	0.719

When Table 4 is examined, it is seen that the reliability index of the “rater” surface is 0.920, the reliability index of the “scored (SFAC)” surface is 0.787, and the reliability index of the “items” surface is 0.719. These values show that the developed SFAC_Rubric is a reliable measurement tool.

Spearman correlation calculated for the total score of SFAC_Rubric; Krippendorf Alpha Coefficient and Cohen's Kappa values calculated for sub-criteria scores; when the reliability indices of the rater, scored (ACT) and item surfaces were evaluated together, it was seen that the SFAC_Rubric created was a reliable scoring key.

CONCLUSION and DISCUSSION

In this study, an analytical rubric was developed to evaluate mathematics teachers' ability to prepare “SFAC” and validity and reliability studies were conducted. The subject is included in SFAC_Rubric Annex 1.

The prepared SFAC_Rubric consists of four criteria to evaluate the ability to prepare SFAC. According to expert opinion, the content, structure and criteria of SFAC_Rubric are appropriate and sufficient to evaluate the ability to prepare SFAC. Also, the high consistency among the expert opinions shows that the prepared SFAC_Rubric is valid. The content validity indices calculated for examining the validity in more detail were also higher than the threshold value for each dimension.

It was observed that there was a positive, significant and high ($Rho = 0.884$, $p < 0.01$) relationship between the total scores given by the two evaluators to the SFAC_Rubric. While the Krippendorf Alpha Coefficient values for its sub-criteria varied between 0.612 and 0.879, Cohen's Kappa results were found to be between 0.557 and 0.943. Also, reliability calculations for rater scored, and item surfaces are at acceptable levels. Although these results show that the created SFAC_Rubric is a reliable scoring key, it also shows that the raters are not fully compatible. But it can be concluded that SFAC_Rubric increases objectivity by increasing the consistency between raters. In similar studies (Bilgen & Dogan, 2017; Buyukkidik & Anil, 2015; Knoch, 2009), it was stated that there was a high agreement among the rubric scores, although not 100 percent.

The difference between raters may also have varied according to the level of knowledge, perceptions and interpretations of raters about the SFAC. This is called the rater effect (Cronbach, 1990). In various studies, it is stated that rubric is insufficient in eliminating rater effects such as severity or generosity of rater (Guler & Gelbal, 2011; Ilhan, 2015). Raters' perception of the criteria differently may also be the reason for the difference between the scores. Bilgen and Dogan (2017) emphasized why the interaction between raters and criteria could lead to the difference between scores while examining the reliability of the rubric they developed in their study.

One of the teaching materials that prospective teachers prefer to use most is SFA (Karacengel et al., 2019). Secondary school students also say that they frequently encounter SFA in their classes (Soylu & Memisoglu, 2020). It is also important for teachers to develop different teaching materials such as SFA (Kutlu & Gokdere,

2020). In this context, it is important to evaluate SFAs developed by teachers by using SFAC_Rubric developed within the scope of the research.

RECOMMENDATIONS

Based on the findings of this research, some suggestions have been made to practitioners and other researchers. Developed SFAC_Rubric is made by considering the points given by two raters. Reliability studies can be conducted in which there are more raters and therefore different analysis techniques should be used. Since the developed SFAC_Rubric is used in the evaluation of the SFAC created by mathematics teachers, it can be investigated whether it can be used in the evaluation of the SFAC prepared by teacher candidates.

ETHICAL TEXT

In this article, journal writing rules, publishing principles, research and publishing ethics rules, journal ethics rules are followed. The authors are responsible for any violations that may arise regarding the article.

Author(s) Contribution Rate: The first author's contribution rate to this article is 40%. The second author's contribution rate to this article is 30%. The third author's contribution rate to this article is 30%.

REFERENCES

- Anders, R. L., & Boys, C. S. (1986). Semantic feature analysis: An interactive strategy for vocabulary development and text comprehension. *Journal of Reading*, 29(7), 610-616.
- Arter, J. A., & McTighe, J. (2000). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Corwin Press.
- Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. Holt.
- Biemer, L. (1993). Authentic assessment. *Educational Leadership*, 50(8), 81-82.
- Bilgen, B. Ö., & Doğan, N. (2017). Puanlayıcılar arası güvenilirlik belirleme tekniklerinin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(1), 63-78. <https://doi.org/10.21031/epod.294847>
- Bodner, G. M. (1986). Constructivism: A theory of knowledge. *Journal of Chemical Education*, 63(10), 873.
- Boston, C. (2002). *Understanding scoring rubrics: a Guide for teachers*. ERIC.
- Buehl, M. M., & Alexander, P. A. (2001). Beliefs about academic knowledge. *Educational Psychology Review*, 13(4), 385-418. <https://doi.org/10.1023/A:1011917914756>
- Büyükkıdık, S., & Anıl, D. (2015). Performansa dayalı durum belirlemede güvenilirliğin genellenebilirlik kuramında farklı desenlerde incelenmesi. *Eğitim ve Bilim*, 40(177), 285-296. <http://dx.doi.org/10.15390/EB.2015.2454>
- Büyükoztürk, Ş. (2007). *Sosyal bilimler için veri analizi el kitabı*. PegemA Yayıncılık.
- Karaçengel, H., Ceylan, F., Gedik, K., & Akbulut, H. İ. (2019, 12-14 April). *Fen bilimleri öğretmen adaylarının öğretim materyalleri ve kullanımı hakkındaki görüşlerinin belirlenmesi* [Conference presentation]. Uluslararası Fen, Matematik, Girişimcilik ve Teknoloji Eğitimi Kongresi, İzmir, Turkey.

- Cronbach, L. I. (1990). *Essentials of psychological testing*. Harper and Row.
- Çetinkaya, M., & Taş, E. (2011). Canlıların sınıflandırılması konusu için web destekli kavram haritaları ve anlam çözümlene tablolarının öğrenme üzerindeki etkisinin araştırılması. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, 16, 180-195.
- Dunbar, N. E., Brooks, C. F., & Miller, K. T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education*, 31(2), 115–28. <https://doi.org/10.1007/s10755-006-9012-x>
- Goodrich, A. H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 27-31.
- Güler, N., & Gelbal, S. (2010). Klasik test kuramı ve çok değişkenlik kaynaklı Rasch modeli üzerine bir çalışma. - *Eurasian Journal of Educational Research*, 38, 108-125.
- Haladyna, T. M. (1997). *Writing test item to evaluate higher order thinking*. Allyn & Bacon.
- İlhan, M. (2015). *Standart ve SOLO taksonomisine dayalı rubrikler ile puanlanan açık uçlu matematik sorularında puanlayıcı etkilerinin çok yüzeysel Rasch modeli ile incelenmesi* [Yayımlanmamış Doktora Tezi]. Gaziantep Üniversitesi.
- Johnson, D. D., & Pearson, P. D. (1984). *Teaching reading vocabulary*. Holt Rinehart & Winston.
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(20), 275-304. <https://doi.org/10.1177/0265532208101008>
- Krippendorff, K. (1995). On the reliability of unitizing continuous data. *Sociological Methodology*, 25, 47-76.
- Kutlu A. N., & Gökdere, M. (2020). Üstün yeteneklilere yönelik farklılaştırılmış fen öğretim modülü hakkında sınıf öğretmeni adaylarının kavramsal algıları ve değerlendirmeleri. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 17(1), 768-798. <https://doi.org/10.33711/yyuefd.751848>
- Mantasiah, R., Yusri, & Jufri. (2020). Semantic feature analysis model: Linguistics approach in foreign language learning material development. *International Journal of Instruction*, 13(1), 185-196. <https://doi.org/10.29333/iji.2020.13112a>
- McMillan, J. W., & Schumacher, S. (2014). *Research in education: Evidence-based inquiry*. Pearson.
- Mertler, C. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25), 1-8. <https://doi.org/10.7275/gcy8-0w24>
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*. Thousand.
- Moskal, B. M. (2000). Scoring rubric: What when and how? *Practical Assesment, Research & Evaluation*, 7(1), 1-5. <https://doi.org/10.7275/a5vq-7q66>
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation*, 7(10), 1-8. <https://doi.org/10.7275/q7rm-gg74>
- Olsen, M. B. (2016). *A semantic and pragmatic model of lexical and grammatical aspect*. Routledge.
- Plano-Clark, V. L., & Creswell, J. W. (2015). *Understanding research: A consumer's guide*. Pearson Education.
- Popham, W. J. (1997). What's wrong and what's right with rubrics. *Educational Leadership*, 55(2), 72-75.
- Soylu, T., & Memişoğlu, H. (2020). Ortaokul öğrencilerinin kavram öğretimine ilişkin görüşlerinin belirlenmesi. *Electronic Turkish Studies*, 15(2), 1307-1323. <http://dx.doi.org/10.29228/TurkishStudies.40142>

Stahl, S. A., & Vancil, S. J. (1986). Discussion is what makes semantic maps work in vocabulary instruction. *The Reading Teacher*, 40(1), 62-67.

Stiggins, R. J. (1994). *Student centered classroom assessment*. Macmillan Publishing Company.

Taşkın, Ö. (2008). *Fen ve teknoloji öğretiminde yeni yaklaşımlar*. Pegem Akademi Yayınları.

Tavşancıl, E., & Aslan E. (2001). *İçerik analizi ve uygulama örnekleri*. Epsilon Yayınları.

Tekin, H. (2009). *Eğitimde ölçme ve değerlendirme*. Yargı Yayınevi.

Wickens, C. M., & Parker, J. (2019) Supporting vocabulary acquisition in physical education settings. *Journal of Physical Education, Recreation & Dance*, 90(5), 16-22.

<https://doi.org/10.1080/07303084.2019.1580635>

Yıldırım, A., & Şimşek, H. (2013). *Sosyal bilimlerde nitel araştırma yöntemleri*. Seçkin Yayıncılık.

ANNEX 1. Rubric for the Evaluation of Semantic Feature Analysis Chart

	Beginner performance (1 point)	Performance to be improved (2 points)	Successful performance (3 points)	Excellent performance (4 points)	Point
Instruction Writing	The subject of the table has not been disclosed. How to fill in the table is not specified. How the table will be scored is not included.	The subject to which the table is related has been specified. How to fill in the table is not specified. How the table will be scored is not included.	The subject to which the table is related has been specified. How to fill in the table is explained. How the table will be scored is not included.	The subject to which the table is related has been specified. How to fill in the table is explained. It is explained how the table will be scored	
Table Design	Both concepts and attributes are not shown together in the same row or column. No blank cells are left to add concepts / features.	Concepts or features are not shown together in the same row or column. No blank cells are left to add concepts / features.	Concepts are shown together in the same row or column. Properties are shown together in the same row or column. No blank cells are left to add concepts / features.	Concepts are shown together in the same row or column. Properties are shown together in the same row or column. Blank cells are left for adding concept / features.	
Subject-Concept-Feature relationship	The chosen subject consists of concepts that do not have many common features. Concepts not related to the subject are also included in the table. Features not related to the concepts are also included in the table.	The chosen topic includes concepts with more common features. Concepts not related to the subject are also included in the table. Features not related to the concepts are also included in the table.	The chosen topic includes concepts with more common features. The concepts specified in the table are relevant to the subject. Features not related to the concepts are also included in the table.	The chosen topic includes concepts with more common features. The concepts specified in the table are relevant to the subject. The features specified in the table are related to the concepts.	
Text Writing	Scientific terms are not included. The terms are not clear, unambiguous or consistent. Language spelling rules were not followed.	Scientific terms are included. The terms are not clear, unambiguous or consistent. Language spelling rules were not followed.	Scientific terms are included. The terms are clear, concise, understandable and consistent. Language spelling rules were not followed.	Scientific terms are included. The terms are clear, concise, understandable and consistent. Language spelling rules were followed.	
TOTAL POINTS					

If nothing was done on any criterion, zero points were awarded for this criterion.

MATEMATİKSEL ANLAM ÇÖZÜMLEME TABLOSUNUN DEĞERLENDİRİLMESİNE YÖNELİK PUANLAMA ANAHTARININ GELİŞTİRİLMESİ*

Öz

Doğru bir şekilde hazırlanan anlam çözümlene tabloları soyut matematiksel kavramların somutlaştırılmasında kullanılabilir. Bu araştırmada matematik öğretmenlerinin oluşturdukları anlam çözümlene tablolarının doğruluğu değerlendirmede kullanılabilecek geçerli ve güvenilir bir dereceli puanlama anahtarı geliştirmek amaçlanmıştır. Araştırmanın verileri anlam çözümlene tablosu hazırlama ile ilgili eğitim almış olan matematik öğretmenlerinin oluşturduğu 36 adet anlam çözümlene tablosundan oluşmaktadır. Geçerlik çalışması kapsamında uzmanların hazırlanan dereceli puanlama anahtarına yönelik görüşleri içerik, yapı ve ölçüt bağlamında tek tek incelenmiştir. Alt boyutlara ilişkin uzman görüşleri arasındaki uyum yüzdelerinin %83 ile %91 arasında değişmektedir. Dereceli puanlama anahtarının tamamına ilişkin uzman görüşü uyum yüzdesi ise %87'dir. Hesaplanan bu uyum yüzdeleri, geliştirilen dereceli puanlama anahtarının geçerli olduğunu göstermektedir. Güvenirlik çalışması kapsamında ise iki matematik uzmanının anlam çözümlene tablolarına verdikleri puanlar arasındaki uyum hesaplanmıştır. Puanlayıcıların alt ölçütlerdeki uyumları "Cohens's Kappa" ile hesaplanmıştır. Yönerge yazımı ve metin yazımı alt boyutlarında mükemmel uyuma, tablo tasarımı alt boyutunda ise kabul edilebilir uyuma görülmüştür. İki puanlayıcının toplam puanları arasındaki uyum ise "Spearman Korelasyon Katsayısı" ile hesaplanarak pozitif, anlamlı ve yüksek düzeyde uyum görülmüştür. Elde edilen bu sonuçlar, geliştirilen dereceli puanlama anahtarının matematik dersinde kullanılan anlam çözümlene tablolarının değerlendirilmesinde geçerli ve güvenilir bir araç olduğunu göstermektedir. Geliştirilen dereceli puanlama anahtarı gerekli değişiklikler yapılarak diğer alanlarda da kullanılabilir.

Anahtar kelimeler: Matematik eğitimi, anlam çözümlene tablosu, performans değerlendirme, dereceli puanlama anahtarı.

* Bu çalışma, birinci yazarın ikinci ve üçüncü yazarların danışmanlığında Gazi Üniversitesi Eğitim Bilimleri Enstitüsü Matematik Eğitimi Anabilim Dalı'nda gerçekleştirdiği doktora tezinden türetilmiştir.

GİRİŞ

Matematik soyut bir yapıya sahiptir. Bu soyut yapının en önemli nedeni de soyut kavramların çokluğundan kaynaklanmaktadır. Çok sayıda kavramın bir birleriyle ilişkilmesi sonucu oluşan şemalar bu soyut yapıyı daha da karmaşık bir hale getirmektedir. Bu nedenle matematik öğretiminde öğrencilerin soyut kavramlar arasındaki ilişkileri daha iyi görerek anlamlı ve kalıcı öğrenmeyi sağlayan alternatif yaklaşımların, teknolojilerin ve materyallerin kullanılması gerekmektedir. Bu alternatif materyallerden bir tanesi de anlam çözümleme tablosudur (AÇT).

AÇT’de öğrencilerin bir dizi kavramın birbiriyle nasıl ilişkili olduğunu keşfetmelerine yardımcı olmak için bir tablo kullanır (Mantasiah vd., 2020). Tablonun bir satırında/sütununda kavramlar veya nesnelere bulunurken diğer sütununda/satırında da özellikler yer almaktadır (Taşkın, 2008). Ausubel (1968), anlamlı öğrenmenin gerçekleşmesinde bir kavramın ve bu kavramla ilgili alt kavramların tanımlayıcı ve ayırt edici özelliklerinin de açıklanması önemli olduğunu belirtir. Bu bağlamda AÇT anlamlı öğrenmenin gerçekleşmesini sağlayabilir.

AÇT, öğrencilerin ilgili kavramları ve özellikleri incelemelerine ve aralarındaki ilişkiyi kurmalarına yardımcı olur. Tamamlanan tabloyu analiz ederek, öğrencileri kavramları ve özellikleri öğrenmeye teşvik eder. Bu da öğrencilerin bağlantıları somutlaştırabilecekleri, tahmin yapabilecekleri ve önemli kavramları daha iyi anlayabilmeleri demektir. AÇT, öğrencilerin tanıdık ve daha az tanıdık terimler arasındaki ilişkileri öğrenmelerini sağlarken, aynı zamanda da bilinmeyen bir terimin anlamını analogiler ve örnekler oluşturarak öğrenmelerini sağlar (Olsen, 2016).

AÇT, öğrencileri bilişsel olarak karşılaştırmaya teşvik eder. Öğrencilerin anahtar terimleri örnekler, işlevler, özellikler veya niteliklerle ilişkilendirerek karşılaştırmalarına yardımcı olur. AÇT’de yer alan kavramlarda belirsizlik olsa bile tablonun varlığı konuyu daha derin bir kavrayışa ve ilgiye katkı sunar (Wickens & Parker, 2019).

AÇT aktivitelerinde kullanılan tablo beynin organize ettiği bilgilerin bir temsilinin ortaya çıkarılmasını sağlar. Bu tablo yardımıyla kavramların benzerliklerini ve farklılıklarını bir arada gören öğrenciler kavramlar arasındaki ilişkileri analiz edebilme ve özelliklerini tartışabilme imkanı bulmuş olurlar (Buehl & Alexander, 2001). Tartışma boyunca öğrenciler kavramların benzerliklerini, farklılıklarını ve aralarındaki hiyerarşik yapıyı görmeye teşvik edilmiş olurlar. Böylece öğrenciler kavramlar arasındaki ilişkileri görme ve önemli kavramlara hakim olma yeterliği kazanmış olurlar (Stahl & Vancil, 1986).

AÇT kullanan öğrenci öğrendiği sözcüklerin anlamlarını önceden öğrendiği sözcüklerle bağlayarak kavram öğrenimini gerçekleştirmiş olur. AÇT, ayrıca kavramların tanımlayıcı veya ayırt edici özelliklerinin öğrenilmesinde de etkili bir araç olarak kullanılabilir. Kavram ile ilgili öğrencilerin ön bilgilerini harekete geçiren AÇT öğrencilerin öğrenecekleri konuya daha fazla ilgi duymalarını da sağlar (Anders & Boys, 1986).

AÇT kullanarak kavramları ifade eden öğrenciler kavramlar arası ilişkileri görme ve kendi bilgilerini yapılandırma fırsatını bulmuş olurlar. Böylece öğrencinin kavram bilgisinde bir eksiklik, yanlışlık veya yanlış varsa ortaya çıkmış olur. Gerçekleştirilen tartışmalar neticesinde öğrencinin yanlışlığı veya kavram yanlışlığı ortadan kaldırılmış olur. Unutulmamalıdır ki AÇT hiçbir zaman öğretmenin yerini tutamaz. Yani kavramların öğretiminde tek başına yeterli değildir. Öğrenci seviyesine uygun olmayan AÇT istenilen amaca hizmet etmez. Ayrıca AÇT'nin çok sık kullanımı öğrencilerin sıkılmalarına sebep olabilir (Çetinkaya & Taş, 2011).

Öğrenciler hakkında fikir üretebilmek için ölçme araçları kullanılarak çeşitli bilgiler toplanabilir. Ölçülecek olan öge (bilgi, beceri, tutum, algı vb.); gerçekleştirilecek ölçmenin amaçları ve öğrencilerin özellikleri gibi kriterler ölçme aracının belirlenmesinde etkilidir. Öğrencilerin herhangi bir konudaki akademik başarısını ölçmek için yazılı, kısa yanıtı ya da çoktan seçmeli gibi bilgi testleri kullanılabilir. Ancak bir işin yerine getirilmesi (örneğin anlam çözümlene tablosu oluşturma) diğer bir ifadeyle sergilenen bir performansın ölçümü bilgi testleri ile zor olabilmektedir. Çünkü bir işi nasıl yapılacağını bilmek ile bu işi belirlenen ölçütlere göre yapabilmek anlamına gelmeyebilir. Bu nedenle eğitimde öğrenci performansları sergilenen performansla ilişkili değerlendirme araçları yardımıyla dolaylı olarak ölçülebilmektedir. Bu ölçüm işlemine de performans değerlendirme denilmektedir (Tekin, 2009).

Performans değerlendirmede öğrenciden bilgi ve becerisini bir araya kullanıp bir ürün oluşturarak göstermesi beklenmektedir (Stiggins, 1994). Böylelikle bilgilerin ne derece edinildiğinin yanı sıra bu bilgileri yeni durumlara ya da gerçek yaşam durumlarına ne derece aktarabildiği de ölçülebilmektedir (Biemer, 1993). Performans değerlendirmede kullanılacak yaygın ölçme araçlarından birisi de dereceli puanlama anahtarıdır [DPA]. DPA, her bir performans için ölçütleri (kriterleri) listeleme ve performansta nelerin yerine getirilmesi gerektiğini belirten dolaylı bir puanlama aracıdır (Popham, 1997). DPA'da, öğrenenin yerine getirmesi gereken performans ve bu performansın değerlendirilmesinde hangi davranışların daha ön planda olduğu açıklanmaktadır (Arter & McTighe, 2000).

DPA'lar, ölçme işlemi değerlendiriciden ve zamandan bağımsız hale getirmelerinden dolayı geçerli ve güvenilir ölçme araçlarıdır (Moskal & Leydens, 2000). Bu sayede öğrenenlerin performanslarını değerlendirmede DPA kullanımı puanlayıcı kaynaklı ölçme hatalarını asgari düzeye indirerek ölçme işlemi daha güvenilir kılabilir (Dunbar ve diğeri, 2006).

DPA'da hem süreç hem de sonuç birlikte değerlendirilir. Süreç boyunca öğrenenler neler yapmaları gerektiğini DPA'yı kullanarak tespit edebilirler. Süreç sonunda da öğretmen, ortaya çıkan performansı belirlenen ölçütlere göre değerlendirebilmektedir (Mertler, 2001; Popham, 1997).

DPA üç bölümden oluşmaktadır (Popham, 1997). Bunlar;

- Değerlendirme Ölçütleri: Öğrencilerin kabul edilebilir performanslarını kabul edilemez performanslarından ayırmak için kullanılır.

- Ölçüt Tanımlamaları: Öğrencilerin performanslarındaki niteliksel ve niceliksel farklılıkları belirtmek için kullanılır.
- Puanlama Stratejisi: Değerlendirilecek performans daha alt boyutlara ayrıştırılabildiğinde ve daha detaylı bir puanlama yapmak istenildiğinde analitik DPA kullanılabilir. Performans daha alt boyutlara ayrıştırılmıyor ise bütüncül DPA kullanılabilir (Haladyna, 1997).

DPA'ların yukarıda bahsedilen özellikleri dikkate alınarak matematik öğretmenlerinin anlam çözümü tablosu oluşturma becerilerini değerlendirmeye yönelik analitik bir DPA geliştirilmesi bu çalışmanın amacıdır. Anlam çözümü tablolarının değerlendirilmesine yönelik yapılan literatür incelemesinde herhangi bir dereceli puanlama anahtarıyla karşılaşılmaştır. Bu nedenle de çalışma kapsamında geliştirilecek olan analitik dereceli puanlama anahtarının matematik eğitiminde hazırlanacak anlam çözümü tablolarının değerlendirilmesinde önemli bir araç olacağı düşünülmektedir.

YÖNTEM

Bu kısımda araştırma deseni, çalışma grubu, veri toplama araçları ve veri analizi ile ilgili detaylı bilgiler yer almaktadır.

Araştırma Deseni

Bu araştırma, anlam çözümü tablolarını değerlendirmek için kullanılacak bir dereceli puanlama anahtarının geliştirildiği betimsel bir çalışmadır. Betimsel araştırmalar bir davranışın ya da performansın düzeyini tespit etmek için de kullanılabilir (McMillan & Schumacher, 2014). Bu çalışmada, ileride başka çalışmalarda da kullanılacak bir dereceli puanlama anahtarı geliştirilmesinden dolayı betimsel araştırma modeli tercih edilmiştir.

Çalışma Grubu

Araştırmanın çalışma grubu, seçkisiz olmayan örnekleme yöntemlerinden uygun örnekleme ve kolay ulaşılabilirlik dikkate alınarak belirlenmiştir. Amaçlı örneklemede, araştırmanın amaçları göz önünde bulundurularak çalışma için en uygun kişilere ulaşılmaya çalışılır (Plano-Clark & Creswell, 2015). Kolay ulaşılabilir örnekleme yöntemi ise araştırmacıya ulaşılması yakın ve kolay olan durumları seçmesine imkan sağlayarak araştırmaya hız ve pratik kazandırır (Yıldırım & Şimşek, 2013). Bu bağlamda, Sivas il merkezindeki ortaokullarda görev yapan ve anlam çözümü tablosu hazırlama ile ilgili eğitim almış, gönüllü matematik öğretmenleri araştırmaya dahil edilmiştir.

Veri Toplama Araçları

Araştırmada veri toplama aracı olarak; öğretmenlerin çizdikleri anlam çözümü tabloları ve bu anlam çözümü tablolarını değerlendirmek için hazırlanan dereceli puanlama anahtarı kullanılmıştır.

Anlam çözümlene tablosu hazırlama ile ilgili eğitim alan öğretmenler, herhangi bir sınıf düzeyi ya da öğrenme alanı kısıtlaması olmadan anlam çözümlene tablosu çizmişlerdir. Öğretmenler, 2018-2019 eğitim-öğretim yılı güz döneminde toplam 36 anlam çözümlene tablosu çizmişlerdir.

Araştırma verilerini elde etmek için anlam çözümlene tablolarını değerlendirebilmek için dereceli puanlama anahtarı oluşturulmuştur. DPA geliştirilirken Goodrich (2000), Haladyna (1997) ve Moskal'ın (2000) önerdikleri aşamalar takip edilmiştir. Bu aşamalara dair bilgiler aşağıda detaylı bir şekilde yer almaktadır;

1. İncelenen performans uygun puanlama anahtarının kullanım amacının belirlenmesi: Oluşturulan DPA'ların amacı matematik öğretmenlerinin oluşturdukları anlam çözümlene tablolarını değerlendirmektir.
2. Puanlama anahtarı türünün (analitik ya da bütünsel) belirlenmesi: Hangi tür dereceli puanlama anahtarının kullanılacağı belirlenirken değerlendirmenin amacı öncelikle dikkate alınmıştır. Ayrıca, anlam çözümlene tablosunun içerisinde birçok öge bulundurmasından dolayı daha alt boyutlara ayrıştırılabilmektedir. Bu nedenle anlam çözümlene tablolarının hem her bir alt boyutta inceleyebilmek hem de daha detaylı bir puanlama yapmak istenildiğinden analitik DPA kullanılmıştır.
3. Beklenen performans göre puanlanacak özelliklerin (ölçüt) belirlenmesi: Anlam çözümlene tablolarının değerlendirmesinde kullanılan dereceli puanlama anahtarında (AÇT_DPA) yer alacak ölçütler, çeşitli çalışmaların (Mantasiah vd., 2020; Olsen, 2016; Taşkın, 2008; Wickens & Parker, 2019) incelenmesiyle (i) Yönerge Yazımı, (ii) Tablo Tasarımı, (iii) Konu-Kavram-Özellik ilişkisi ve (iv) Metin Yazımı olacak şekilde belirlenmiştir.
4. Ölçütler için başarı düzeylerinin belirlenmesi ve puanlandırılması: Öğretmen performans düzeyleri ve puanları; Başlangıç düzeyinde performans (1 puan), Geliştirilmesi gereken performans (2 puan), Başarılı performans (3 puan) ve Mükemmel performans (4 puan) olacak şekilde düzenlenmiştir. Herhangi bir ölçütte hiçbir şey yapılmamışsa bu ölçüt için sıfır puan verilmiştir.
5. Ölçütlerin her bir başarı düzeyi için performans tanımlarının yapılması: Ölçütlerin özellikleri dikkate alınarak tüm başarı düzeyleri için performans tanımlamaları yapılmıştır. Bu tanımlamalar; Mantasiah vd. (2020), Olsen (2016), Taşkın (2008) ile Wickens ve Parker (2019) çalışmalarında belirttikleri, anlam çözümlene tablosunun hazırlanmasında dikkat edilmesi gereken hususlar göz önüne alınarak hazırlanmıştır. Başlangıç düzeyi önemli eksiklikleri olan performans olarak tanımlanırken eksiksiz ve örnek gösterilebilecek nitelikte olan performanslar mükemmel performans olarak tanımlanmıştır.
6. Uzman görüşlerinin alınması: Hazırlanan DPA'lar öncelikle bir Türkçe öğretmeni tarafından Türkçe dil kuralları bakımından incelenmiştir. Daha sonra da matematik eğitimcisi üç akademisyenden uzman görüşü alınmıştır.
7. Geliştirilen puanlama anahtarının pilot çalışması: Geliştirilen DPA'nın pilot uygulaması, matematik eğitimi literatüründe yer alan çeşitli anlam çözümlene tablolarının değerlendirilmesiyle gerçekleştirilmiştir.

8. Puanlama anahtarının geçerlik ve güvenilirlik çalışmaları: Katılımcı öğretmenlerce hazırlanan AÇT'ler kullanılarak AÇT_DPA'nın geçerlik ve güvenilirlik çalışmaları yapılmıştır.

DPA'ların geçerlikleri içerik, yapı ve ölçüt bağlamında ele alınmaktadır (Moskal & Leydens, 2000). Bu nedenle DPA'nın geçerliğini sınamak için Boston'ın (2002) belirttiği sorular dikkate alınarak alanında uzman akademisyenlere sorulmuştur. Bu sorular;

- *İçerik:* Değerlendirme ölçütleri belirlenen performansın dışında bir içeriğe sahip mi? Değerlendirme ölçütleri performansın tüm yönlerini kapsıyor mu?
- *Yapı:* Puanlama ölçütleriyle değerlendirilen performansın tüm önemli öğeleri mevcut mu? İlgili performansla ilişkili olmayan herhangi bir değerlendirme ölçütü var mı?
- *Ölçüt:* Değerlendirilecek ilgili performansın önemli öğeleri nelerdir? Puanlama ölçütleri, ilgili performansın öğelerini nasıl yansıtmaktadır? Puanlama ölçütleri ilgili performansın önemli öğelerini nasıl ölçmektedir? Şeklinde dir.

DPA'nın güvenilirlik çalışması ise puanlayıcılar arasındaki uyumun hem klasik tepki kuramı hem de madde tepki kuramı bağlamında incelenmesiyle belirlenmiştir. Alt boyutlara ilişkin uzman görüşleri arasındaki uyum yüzdeleri %83 ile %91 arasında değişmektedir. DPA'nın tamamına ilişkin uzman görüşü uyum yüzdesi ise %87'dir. Geliştirilen DPA'nın tüm boyutlardaki kapsam geçerlik indeksi, eşik değer olan 0.80'den büyüktür.

Veri Toplama Süreci

Araştırma kapsamında birinci yazarın doktora tez çalışması kapsamında katılımcı öğretmenler 2018-2019 eğitim-öğretim yılı güz döneminde anlam çözümlene tablolarını oluşturmuşlardır. Çizilen anlam çözümlene tabloları 2018-2019 eğitim-öğretim yılı güz döneminde puanlayıcılar tarafından değerlendirilmiştir. Puanlayıcılar, matematik eğitimi alanında yüksek lisans yapmış ve anlam çözümlene tablosu hazırlama ile ilgili eğitim almış iki uzmandan oluşmaktadır. Verilerin toplanmasında ve çalışmanın yapılmasında yayın etiğine uyulmuştur. Katılımcıların araştırmaya gönüllü olarak katılabileceği ve istedikleri zaman araştırmadan ayrılacakları belirtilmiştir. Katılımcıların isimlerinin ve kurumlarının gizli tutulacağı taahhüt edilmiştir.

Verilerin Analizi

Uzman akademisyenlerden gelen değerlendirmeler arasındaki uyum Miles ve Huberman (1994) tarafından önerilen güvenilir formülü ile hesaplanmıştır.

$$Uyum\ Yüzdesi = \frac{Uzlaşma\ sayısı}{Uzlaşma\ sayısı + Uzlaşmama\ sayısı}$$

Bu hesaplama sonucunda elde edilen uyum yüzdesinin %70'in üzerinde olması kabul edilebilir düzey olarak görülmektedir (Tavşancıl & Aslan, 2001).

Güvenirlik çalışması iki puanlayıcının katılımıyla AÇT_DPA'ların her bir ölçütüne ve toplamına verilen puanlar dikkate alınarak gerçekleştirilmiştir. Güvenirlik çalışması hem klasik test kuramı hem de madde tepki kuramı dikkate alınarak yapılmıştır. Klasik test kuramı kapsamında, iki puanlayıcı arasındaki tutarlılık düzeyi puanların normal dağılım göstermemesinden dolayı "Spearman Korelasyon Katsayısı" ile belirlenirken her bir alt ölçütteki tutarlılık düzeyi de "Krippendorff Alfa" ve "Cohens's Kappa" ile belirlenmiştir. Madde tepki kuramı bağlamında ise çok yüzeysel Rasch ölçme modeli kullanılmıştır. Klasik tepki kuramı kapsamındaki analizler SPSS ile yapılırken madde tepki kuramı kapsamındaki analizler FACET programı ile yapılmıştır.

Spearman korelasyonu (Rho) 0.70'ten büyükse puanlayıcılar arasında yüksek düzeyde bir ilişkinin olduğu söylenebilir. Cohen's kappa değeri için ise 0.00 - 0.20 önemsiz uyuma; 0.21-0.40 orta derecede uyuma; 0.41 - 0.60 kabul edilebilir uyuma; 0.61 - 0.80 önemli derecede uyuma; ve 0.81 - 1.00 mükemmel uyuma olarak ifade edilir (Büyüköztürk, 2007). Krippendorff Alfa değeri için ise 0.67'den küçük değerler için zayıf düzeyde uyum, 0.64-0.80 arası orta düzeyde uyum, 0.80 ve üzeri ise yüksek düzeyde uyum olarak değerlendirilmektedir (Krippendorff, 1995).

BULGULAR

Geçerlik ile İlgili Bulgular

Geliştirilen AÇT_DPA'nın geçerliğini değerlendirme amacıyla uzmanlardan gelen cevaplar ve öneriler dikkate alınmıştır. Uzmanların AÇT_PDA hakkındaki görüşleri içerik, yapı ve ölçüt bağlamında tek tek incelenmiştir. Alt boyutlara ilişkin uzman görüşleri arasındaki uyum yüzdeleri %83 ile %91 arasında değişmektedir. DPA'nın tamamına ilişkin uzman görüşü uyum yüzdesi ise %87'dir. Hesaplanan bu uyum yüzdeleri, geliştirilen AÇT_DPA'nın içerik, yapı ve ölçüt alt boyutları bakımından hem de tamamı bakımından geçerli olduğunu kanıtlar değerdedir.

Geçerliğe ilişkin detaylı bilgi edinmek için DPA'nın kapsam geçerlik indeksi hesaplanarak sonuçlara Tablo 1'de yer verilmiştir.

Tablo 1. Puanlama Anahtarı Kapsam Geçerlik İndeksleri

Boyutlar	Sorular	Kapsam Geçerlik İndeksi
İçerik	Değerlendirme ölçütleri performansın tüm yönlerini kapsıyor mudur?	1.00
	Değerlendirme ölçütleri performans dışı bir içeriğe sahip midir?	1.00
Yapı	Puanlama ölçütleriyle değerlendirilen performansın tüm önemli öğeleri mevcut mudur?	1.00
	İlgili performansla ilişkili olmayan herhangi bir değerlendirme ölçütü var mıdır?	1.00
Ölçüt	Değerlendirilecek ilgili performansın önemli öğelerine yer verilmiş midir?	1.00
	Puanlama ölçütleri, ilgili performansın öğelerini doğru olarak yansıtmakta mıdır?	1.00

Tablo 1 incelendiğinde geliştirilen DPA'nın tüm boyutlardaki kapsam geçerlik indeksinin eşik değer olan 0.80'den büyük olduğu görülmektedir. Bu sonuç, geliştirilen DPA'nın anlam çözümleme tablolarını değerlendirmek için geçerli olduğunu kanıtlar niteliktedir.

Ayrıca, geçerliği artırmak adına bazı gerekli düzeltmeler de yapılmıştır. Örneğin, anlam çözümleme tablosunun herhangi bir alt boyutunda barındırması gereken özelliğe ait hiçbir çizim yoksa bu alt boyut performansına sıfır puan verilecek şekilde AÇT_DPA güncellenmiştir.

Güvenirlilik ile İlgili Bulgular

Öğretmenlerin oluşturdukları anlam çözümleme tabloları iki ayrı puanlayıcı tarafından değerlendirilmiştir. Güvenirlilik çalışmaları hem AÇT_DPA'dan elde edilen toplam puanlar hem de her bir alt boyuttan elde edilen puanlar hesaplanarak gerçekleştirilmiştir.

İki farklı puanlayıcının AÇT_DPA'ya verdikleri toplam puanlar arasındaki tutarlığın belirlenmesi için yapılan Spearman korelasyon katsayısı analizine ilişkin sonuçlar Tablo 2' de yer almaktadır.

Tablo 2. AÇT_DPA'nın Toplam Puanlarına İlişkin Spearman Korelasyon Sonuçları

		Birinci değerlendiricinin puanları	İkinci değerlendiricinin puanları
Birinci değerlendiricinin puanları	Spearman's Rho	1.000	0.884*
	P		0.000
	N	36	36
İkinci değerlendiricinin puanları	Spearman's Rho	0.884*	1.000
	P	0.00	
	N	36	36

Tablo 2 incelendiğinde anlam çözümleme tablolarının değerlendirilmesinde kullanılacak olan DPA'ya iki değerlendiricinin verdiği puanlar arasında pozitif, anlamlı ve yüksek ($Rho=0,884$, $p<0.01$) bir ilişkinin olduğu görülmektedir. Bu sonuç hazırlanan AÇT_DPA'nın anlam çözümleme tablolarının değerlendirilmesinde yeterli düzeyde güvenilir olduğunu göstermektedir.

Ayrıca iki farklı puanlayıcının AÇT_DPA alt ölçütlerine verdikleri puanlar arasındaki uzlaşma Krippendorf Alfa Katsayısı ile belirlenirken bu uzlaşmanın güvenirliliği Cohen'in Kappa katsayısı ile analiz edilmiştir. Elde edilen sonuçlara Tablo 3'te yer verilmiştir.

Tablo 3. AÇT_DPA Alt Ölçütlerine Ait Puanlayıcılar Arası Uyum

Alt Ölçütler	Alfa Katsayısı	Cohen's Kappa
Yönerge Yazımı	0.879	0.943
Tablo Tasarımı	0.612	0.557
Konu-Kavram-Özellik ilişkisi	0.695	0.743
Metin Yazımı	0.875	0.840

Tablo 3'te bulunan alfa katsayıları incelendiğinde "konu-kavram-özellik ilişkisi" alt boyutunda orta düzeyde, "yönerge yazımı" ve "metin yazımı" alt boyutlarında yüksek düzeyde uyum olduğu görülmektedir. Cohen's Kappa sonuçları incelendiğinde ise "Yönerge yazımı" ve "metin yazımı" alt boyutlarında mükemmel uyuma, "Konu-kavram-özellik ilişkisi" alt boyutunda önemli derecede uyuma düzeyindedir. Her ne kadar Alfa katsayısı

dikkate alındığında “tablo tasarımı” alt boyutundaki uyum zayıf düzeyde olsa da Cohen’s Kappa değerleri “Tablo tasarımı” alt boyutunda kabul edilebilir uyuma düzeyindedir.

Madde tepki kuramı kapsamında çok yüzeyle Rasch modeli kapsamında yüzeylere ilişkin güvenilirlik indeksi sonuçları Tablo 4’te yer almaktadır.

Tablo 4. AÇT_DPA Yüzeylere Ait Güvenirlik İndeksleri

Yüzeyler	Güvenirlik İndeksi
Puanlayıcı	0.920
Puanlanan (AÇT)	0.787
Madde	0.719

Tablo 4 incelendiğinde “puanlayıcı” yüzeyinin güvenilirlik indeksinin 0.920, “puanlanan (AÇT)” yüzeyinin güvenilirlik indeksinin 0.787 ve “madde” yüzeyinin güvenilirlik indeksinin de 0.719 olduğu görülmektedir. Bu değerler geliştirilen AÇT_DPA’nın güvenilir bir ölçme aracı olduğunu göstermektedir.

AÇT_DPA’nın toplam puan için hesaplanan Spearman korelasyonu; alt ölçüt puanları için hesaplanan Krippendorff Alfa Katsayısı ile Cohen’s Kappa değerleri; puanlayıcı, puanlanan (AÇT) ve madde yüzeylerine ait güvenilirlik indeksleri birlikte değerlendirildiğinde oluşturulan AÇT_DPA’nın güvenilir bir puanlama anahtarı olduğu görülmüştür.

TARTIŞMA ve SONUÇ

Bu araştırmada matematik öğretmenlerinin anlam çözümlene tablosu hazırlama becerilerini değerlendirmek için analitik bir dereceli puanlama anahtarı geliştirilerek geçerlik ve güvenilirlik çalışmaları yapılmıştır. Söz konusu AÇT_DPA Ek-1 de yer almaktadır.

Hazırlanan AÇT_DPA, anlam çözümlene tablosu hazırlama becerisini değerlendirmek üzere dört ölçütten oluşmaktadır. Uzman görüşüne göre AÇT_DPA’nın içeriği, yapısı ve ölçütleri anlam çözümlene tablosu hazırlama becerisini değerlendirmede uygun ve yeterlidir. Ayrıca, uzman görüşleri arasındaki yüksek tutarlılık hazırlanan AÇT_DPA’nın geçerliği olduğunu göstermektedir. Geçerliği daha detaylı incelemek amacıyla hesaplanan kapsam geçerlik indeksleri de her bir boyut için eşik değerden yüksek çıkmıştır.

İki değerlendiricinin AÇT_DPA’ya verdikleri toplam puanlar arasında pozitif, anlamlı ve yüksek ($Rho=0,884$, $p<0.01$) bir ilişkinin olduğu görülmüştür. Alt ölçütlerine ilişkin Krippendorff Alfa Katsayısı değerleri 0.612 ile 0.879 arasında değişirken, Cohen’s Kappa sonuçlarının 0.557 ve 0.943 arasında yer aldığı görülmüştür. Ayrıca puanlayıcı, puanlanan ve madde yüzeyleri için güvenilirlik hesaplamaları da kabul edilebilir düzeydedir. Bu sonuçlar oluşturulan AÇT_DPA’nın güvenilir bir puanlama anahtarı olduğu göstermekle birlikte puanlayıcıların tamamen uyumlu olmadığını da göstermektedir. Ama AÇT_DPA’nın puanlayıcılar arasındaki tutarlığı artırarak objektifliği artırdığı sonucuna varılabilir. Benzer çalışmalarda da (Bilgen & Doğan, 2017; Büyükkıdık & Anıl, 2015; Knoch, 2009) DPA puanlayıcıları arasında yüzde yüz uyum olmasa da yüksek bir uyumun olduğu belirtilmiştir.

Puanlayıcılar arasındaki farklılık puanlayıcıların anlam çözümleme tablosu hakkındaki bilgi düzeyleri, algıları ve yorumlamalarına göre de değişiklik göstermiş olabilir. Buna puanlayıcı etkisi denilmektedir (Cronbach, 1990). Çeşitli araştırmalar da puanlayıcının katılığı ya da cömertliği gibi puanlayıcı etkilerini ortadan kaldırmada DPA'nın yetersiz olduğu belirtilmektedir (Güler & Gelbal, 2011; İlhan, 2015).

Puanlayıcıların ölçütleri farklı algılamış olmaları da puanlar arasındaki farklılığın nedeni olabilir. Bilgen ve Doğan (2017) yaptıkları çalışmada geliştirdikleri DPA'nın güvenilirliğini incelerken puanlayıcılar ile ölçütler arasındaki etkileşimin puanlar arasındaki farklılığa ne den olabildiğini vurgulamışlardır.

Öğretmen adaylarının en çok kullanmayı tercih ettikleri öğretim materyallerinden birisi de SFA'dır (Karaçengel vd., 2019). Ortaokul öğrencileri de derslerinde sık sık SFA ile karşılaştıklarını söylemektedirler (Soylu & Memişoğlu, 2020). Ayrıca öğretmenlerin SFA gibi farklı öğretim materyalleri geliştirmeler önemlidir (Kutlu & Gökdere, 2020). Bu bağlamda öğretmenlerin geliştirdikleri SFA'ların araştırma kapsamında geliştirilen dereceli puanlama anahtarı kullanılarak değerlendirilmesi önem kazanmaktadır.

ÖNERİLER

Bu araştırmanın bulgularından hareketle uygulayıcılara ve diğer araştırmacılara bazı önerilerde bulunulmuştur. Geliştirilen AÇT_DPA iki puanlayıcının verdikleri puanlar dikkate alınarak yapılmıştır. Geliştirilen AÇT_DPA, matematik öğretmenlerinin oluşturdukları anlam çözümleme tablolarının değerlendirilmesinde kullanıldığı için öğretmen adaylarının hazırladıkları anlam çözümleme tablolarının değerlendirilmesinde de kullanılabilir olup olmadığı araştırılabilir.

Etik Metni

Bu makalede dergi yazım kurallarına, yayın ilkelerine, araştırma ve yayın etiği kurallarına, dergi etik kurallarına uyulmuştur. Makale ile ilgili doğabilecek her türlü ihlallerde sorumluluk yazarlara aittir.

Yazar(lar)ın Katkı Oranı Beyanı: Çalışmaya tüm araştırmacılar katkı sunmuştur. Birinci yazarın bu makaleye katkı oranı %40'dır. İkinci yazarın bu makaleye katkı oranı %30'dur. Üçüncü yazarın bu makaleye katkı oranı ise %30'dur.

KAYNAKÇA

Anders, R. L., & Boys, C. S. (1986). Semantic feature analysis: An interactive strategy for vocabulary development and text comprehension. *Journal of Reading, 29*(7), 610-616.

Arter, J. A., & McTighe, J. (2000). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Corwin Press.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. Holt.

Biemer, L. (1993). Authentic assessment. *Educational Leadership, 50*(8), 81-82.

- Bilgen, B. Ö., & Doğan, N. (2017). Puanlayıcılar arası güvenilirlik belirleme tekniklerinin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(1), 63-78. <https://doi.org/10.21031/epod.294847>
- Bodner, G. M. (1986). Constructivism: A theory of knowledge. *Journal of Chemical Education*, 63(10), 873.
- Boston, C. (2002). *Understanding scoring rubrics: a Guide for teachers*. ERIC.
- Buehl, M. M., & Alexander, P. A. (2001). Beliefs about academic knowledge. *Educational Psychology Review*, 13(4), 385-418. <https://doi.org/10.1023/A:1011917914756>
- Büyükkıdık, S., & Anıl, D. (2015). Performansa dayalı durum belirlemede güvenilirliğin genellenebilirlik kuramında farklı desenlerde incelenmesi. *Eğitim ve Bilim*, 40(177), 285-296. <http://dx.doi.org/10.15390/EB.2015.2454>
- Büyüköztürk, Ş. (2007). *Sosyal bilimler için veri analizi el kitabı*. PegemA Yayıncılık.
- Karaçengel, H., Ceylan, F., Gedik, K., & Akbulut, H. İ. (2019, 12-14 April). *Fen bilimleri öğretmen adaylarının öğretim materyalleri ve kullanımı hakkındaki görüşlerinin belirlenmesi* [Conference presentation]. Uluslararası Fen, Matematik, Girişimcilik ve Teknoloji Eğitimi Kongresi, İzmir, Turkey.
- Cronbach, L. I. (1990). *Essentials of psychological testing*. Harper and Row.
- Çetinkaya, M., & Taş, E. (2011). Canlıların sınıflandırılması konusu için web destekli kavram haritaları ve anlam çözümleme tablolarının öğrenme üzerindeki etkisinin araştırılması. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, 16, 180-195.
- Dunbar, N. E., Brooks, C. F., & Miller, K. T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education*, 31(2), 115-28. <https://doi.org/10.1007/s10755-006-9012-x>
- Goodrich, A. H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 27-31.
- Güler, N., & Gelbal, S. (2010). Klasik test kuramı ve çok değişkenlik kaynaklı Rasch modeli üzerine bir çalışma. - *Eurasian Journal of Educational Research*, 38, 108-125.
- Haladyna, T. M. (1997). *Writing test item to evaluate higher order thinking*. Allyn & Bacon.
- İlhan, M. (2015). *Standart ve SOLO taksonomisine dayalı rubrikler ile puanlanan açık uçlu matematik sorularında puanlayıcı etkilerinin çok yüzeysel Rasch modeli ile incelenmesi* [Yayımlanmamış Doktora Tezi]. Gaziantep Üniversitesi.
- Johnson, D. D., & Pearson, P. D. (1984). *Teaching reading vocabulary*. Holt Rinehart & Winston.
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(20), 275-304. <https://doi.org/10.1177/0265532208101008>
- Krippendorff, K. (1995). On the reliability of unitizing continuous data. *Sociological Methodology*, 25, 47-76.
- Kutlu A. N., & Gökdere, M. (2020). Üstün yeteneklilere yönelik farklılaştırılmış fen öğretim modülü hakkında sınıf öğretmeni adaylarının kavramsal algıları ve değerlendirmeleri. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 17(1), 768-798. <https://doi.org/10.33711/yyuefd.751848>
- Mantasiah, R., Yusri, & Jufri. (2020). Semantic feature analysis model: Linguistics approach in foreign language learning material development. *International Journal of Instruction*, 13(1), 185-196. <https://doi.org/10.29333/iji.2020.13112a>

- McMillan, J. W., & Schumacher, S. (2014). *Research in education: Evidence-based inquiry*. Pearson.
- Mertler, C. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25), 1-8. <https://doi.org/10.7275/gcy8-0w24>
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*. Thousand.
- Moskal, B. M. (2000). Scoring rubric: What when and how? *Practical Assessment, Research & Evaluation*, 7(1), 1-5. <https://doi.org/10.7275/a5vq-7q66>
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation*, 7(10), 1-8. <https://doi.org/10.7275/q7rm-gg74>
- Olsen, M. B. (2016). *A semantic and pragmatic model of lexical and grammatical aspect*. Routledge.
- Plano-Clark, V. L., & Creswell, J. W. (2015). *Understanding research: A consumer's guide*. Pearson Education.
- Popham, W. J. (1997). What's wrong and what's right with rubrics. *Educational Leadership*, 55(2), 72-75.
- Soylu, T., & Memişoğlu, H. (2020). Ortaokul öğrencilerinin kavram öğretimine ilişkin görüşlerinin belirlenmesi. *Electronic Turkish Studies*, 15(2), 1307-1323. <http://dx.doi.org/10.29228/TurkishStudies.40142>
- Stahl, S. A., & Vancil, S. J. (1986). Discussion is what makes semantic maps work in vocabulary instruction. *The Reading Teacher*, 40(1), 62-67.
- Stiggins, R. J. (1994). *Student centered classroom assessment*. Macmillan Publishing Company.
- Taşkın, Ö. (2008). *Fen ve teknoloji öğretiminde yeni yaklaşımlar*. Pegem Akademi Yayınları.
- Tavşancıl, E., & Aslan E. (2001). *İçerik analizi ve uygulama örnekleri*. Epsilon Yayınları.
- Tekin, H. (2009). *Eğitimde ölçme ve değerlendirme*. Yargı Yayınevi.
- Wickens, C. M., & Parker, J. (2019) Supporting vocabulary acquisition in physical education settings. *Journal of Physical Education, Recreation & Dance*, 90(5), 16-22. <https://doi.org/10.1080/07303084.2019.1580635>
- Yıldırım, A., & Şimşek, H. (2013). *Sosyal bilimlerde nitel araştırma yöntemleri*. Seçkin Yayıncılık.

EK 1. Anlam Çözümleme Tablosu için Analitik Dereceli Puanlama Anahtarı

ÖLÇÜTLER	Başlangıç Düzeyindeki Performans (1)	Geliştirilmesi Gereken Performans (2)	Başarılı performans (3)	Mükemmel performans (4)	Puan
Yönerge Yazımı	Anlam çözümleme tablosunun ilgili olduğu konu açıklanmamıştır. Anlam çözümleme tablosunun nasıl doldurulacağı belirtilmemiştir. Anlam çözümleme tablosunun nasıl puanlandırılacağına yer verilmemiştir.	Anlam çözümleme tablosunun ilgili olduğu konu belirtilmiştir. Anlam çözümleme tablosunun nasıl doldurulacağı belirtilmemiştir. Anlam çözümleme tablosunun nasıl puanlandırılacağına yer verilmemiştir.	Anlam çözümleme tablosunun ilgili olduğu konu belirtilmiştir. Anlam çözümleme tablosunun nasıl doldurulacağı açıklanmıştır. Anlam çözümleme tablosunun nasıl puanlandırılacağına yer verilmemiştir.	Anlam çözümleme tablosunun ilgili olduğu konu belirtilmiştir. Anlam çözümleme tablosunun nasıl doldurulacağı açıklanmıştır. Anlam çözümleme tablosunun nasıl puanlandırılacağı açıklanmıştır.	
Tablo Tasarımı	Hem kavramlar hem de özellikler bir arada aynı satır ya da sütunda gösterilmemiştir. Kavram / özellik ekleyebilmek için boş hücreler bırakılmamıştır.	Kavramlar ya da özellikler bir arada aynı satır ya da sütunda gösterilmemiştir. Kavram / özellik ekleyebilmek için boş hücreler bırakılmamıştır.	Kavramlar bir arada aynı satır ya da sütunda gösterilmiştir. Özellikler bir arada aynı satır ya da sütunda gösterilmiştir. Kavram / özellik ekleyebilmek için boş hücreler bırakılmamıştır.	Kavramlar bir arada aynı satır ya da sütunda gösterilmiştir. Özellikler bir arada aynı satır ya da sütunda gösterilmiştir. Kavram / özellik ekleyebilmek için boş hücreler bırakılmamıştır.	
Konu-Kavram-Özellik ilişkisi	Seçilen konu ortak özelliği fazla olmayan kavramlardan oluşmaktadır. Konu ile ilgili olmayan kavramlar da tabloda yer almaktadır. Kavramlar ile ilgili olmayan özellikler de tabloda yer almaktadır.	Seçilen konu ortak özelliği fazla olan kavramları içerir. Konu ile ilgili olmayan kavramlar da tabloda yer almaktadır. Kavramlar ile ilgili olmayan özellikler de tabloda yer almaktadır.	Seçilen konu ortak özelliği fazla olan kavramları içerir. Tabloda belirtilen kavramlar konu ile ilgilidir. Kavramlar ile ilgili olmayan özellikler de tabloda yer almaktadır.	Seçilen konu ortak özelliği fazla olan kavramları içerir. Tabloda belirtilen kavramlar konu ile ilgilidir. Tabloda belirtilen özellikler kavramlar ile ilgilidir.	
Metin Yazımı	Bilimsel terimlere yer verilmemiştir. Terimler açık, net, anlaşılır ve tutarlı değildir. Dil yazım kurallarına uyulmamıştır.	Bilimsel terimlere yer verilmiştir. Terimler açık, net, anlaşılır ve tutarlı değildir. Dil yazım kurallarına uyulmamıştır.	Bilimsel terimlere yer verilmiştir. Terimler açık, net, anlaşılır ve tutarlıdır. Dil yazım kurallarına uyulmamıştır.	Bilimsel terimlere yer verilmiştir. Terimler açık, net, anlaşılır ve tutarlıdır. Dil yazım kurallarına uyulmuştur.	
Toplam Puan					

Herhangi bir ölçütte hiçbir şey yapılmamışsa bu ölçüt için sıfır puan verilir.