




Developing Achievement Test on Cell Subject for 6th Grade: A Validity and Reliability Study

Serkan TİMUR^{a*}, Fatih DOĞAN^a, Nagihan İMER ÇETİN^a, Betül TİMUR^a, Rüveyda IŞIK^b

^aÇanakkale Onsekiz Mart Üniversitesi Eğitim Fakültesi, Çanakkale/Türkiye

^bÇanakkale Onsekiz Mart Üniversitesi Eğitim Bilimleri Enstitüsü, Çanakkale/Türkiye 

Article Info

DOI: 10.14812/cufej.602535

Article History:

Received 07.08.2019

Revised 15.10.2019

Accepted 19.10.2019

Keywords:

Achievement test

Test development,

Cell,

Secondary school

Abstract

In this study, it is aimed to develop a valid and reliable achievement test for the subject of "Cell" belonging to the 6th grade "Systems in our Body" unit within the scope of Science course. The achievement test was prepared as a 25-item multiple-choice test, considering the cell-related objectives in the 6th grade Science curriculum. The pilot study of the test was carried out with 62 students attending 6th grade in 2017-2018 academic year. The comprehensibility and the consistency with the scientific knowledge of the test items were prepared by the opinions of various experts in the field of science education. The item analysis was performed according to the answers of the students in the achievement test, and the discriminant and difficulty indexes were calculated for each item. After the item analysis, 3 items were excluded from the test and a Cell Achievement Test (CAT) consisting of 22 questions was obtained. The mean item difficulty of CAT was 0.652 and the mean discrimination index was 0.507.

6. Sınıf Hücre Konusuna İlişkin Başarı Testi Geliştirilmesi: Geçerlik ve Güvenirlik Çalışması

Makale Bilgisi

DOI: 10.14812/cufej.602535

Makale Geçmişi:

Geliş 07.08.2019

Düzeltilme 15.10.2019

Kabul 19.10.2019

Anahtar Kelimeler:

Başarı testi,

Test geliştirme,

Hücre,

Ortaokul

Öz

Bu çalışmada Fen Bilimleri dersi kapsamında 6. Sınıf "Vücudumuzdaki Sistemler" ünitesine ait "Hücre" konusuna yönelik geçerli ve güvenilir bir başarı testinin geliştirilmesi amaçlanmıştır. Başarı testi, 6. sınıf Fen Bilimleri dersi öğretim programında yer alan hücre konusuna ait kazanımlar dikkate alınarak 25 madde çoktan seçmeli test olarak hazırlanmıştır. Testin pilot çalışması, 2017-2018 eğitim öğretim yılında 6. Sınıfında öğrenim gören 62 öğrenci ile yapılmıştır. Hazırlanan test maddelerinin anlaşılabilirliği, bilimsel bilgilerle tutarlılığı fen bilgisi eğitimi alanında çeşitli uzmanların görüşleri alınarak düzenlenmiştir. Öğrencilerin başarı testinde bulunan maddelere verdikleri cevaplara göre madde analizi yapılmış ve böylece her bir maddeye ilişkin ayırt edicilik ve güçlük indeksleri hesaplanmıştır. Madde analizi yapıldıktan sonra 3 madde testten çıkartılmış ve 22 sorudan oluşan Hücre Başarı Testi (HBT) elde edilmiştir. HBT'nin ortalama madde gücü 0.652, ortalama madde ayırt ediciliği ise 0.507 olarak hesaplanmıştır.

* Author: serkantimur42@gmail.com

Introduction

In education system, we answer the question “How much did we teach?” through measurement and evaluation (Temizkan & Sallabaş, 2011). Therefore, measurement and evaluation is an inevitable element for education. To what extent a student has achieved intended or desired teaching achievement can only be determined by measurement and evaluation. In other words, to what extent students are gained the desired qualifications which are taught in the school is revealed in this way (Atılğan, 2013). Baykul (2000) stated the importance of measurement and evaluation is to determine the durability of educational programs, to specify the effectiveness of the methods and techniques used in teaching, most importantly to determine the learning difficulties in students and to measure the success of the students and whether they have achieved the desired objectives. These descriptions show that measurement and evaluation is the most important and complementary element of the education.

Traditional measurement and evaluation tools in literature are stated as open-ended written surveys, multiple choice questions, true-false questions, gap filling questions and matching questions (Karip, 2012). Open ended written surveys are measurement tools in which students are asked one or more questions and are asked to indicate the answer by writing in a certain order (Temizkan & Sallabaş, 2011). However, due to the implementation and scoring of the written surveys’ requiring a lot of time, they are not preferred so much. Therefore, practitioners limit the number of questions in written surveys examinations. But this reduces the scope validity of the written surveys because the questions do not represent the population (Tekin, 2010). Another assessment tool is true-false questions. They are based on questions given to the student in which students classify the sentence or proposition as “True (T)” or “False (F). One of the advantages of this technique is that it is easy to prepare and answer questions. The high number of questions increases the scope validity of this test. However, the biggest disadvantage of this measurement tool is that the student has a high chance of finding the right answer (50%) (Erkuş, 2006). The gap filling questions are measurement tool based on subtraction of a word from the text and in which students are asked to fill this extracted word writing as same or synonym. (Akyol, 2012). Gap filling questions, which can be applied both in groups and individually, are a useful technique. In addition, due to the student's responses to the extracted word is short, scoring these tests are easy and objective (Erkuş, 2006). Matching questions are tests in which students are asked to make the right match with the words given in one column to the other words given in the other column. Preparing and scoring these tests are very easy The most important feature of this measurement tool is that it tests behaviors at the level of remembering (Erkuş, 2006). One of the most commonly used assessment tools in the education system is multiple choice test. Multiple-choice tests, due to that they nclude too many questions that increase the validity of the test and that they are practical and are not too long, are preffered too much. However, multiple-choice tests had been qualified as being the most superior measurement tool ever found (Özçelik, 1998). In multiple choice tests, the correct answer is given by the tester among the answer options and the test respondents try to find this correct answer (Güler, 2017). As it can be understood from this definition, multiple-choice tests are one of the most widely used measurement tools in the education system, although they limit the thoughts, comments and concepts of the respondent. Because these tests save time especially for crowded groups. Easy scoring is another advantage for these tests (Birgili, 2014). However, it is easy to ensure the reliability of scoring in these tests (Klufa, 2015).

When the related literature is examined, it is seen that multiple choice tests developed to measure students' success are quite a lot (Akbulut & Çepni 2013; Altunkaya, 2016; Ayvaci & Durmuş, 2016; Demir, Kızılay & Bektaş, 2016; Fidan, 2013; Gönen, Kocakaya & Kocakaya, 2011; Narlı & Başer, 2008; Singh & Rosengrant, 2003; Saylan Kırmızıgül & Kaya, 2019; Şahin & Keşan, 2017). Some of these tests developed in relation to science education are as follows; acids and bases (Demirci, 2011), buoyancy and swimming (Akbulut, 2010), electrical circuits (Can, Sen & Eryılmaz, 2011), environmental problems (Güven, 2013), solutions (Çalik, 2006; Demir, Kızılay & Bektaş, 2016), dynamic (Gönen, Kocakaya & Kocakaya, 2011), genetics (Çakır & Aldemir, 2011), solar system (Karamustafaoğlu & Tutar, 2016), heat and temperature (Ayvaci & Durmuş, 2016), force and motion (Akbulut & Çepni, 2013; İdin & Aydoğdu, 2016), change of matter (Saraç, 2018), the matter (Küçükaydın, Mem & Karamustafaoğlu, 2014), the granular structure of matter (Kenan & Özmen, 2014), sound (Sözen & Bolat, 2016)), respiration and photosynthesis (Haslam &

Treagust, 1987), Earth's crust (Sontay & Karamustafaoğlu, 2017). However, it is seen that there are limited tests to measure the achievement of secondary school students in biology subjects in science education. However, biology is accepted as the basic science that examines living and natural sciences (Kiziroğlu, 1988), and "cell" is one of the most difficult subject to understand by the students as it contains intangible concepts such as cell membrane, cell organelles and functions. A good learning of the cell subject will lead to a better understanding of the advanced biological subjects by the students and also it will lead to students to gain a positive attitude towards this course. For this reason, in this study the multiple choice test was developed to measure the academic achievement of 6th grade students regarding the "Cell" subject of the "Systems in Our Body" unit. It is thought that this test will be beneficial for teachers or researchers who want to measure the success of 6th grade students in "Cell" subject.

Method

Survey design was used in this study. The survey design is one of the quantitative research methods. In the literature, survey design studies are stated as the studies which allow us collect data in order to determine the characteristics of the research group and allow to make verbal or numerical descriptions about the characteristics such as opinions, beliefs or talents about the universe through these data.

Participants

For the pilot study of CAT, the sample group consisted of 62 students (3 branches, 34 girls, 28 boys) randomly selected from the 6th grade of a large secondary school in Ankara in 2017-2018 academic year. The sample group of the study consisted of 138 students in the 6th grade in the same secondary school except the branches included in the pilot study. The males are 64 and the females are 74 of the 138 students.

The Development Process of Cell Achievement Test (CAT)

In the first stage of the development of CAT, a table of objectives was prepared by considering the 6th grade cell subject's objectives. CAT with 25 multiple choice item was developed considering the targets and behaviors in this table. Questions in the test were examined by 2 instructors and 1 language expert for competence and comprehensibility in terms of language and in accordance with the achievements to be measured.

In the second stage of the study, a pilot study was conducted with 62 students in 6th grade of secondary school to test the applicability of the test. As a result of the analyzes performed after the pilot application, the difficulty level and discrimination of each item in the test were calculated and it was decided to remove 3 items from the test.

At the last stage, the test was applied to 138 students with 22 items in which 3 items removed from the test. The new 22-item test was re-tested for reliability and it was seen that the test was applicable.

TAP (Test Analysis Program) program was used to calculate the reliability, discrimination and difficulty indices of the CAT in both pilot and main study. Detailed data regarding these analyzes are presented in the findings section.

Findings

In the first stage of the study, the table of objectives shown in Table 1 was prepared by considering the objectives related to the subject of "Cell" in the science curriculum for the development of CAT.

Table 1.*Objective table prepared for CAT*

Cell Subjects	Objectives	Item No
1.The basic structure of living organisms: Cell	1.1. Explains the cell	Q2, Q9, Q16, Q3, Q4
2. Parts of the cell	2.1. Explains the basic parts of the cell	Q1, Q2, Q8, Q10, Q25, Q3, Q7, Q15, Q17, Q22, Q4, Q9, Q16, Q11, Q14
3. Similarities and differences between plant and animal cells	3.1. Compares animal and plant cells in terms of basic parts. 3.2. Compares animal and plant cells in terms of tasks	Q23, Q6, Q20, Q12, Q13, Q14
4. Developing technology and cell	4.1. Discusses the ideas about the structure of the cell from past to present by associating with technological developments.	Q5, Q17
5. Cell-tissue-organ-system-organism relationship in multicellular organisms	5.1. Explains the relationship between cell-tissue-organ-system-organism.	Q24, Q19, Q6, Q18, Q21

Before the pilot study, CAT consisting of 25 multiple-choice (4-choice) questions was developed in line with the objectives indicated in Table 1 using various sources by the researcher. The CAT was developed as multiple choice questions with 4 choice since the questions asked in the central exams were multiple choice (test) and with 4 choice. In order to determine the relationship and comprehensibility of the items in this test with the scientific knowledge, opinions of 2 instructors specialized in science education were taken. In addition, the accuracy of the test in terms of grammar was determined by consulting a linguist.

The CAT became ready for pilot study after various arrangements. The results of the analysis obtained from the pilot study conducted with 62 students are shown in Table 2.

Table 2.*Some statistical data for item analysis of CAT (out of 25 items)*

Participants	62	Number of questions	25	KR-20	0.817
Standard deviation	4.86	Skewness	-0.352	Kurtosis	-0.929
Mean item difficulty	0.674	Mean discrimination Index			0.477

When Table 2 was examined, it was seen that the mean item difficulty of CAT was 0.674 and the mean discrimination index was 0.477. However, Skewness and Kurtosis values of CAT were -0.352 and -0.929, respectively. These values indicated that the data exhibited normal distribution, i.e. a homogeneous scattering. The Skewness data indicated the degree of symmetrical distribution, i.e. the homogeneity, while the Kurtosis value indicated how flat or steep the CAT was in relation to the normal distribution curve. Accordingly, a Skewness value of <0 indicated a negative skewness to the left. The fact that the average was smaller than the media supports this statement. It could be said here that more than half of the scores were collected above the arithmetic mean. In addition, it could be concluded that the CAT questions were relatively easy for the sample group and the success was high. Therefore, CAT could be qualified as a test that accomplished the program and teaching objectives.

The prepared CAT was applied to a sample group which consists of 62 students who were 6th grade and based on the responses, the degree of difficulty and discrimination of each item in the test was calculated by the TAP (Test Analysis Program) program. The data obtained from the reliability analysis were shown in Table 3.

Table 3.

Item analysis of the test according to the answers of the students in the upper and lower groups (out of 25 items)

Item No	N	P	R _{ijx}	Upper Group Correct Answer	Lower Group Correct Answer	Result
1	38	0.61	0.60	15 (0.83)	4 (0.24)	Very good
2	35	0.56	0.59	16 (0.89)	5 (0.29)	Very good
3	44	0.71	0.53	17 (0.94)	7 (0.41)	Very good
4	44	0.71	0.42	16 (0.89)	8 (0.47)	Very good
5	53	0.85	0.35	18(1.00)	11 (0.65)	Good
6	24	0.39	0.43	13 (0.72)	5 (0.29)	Very good
7	28	0.45	0.37	12 (0.67)	5 (0.29)	Good
8	27	0.44	0.38	11 (0.61)	4 (0.24)	Good
9	45	0.73	0.65	17 (0.94)	5 (0.29)	Very good
10	33	0.53	0.42	15 (0.83)	7 (0.41)	Very good
11	45	0.73	0.65	17 (0.94)	5 (0.29)	Very good
12	46	0.74	0.53	17 (0.94)	7 (0.41)	Very good
13	48	0.77	0.36	17 (0.94)	10 (0.59)	Good
14	38	0.61	0.60	15 (0.83)	4 (0.24)	Very good
15	42	0.68	0.42	17 (0.94)	9 (0.53)	Very good
16#	53	0.85	0.29	18 (1.00)	12 (0.71)	Weak
17#	51	0.82	0.12	16 (0.89)	13 (0.76)	Weak
18	41	0.66	0.54	16 (0.89)	6 (0.35)	Very good
19	48	0.77	0.59	18 (1.00)	7 (0.41)	Very good
20	43	0.69	0.53	18 (1.00)	8 (0.47)	Very good
21	38	0.61	0.71	17 (0.94)	4 (0.24)	Very good
22	38	0.61	0.48	16 (0.89)	7 (0.41)	Very good
23#	50	0.81	0.24	16 (0.89)	11 (0.65)	Weak
24	42	0.68	0.53	17 (0.94)	7 (0.41)	Very good
25	50	0.81	0.59	18 (1.00)	7 (0.41)	Very good

p: Item difficulty r_{ijx}: Item discrimination; N: Total number of students who answered the question correctly

When Table 3 was examined, the number of correct answers given in the lower and upper groups for each item in CAT showed the index of discrimination and difficulty index of the test. When these values were examined, it was concluded that some items in the test should be removed from the test. For example, the difficulty index of item 16 in the test was 0.85 and the index of discrimination was 0.29. These values indicated that the item was easy and did not discriminate. For these reasons, it was decided to exclude item 16 from the test. Similarly, it was found that the item difficulty index value of item 17 was 0.82 and the discrimination index was 0.12. Since the discriminative property was smaller than 0.30, it was deemed appropriate to remove item 17 from the test. It was decided to remove item 23 because of the low discrimination of item and the close number of students who answered correctly in the lower group upper group. In summary, items 16, 17 and 23 were excluded from the test, but the removal of these items from the test did not impair the scope validity of the test. As a result, after 3 items were

removed from the test, reliability analysis was performed again on the new 22-item test and the findings obtained were presented in Table 4.

Table 4.
Some statistical results for item analysis of CAT (out of 22 items)

Participants	62	Number of questions	22	KR-20	0.818
Standard deviation	4.62	Skewness	-0.284	Kurtosis	-1.137
Mean item difficulty	0.652	Mean discrimination index			0.507

When Table 4 was examined, it was seen that there was not much change in the reliability values when 16, 17 and 23 items removed from the CAT, but the mean item difficulty decreased from 0.674 to 0.652, and the mean discrimination index increased from 0.477 to 0.507. Decrease in item difficulty indicated that the CAT became a little bit more difficult with the removal of these items, but the increase in the mean discrimination index indicated that the students in the sample group could be separated better.

As a result of the detailed analyzes given above, CAT was organized as 22 items and the page layout was finalized. The final version of the test was presented in Appendix-1.

Conclusion and Discussion

In this study, it is aimed to develop a “Cell Achievement Test (CAT)” which is compatible with the objectives the “Cell” subject of the “Systems in Our Body” unit in Science curriculum.

While developing a test, firstly the purpose and scope of the test are determined, then the items are written and arranged, and then the items are analyzed by pre-application and finally the test is created and made ready for implementation (Güler, 2012; Özçelik, 2011). In this study, these steps were followed while developing CAT. When the relevant literature is examined, it is seen that many tests have been developed through this way by the researchers (Akbulut & Çepni, 2013; Çakır & Aldemir, 2011; Fidan, 2013; Gönen et al., 2011; Kingır, 2011; Kızılcık & Tan, 2011; Şen & Eryılmaz, 2011; Tosun & Taşkesenligil, 2011). Firstly, CAT, which was prepared as 25 questions, was subjected to item analysis after the pilot study and decreased to 22 items. As the other studies in the literature, in this study, expert opinion was used to test the scope validity of the test (Ağgöl Yalçın & Bayrakçeken, 2010; Aksoy, 2010; Çalık & Ayas, 2003; Fidan, 2013; Güler, 2012; Kingır, 2011; Özçelik, 2011; Tosun & Taşkesenligil, 2011). The mean item difficulty and discriminant values of CAT obtained in this study were found to be 0.674 and 0.477, respectively. It was stated in the literature that the average item difficulty index of a test should be 0.50 (Atılğan, 2009; Gömleksiz & Erkan, 2010; Tekin, 2010). Accordingly, the average item difficulty index of the achievement test applied to the students in the study was 0.674, indicating that the test was easy for the students. This can be explained by the fact that the students have high grade point averages. The fact that the item discrimination index value of CAT was found to be 0.477 shows that the test has a high item discrimination power, since a test's item discrimination index is 0.40 and above, which indicates that the test has a high item discrimination property (Tekin, 2010). Apart from these features for an improved test, an important element that should be included in the test is the reliability of the scores obtained from that test. One of the most preferred methods to determine the reliability of scores obtained from a test is the Cronbach Alpha reliability coefficient (Şimşek, 2007; Taşdemir, 2004; Yalçın, 2012). In this study, the reliability value for CAT results applied to 6th grade students was found as 80. The scores obtained from a test should be between 0.60-0.90 in order to be considered reliable (Can, 2014). Accordingly, it can be said that CAT is a highly reliable measurement tool. When all these features that are required in a test are examined for CAT, it can be said that the values obtained are meaningful and therefore, it can be said that a valid and reliable achievement test has been developed which can contribute to the related literature including the objectives on Cell at 6th grade level.

Although this study provides contribution to the literature as a validity and reliability study of an achievement test on the subject of cell in the 6th grade of secondary school, the test consists of multiple choice questions and it is limited in determining students' learning about the cell. Therefore; it can be

suggested that this test ought to be developed in two stages for the future studies. Thus, the test whose first stage consists of a multiple-choice form and second stage is formed as open-ended question, will allow researchers to examine students' learning about cell in detail. In future studies, considering the use of the achievement test that developed in this study, the effectiveness of this test and its applicability can be determined and thus the test developed in this study can be revised.

Türkçe Sürümü

Giriş

Eğitim ve öğretim sisteminde öğrenciye “Ne kadar öğrettik?” sorusunu ölçme ve değerlendirmearacılığıyla yanıt veririz (Temizkan & Sallabaş, 2011). Bu nedenle ölçme ve değerlendirme eğitim için kaçınılmaz bir unsurdur. Bir öğrencinin hedeflenen veya ulaşılmak istenen öğretim kazanımına ne derece erişip erişmediği ancak ölçme ve değerlendirme aracılığıyla tespit edilebilir. Başka bir ifadeyle okulda kazandırılmak istenen niteliklerin öğrenciye ne ölçüde kazandırılıp kazandırılmadığı bu şekilde belirlenir. (Atılğan, 2013). Baykul (2000) ise eğitimde ölçme ve değerlendirmenin önemini, eğitim programlarının ne derece dayanıklı olup olmadığını saptaması, öğretimde kullanılan yöntem ve tekniklerin ne derece etkili olduğunu belirlemesi, öğrencilerdeki öğrenme güçlüklerini saptaması ve en önemlisi öğrencilerin başarılarını ölçmesi, eğitimin istenen hedeflere ne derece ulaşip ulaşılmadığını görmemizi sağlaması, şeklinde belirtmiştir. Bu ifadeler, eğitim öğretim sürecinin en önemli ve tamamlayıcı unsurunun “ölçme ve değerlendirme” olduğunu göstermektedir.

Literatürde geleneksel ölçme ve değerlendirme araçları; açık uçlu yazılı yoklamalar, çoktan seçmeli sorular, doğru-yanlış soruları, boşluk doldurma soruları ve eşleştirme soruları olarak ifade edilmiştir (Karip, 2012). Yazılı yoklamalar, öğrencilere bir ya da birkaç sorunun yöneltildiği ve belirli bir düzende cevabın yazılı olarak belirtilmesinin istendiği ölçme araçlarıdır (Temizkan & Sallabaş, 2011). Ancak yazılı yoklamaların uygulanması ve puanlanması zaman alıcı olduğundan çok fazla tercih edilmemektedir. Bu nedenle uygulayıcılar bu tür sınavlarda soru sayısını sınırlı tutmaktadırlar. Bu durum da soruların evreni temsil edememesinden yazılı yoklamaların kapsam geçerliğini düşürmektedir (Tekin, 2010). Bir diğer ölçme değerlendirme aracı ise doğru-yanlış sorularıdır. Öğrencinin kendisine verilen cümleyi veya önermeyi “Doğru (D)” veya “Yanlış (Y)” olarak sınıflandırmasına dayanan sorulardır. Bu tekniğin avantajlarından biri soruların hazırlanmasının ve yanıtlanmasının kolay olmasıdır. Soru sayısının fazla olması bu testin kapsam geçerliğini de artırmaktadır. Ancak bu ölçme aracında en büyük dezavantaj öğrencinin doğru yanıtı bulma şansının yüksek (%50) olmasıdır. (Erkuş, 2006). Boşluk doldurma soruları ise metinden bir kelimenin çıkarılması ve çıkarılan bu kelimenin aynısını ya da eş anlamlısını öğrencilerin yazmasının istendiği bir ölçme aracıdır (Akyol, 2012). Hem grupla hem de bireysel olarak uygulanabilen boşluk doldurma soruları kullanışlı bir tekniktir. Ayrıca öğrencinin boşluğa yazacağı yanıtlar kısa olduğundan puanlaması kolay ve nesnelidir (Erkuş, 2006). Eşleştirme soruları, öğrenciye bir sütunda verilen kelimeleri diğer sütunda verilen kelimelerle bağlantılı olacak şekilde uygun eşlemeyi yapmasının istendiği sorulardır. Bu soruların hazırlanması ve puanlanması oldukça kolaydır. Bu ölçüm aracının en önemli özelliği hatırlama düzeyindeki davranışları test etmesidir (Erkuş, 2006). Eğitim sisteminde bu ölçme değerlendirme araçlarından en sık kullanılanlardan birisi çoktan seçmeli testlerdir. Çoktan seçmeli testler, çok soru içermesinden dolayı kapsam geçerliğini fazlasıyla sağlaması, sınav süresinin çok uzun olmaması ve uygulamaya elverişli olması bakımından çok fazla tercih edilmektedir. Bununla birlikte, çoktan seçmeli testlerin şimdiye kadar bulunmuş en üstün ölçme aracı olduğu ifade edilmiştir (Özçelik, 1998). Çoktan seçmeli testlerde doğru cevap testi hazırlayan tarafından cevap seçenekleri arasında verilir ve testi cevaplayanlar bu doğru cevabı bulmaya çalışırlar (Güler, 2017). Bu tanımlamadan da anlaşılacağı üzere çoktan seçmeli testler, testi cevaplayanın o konudaki düşüncelerini, yorumlarını, kavramlarını sınırlasa da eğitim sisteminde en yaygın kullanılan ölçme araçlarından biridir. Çünkü bu testler özellikle kalabalık gruplarda testi uygulayan kişi için zamandan kazanç sağlamaktadır. Kolay puanlanabilmesi de avantajlarından bir diğeridir (Birgili, 2014). Bununla birlikte bu testlerle puanlamanın güvenilirliğinin sağlanması da kolaydır (Klufa, 2015).

İlgili literatür incelendiğinde öğrencilerin başarılarını ölçmeye yönelik geliştirilmiş çoktan seçmeli testlerin oldukça fazla olduğu görülmektedir (Akbulut ve Çepni 2013; Altunkaya, 2016; Ayvaci ve Durmuş, 2016; Demir, Kızılay ve Bektaş, 2016; Fidan, 2013; Gönen, Kocakaya ve Kocakaya, 2011; Narlı ve Başer, 2008; Singh ve Rosengrant,2003; Saylan Kırmızıgül ve Kaya, 2019; Şahin ve Keşan, 2017). Fen eğitimiyle

ilgili olarak geliştirilen bu testlerden bazıları konu dağılımlarına göre şu şekildedir; asit ve bazlar (Demirci, 2011), kaldırma kuvveti ve yüzme (Akbulut, 2010), elektrik devreleri (Can, Şen ve Eryılmaz, 2011), çevre sorunları (Güven, 2013), çözeltiler (Çalık, 2006; Demir, Kızılay ve Bektaş, 2016), dinamik (Gönen, Kocakaya ve Kocakaya, 2011), genetik (Çakır ve Aldemir, 2011), güneş sistemi (Karamustafaoğlu ve Tutar, 2016), ısı ve sıcaklık (Ayvacı ve Durmuş, 2016), kuvvet ve hareket (Akbulut ve Çepni, 2013; İdin ve Aydoğdu, 2016), maddenin değişimi (Saraç, 2018), maddeyi tanıyalım (Küçükaydın, Mem ve Karamustafaoğlu, 2014), maddenin tanecikli yapısı (Kenan ve Özmen, 2014), ses (Sözen ve Bolat, 2016), solunum ve fotosentez (Haslam ve Treagust, 1987), yer kabuğu (Sontay ve Karamustafaoğlu, 2017). Ancak fen eğitiminde biyoloji konularına yönelik ortaokul öğrencilerin başarılarını ölçmeye yönelik testlerin sınırlı olduğu görülmektedir. Halbuki biyoloji canlı ve doğa bilimlerini inceleyen temel bilim olarak kabul edilmektedir (Kiziroğlu, 1988) ve biyoloji konuları içerisinde yer alan hücre konusu hücre zarı, hücrenin organelleri ve görevleri gibi soyut kavramlar içerdiğinden öğrenciler tarafından anlaşılması zor konuların başında gelmektedir. Hücre konusunun öğrenciler tarafından iyi öğrenilmesi daha ileri düzeydeki biyolojik konuların öğrenciler tarafından daha iyi anlaşılmasına ve derse karşı öğrencilerin olumlu tutum kazanmasına neden olacaktır. Bu nedenle bu çalışmada geliştirilen çoktan seçmeli test, ortaokul 6. sınıf öğrencilerinin “Vücudumuzdaki Sistemler” ünitesine ait “Hücre” konusundaki akademik başarılarını ölçmeye yönelik geliştirilmiştir. Geliştirilen bu test ile Hücre konusunda 6. sınıf öğrencilerinin başarılarını ölçmek isteyen öğretmen ya da araştırmacılara yarar sağlanacağı düşünülmektedir.

Yöntem

Bu çalışmada tarama deseni kullanılmıştır. Tarama deseni nicel araştırma yöntemlerinden biridir. Literatürde, tarama araştırmaları üzerinde çalışılan grubun özelliklerini tespit etmek amacıyla veri toplamayı ve bu verilerden yola çıkarak evren hakkında görüş, inanış veya yetenek gibi özellikler hakkında sözel yada sayısal olarak betimlemeler yapmamıza olanak sağlayan araştırmalar olara ifade edilmiştir (Fraenkel, Wallen, & Hyun, 2012).

Katılımcılar

HBT'nin pilot çalışması için örneklem grubu 2017-2018 eğitim öğretim yılında Ankara ilinin büyük bir ortaokulunun 6. Sınıfında öğrenim gören ve rastgele seçilen 62 (3 şube, kız 34, erkek 28) öğrenciden oluşturulmuştur. Asıl çalışmanın örneklem grubunu ise yine aynı ortaokulun pilot çalışmaya dahil olan şubeler haricindeki 6. Sınıflarında öğrenim gören 138 öğrenci oluşturmuştur. 138 kişilik öğrenci grubunun 64 tanesi erkek öğrenci, 74 tanesi kız öğrenciden oluşmaktadır.

Hücre Başarı Testinin Geliştirilmesi Süreci

HBT'nin geliştirilmesi sürecinde ilk olarak 6. sınıf hücre konusu kazanımları göz önünde bulundurularak belirtke tablosu hazırlanmıştır. Bu tabloda bulunan hedef ve davranışlar göz önünde bulundurularak 25 maddeden oluşan çoktan seçmeli hücre başarı testi oluşturulmuştur. Testte bulunan sorular; dil açısından yeterliği ve anlaşılabilirliği, ölçülmek istenen kazanımlarla uyumu açısından alanında uzman 2 öğretim elemanı ve 1 dil uzmanının görüşüne başvurulmuştur.

Çalışmanın ikinci aşamasında ise oluşturulan testin uygulanabilirliğini test etmek amacıyla ortaokulun 6. sınıfında öğrenim gören 62 öğrenciyle pilot çalışma yapılmıştır. Pilot uygulama sonrası yapılan analizler sonucunda testteki her bir maddenin güçlük derecesi ve ayırt ediciliği hesaplanmıştır ve testten 3 maddenin çıkartılmasına karar verilmiştir.

En son aşamada testten çıkarılan 3 maddeden sonra kalan 22 madde ile test 138 öğrenciye uygulanmıştır. 22 maddelik yeni test üzerinde tekrar güvenilirlik analizleri yapılmıştır ve testin uygulanabilir olduğu kanısına varılmıştır.

Hem pilot çalışmada hem de asıl çalışmada HBT testinin güvenilirliği, maddelerinin ayırt edicilik ve güçlük indekslerinin hesaplanmasında TAP (Test Analiz Programı) programı kullanılmıştır. Bu analizlere ilişkin detaylı veriler bulgular kısmında sunulmuştur.

Bulgular

Çalışmanın ilk aşamasında HBT'nin geliştirilmesi için Fen Bilimleri Dersi öğretim programında yer alan "Hücre" konusuna ilişkin kazanımlar dikkate alınarak aşağıdaki Tablo 1'de gösterilen belirtke tablosu hazırlanmıştır.

Tablo 1.
HBT İçin Hazırlanan Belirtke Tablosu

Hücre Konuları	Hücre Konusuna ait öğrenci Kazanımları	Soru No
1.Canlıların temel yapı birimi: Hücre	1.1. Hücreyi açıklar	S2, S9, S16, S3, S4
2.Hücresinin kısımları	2.1. Hücresinin temel kısımlarını açıklar	S1, S2, S8, S10, S25, S3, S7, S15, S17, S22, S4, S9, S16, S11, S14
3. Bitki ve hayvan hücreleri arasındaki benzerlik ve farklılıklar	3.1.Hayvan ve bitki hücrelerini temel kısımlar açısından karşılaştırır 3.2. Hayvan ve bitki hücrelerini görevleri kısımlar açısından karşılaştırır	S23, S6, S20, S12, S13, S14
4. Gelişen teknoloji ve hücre	4.1. Geçmişten günümüze, hücrenin yapısı ile ilgili olarak ileri sürülen görüşleri teknolojik gelişmelerle ilişkilendirerek tartışır.	S5, S17
5. Çok hücreli canlılarda hücre-doku-organ-sistem-organizma ilişkisi	5.1.Hücre-doku-organ-sistem-organizma ilişkisini açıklar.	S24, S19, S6, S18, S21

Tablo 1'de belirtilen kazanımlar doğrultusunda ve çeşitli kaynaklardan yararlanarak araştırmacı tarafından pilot çalışma öncesi çoktan seçmeli (4 seçenekli) 25 sorudan oluşan HBT geliştirilmiştir. Merkezi yazılılarda sorulan soruların çoktan seçmeli (test) ve 4 seçenekli olmasından dolayı HBT testi 4 seçenekli ve çoktan seçmeli sorulardan oluşturulmuştur. Elde edilen bu testteki maddelerin bilimsel bilgilerle ilişkisini ve anlaşılabilirliğini belirlemek amacıyla fen bilimleri eğitimi alanında uzman 2 öğretim elemanının görüşleri alınmıştır. Ayrıca 1 dil uzmanı görüşüne başvurularak testin dil bilgisi açısından doğruluğu tespit edilmiştir. Çeşitli düzenlemelerden sonra HBT pilot uygulama için hazır hale gelmiştir. 62 öğrenci ile gerçekleştirilen pilot çalışmadan elde edilen verilerin analiz sonuçları Tablo 2'de gösterilmiştir.

Tablo 2.
HBT'nin Madde Analizine Yönelik Bazı İstatistiksel Veriler (25 Madde Üzerinden)

Kişi	62	Soru Sayısı	25	KR-20	0.817
Standart sapma	4.86	Skewness	-0.352	Kurtosis	-0.929
Ortalama madde güçlüğü	0.674	Ortalama ayırt edicilik indeksi			0.477

Tablo 2 incelendiğinde HBT'nin ortalama madde güçlüğü 0.674, ortalama ayırt edicilik indeksinin ise 0.477 olarak bulgulandığı görülmektedir. Bununla birlikte HBT'nin Skewness (çarpıklık) ve Kurtosis (basıklık) değerlerinin sırasıyla -0,352 ve -0,929 olduğu saptanmıştır. Bu değerler verilerin normal dağılım sergilediğini yani homojen bir saçılma yaptığını gösterir. Skewness veri dağılımının ne derece simetrik olduğunu yani homojenliğini gösterirken Kurtosis değeri HBT'nin normal dağılım eğrisine ilişkin ne derece basık ya da dik olduğunu gösterir. Buna göre Skewness değerinin <0 olması sola yani negatif bir çarpıklığı ifade eder. Ortalamanın medyandan küçük olması bu ifadeyi destekler niteliktedir. Buradan puanların yarıdan fazlasının aritmetik ortalamanın üstünde toplandığı söylenebilir. Ayrıca hazırlanan HBT test

sorularının örneklem grubuna nispeten kolay geldiği ve başarının yüksek olduğu sonucu çıkarılabilir. Dolayısıyla HBT testi program ve öğretim hedeflerini gerçekleştiren bir test niteliğindedir.

Hazırlanan HBT testi 6. sınıf öğrencilerinden oluşan 62 kişilik örneklem grubuna uygulanmış ve cevaplara dayanarak testteki her bir maddenin güçlük derecesi ve ayırt ediciliği TAP (Test Analiz Programı) programı ile hesaplanmıştır. Güvenilirlik analizi sonucunda elde edilen veriler Tablo 3’de gösterilmiştir.

Tablo 3.
Üst ve Alt Gruptaki Öğrencilerin Yanıtlarına Göre Testin Madde Analizi (25 Madde Üzerinden)

Soru No	N	P	r _{ij}	Dü	Da	Sonuç
1	38	0.61	0.60	15 (0.83)	4 (0.24)	çok iyi
2	35	0.56	0.59	16 (0.89)	5 (0.29)	çok iyi
3	44	0.71	0.53	17 (0.94)	7 (0.41)	çok iyi
4	44	0.71	0.42	16 (0.89)	8 (0.47)	çok iyi
5	53	0.85	0.35	18(1.00)	11 (0.65)	iyi
6	24	0.39	0.43	13 (0.72)	5 (0.29)	çok iyi
7	28	0.45	0.37	12 (0.67)	5 (0.29)	iyi
8	27	0.44	0.38	11 (0.61)	4 (0.24)	iyi
9	45	0.73	0.65	17 (0.94)	5 (0.29)	çok iyi
10	33	0.53	0.42	15 (0.83)	7 (0.41)	çok iyi
11	45	0.73	0.65	17 (0.94)	5 (0.29)	çok iyi
12	46	0.74	0.53	17 (0.94)	7 (0.41)	çok iyi
13	48	0.77	0.36	17 (0.94)	10 (0.59)	iyi
14	38	0.61	0.60	15 (0.83)	4 (0.24)	çok iyi
15	42	0.68	0.42	17 (0.94)	9 (0.53)	çok iyi
16#	53	0.85	0.29	18 (1.00)	12 (0.71)	Zayıf
17#	51	0.82	0.12	16 (0.89)	13 (0.76)	Zayıf
18	41	0.66	0.54	16 (0.89)	6 (0.35)	çok iyi
19	48	0.77	0.59	18 (1.00)	7 (0.41)	çok iyi
20	43	0.69	0.53	18 (1.00)	8 (0.47)	çok iyi
21	38	0.61	0.71	17 (0.94)	4 (0.24)	çok iyi
22	38	0.61	0.48	16 (0.89)	7 (0.41)	çok iyi
23#	50	0.81	0.24	16 (0.89)	11 (0.65)	Zayıf
24	42	0.68	0.53	17 (0.94)	7 (0.41)	çok iyi
25	50	0.81	0.59	18 (1.00)	7 (0.41)	çok iyi

Dü: Soruya doğru cevap veren üst gruptaki öğrenci sayısı; Da: Soruya doğru cevap veren alt gruptaki öğrenci sayısı; p: Madde güçlüğü r: Madde ayırt ediciliği; N: Soruya doğru cevap veren toplam öğrenci sayısı

Tablo 3’ e bakıldığında, HBT’ndeki her bir maddeye ilişkin alt ve üst grupta verilen doğru cevap sayısı testin ayırt edicilik indeksini ve güçlük indekslerini göstermektedir. Bu değerler incelendiğinde testteki bazı maddelerin testten çıkartılmasının doğru olacağı kanısına varılmıştır. Örneğin, testteki 16. Maddenin güçlük indeksinin 0.85, ayırt edicilik indeksinin ise 0.29 olduğu görülmektedir. Bu değerler maddenin kolay olduğunu, ayırt edicilik özelliği taşımadığını göstermektedir. Bu sebeplerden ötürü 16. Maddenin testten çıkartılmasına karar verilmiştir. Aynı şekilde 17 maddenin de madde güçlük indeksi değerinin 0.82, ayırt edicilik indeksinin 0.12 olduğu tespit edilmiştir. Ayırt edicilik özelliğinin 0.30 da küçük olması sebebiyle 17. Maddenin de testten çıkartılması uygun görülmüştür. HBT’ de 23. maddenin de ayırt ediciliğinin düşük olması ve alt grup üst grupta doğru cevap veren öğrenci sayısının birbirine çok yakın olması nedeniyle bu maddenin de testten çıkartılmasına karar verilmiştir. Özetle testten 16,17 ve 23. Maddeler çıkartılmıştır ancak testten bu maddelerin çıkartılması testin kapsam geçerliğini bozmamıştır. Sonuç itibarıyla testten 3 madde çıkartıldıktan sonra elde edilen 22 maddelik yeni test üzerinde tekrar güvenilirlik analizleri yapılmış ve elde edilen bulgular Tablo 4’de sunulmuştur.

Tablo 4.
HBT'nin Madde Analizine Yönelik Bazı İstatistiksel Sonuçlar (22 Madde Üzerinden)

Kişi	62	Soru Sayısı	22	KR-20	0.818
Standart sapma	4.62	Skewness	-0.284	Kurtosis	-1.137
Ortalama madde güçlüğü	0.652	Ortalama ayırt edicilik indeksi			0.507

Tablo 4 incelendiğinde HBT testinden 16, 17 ve 23 maddelerin çıkarılmasıyla güvenilirlik değerlerinde çok fazla bir değişim görülmemekle birlikte ortalama madde güçlüğü 0,674 den 0,652 düştüğü, ayrıca ortalama ayırt edicilik indeksinin ise 0,477 den 0,507 e yükseldiği görülmektedir. Madde güçlüğü azalması HBT testinin söz konusu maddelerin çıkarılmasıyla biraz daha zorlaştığını, bununla birlikte ortalama madde ayırt ediciliğinin yükselmesi ise örneklem grubunda bulunan öğrencilerin daha iyi ayrılabilirliğini gösterir.

Yukarıda verilen detaylı analizler sonucu HBT 22 madde ve seçenekleri dikkate alınarak sıralanmış ve sayfa düzenlemesi yapılarak son hali verilmiştir. Testin son hali Ek-1'de sunulmuştur.

Sonuç ve Tartışma

Bu çalışmada Fen Bilimleri dersi öğretim programında yer alan “Vücudumuzdaki Sistemler” ünitesine ait “Hücre” konusunun kazanımları ile uyumlu, geçerlik ve güvenilirlik çalışması yapılmış bir “Hücre Başarı Testi (HBT)”nin geliştirilmesi amaçlanmıştır.

Bir test geliştirilirken öncelikle testin amacı ve kapsamı belirlenir sonra maddeler yazılır ve düzenlenir daha sonra ön uygulama yapılarak maddelerin analizi yapılır ve son olarak test oluşturularak uygulanmaya hazır hale getirilir (Güler, 2012; Özçelik, 2011). Bu çalışmada HBT geliştirilirken de bu basamaklar izlenmiştir. İlgili literatür incelendiğinde araştırmacılar tarafından bu şekilde bir çok testin geliştirildiği görülmektedir (Akbulut ve Çepni, 2013; Çakır ve Aldemir, 2011; Fidan, 2013; Gönen ve diğer., 2011; Kingir, 2011; Kızılıcak ve Tan, 2011; Şen ve Eryılmaz, 2011; Tosun ve Taşkesenligil, 2011). İlk olarak 25 soru şeklinde hazırlanan HBT, pilot çalışma sonrası madde analizine tabi tutulmuş ve 22 maddeye inmiştir. Literatürdeki diğer çalışmalarda olduğu gibi bu çalışmada da testin kapsam geçerliğinin sınanması için uzman görüşüne başvurulmuştur (Ağgöl Yalçın ve Bayrakçeken, 2010; Aksoy, 2010; Çalık ve Ayas, 2003; Fidan, 2013; Güler, 2012; Kingir, 2011; Özçelik, 2011; Tosun ve Taşkesenligil, 2011). Bu çalışmada elde edilen HBT'nin ortalama madde güçlüğü ve ayırt edicilik değerleri sırasıyla 0.674 ve 0.477 olarak bulgulanmıştır. Literatürde bir testin ortalama madde güçlük indeksinin 0.50 düzeyinde olması gerektiği belirtilmiştir (Atılğan, 2009; Gömleksiz ve Erkan, 2010; Tekin, 2010). Buna göre araştırmada öğrencilere uygulanan başarı testinin ortalama madde güçlük indeksinin 0.674 çıkması, testin öğrencilere kolay geldiğini göstermektedir. Bu durum öğrencilerin yüksek not ortalamalarına sahip olmalarından dolayı kaynaklandığı şeklinde açıklanabilir. HBT'nin madde ayırt edicilik indeksi değerinin 0.477 olarak bulgulanması testin madde ayırt etme gücünün yüksek olduğunu göstermektedir çünkü bir testin madde ayırt edicilik indeksinin 0.40 ve üzerinde olması o testin madde ayırt edicilik özelliğinin yüksek olduğunu gösterir (Tekin, 2010). Geliştirilen bir test için bahsedilen bu özellikler dışında testte bulunması gereken önemli bir unsur da o testten elde edilen puanların güvenilirliğidir. Bir testten elde edilen puanların güvenilirliğini tespit etmeye yönelik en çok tercih edilen yöntemlerden biri Cronbach Alpha güvenilirlik katsayısıdır (Şimşek, 2007; Taşdemir, 2004; Yalçın, 2012). Bu çalışmada 6. sınıf öğrencilerine uygulanan HBT sonuçları için güvenilirlik değeri 0.80 olarak bulgulanmıştır. Bir testten alınan puanların güvenilir kabul edilebilir olması için 0.60-0.90 arasında değer alması gerekmektedir (Can, 2014). Buna göre HBT'nin güvenilirliği yüksek bir ölçme aracı olduğu söylenebilir. Bir testte olması gereken tüm bu özellikler HBT için incelendiğinde elde edilen değerlerin anlamlı olduğu ve bu nedenle 6. Sınıf seviyesinde Hücre konusundaki kazanımları içeren ilgili literatüre katkı sağlayabilecek geçerli ve güvenilir bir başarı testi geliştirildiği söylenebilir.

Bu çalışma ortaokul 6. sınıf Hücre konusunda geçerliği ve güvenilirliği saptanmış bir başarı testi olarak literatüre katkı sağlasa da, test çoktan seçmeli sorulardan oluştuğundan öğrencilerin hücre konusundaki

kavramlarını derinlemesine belirlemede sınırlı kalmaktadır. Bu nedenle ileride yapılacak çalışmalara bu testin 2 aşamalı olarak geliştirilmesi önerilebilir. Böylece ilk aşaması çoktan seçmeli, ikinci aşaması açık uçlu sorudan oluşacak test, araştırmacılara öğrencilerin hücre konusuna ilişkin kavramlarını detaylı olarak inceleme fırsatı sunacaktır. Çalışmada geliştirilen başarı testinin alanyazında kullanılması sonucu testin ölçme gücü ve uygulanabilirliği tespit edilebilir ve böylece ileride yapılacak çalışmalarda geliştirilen test revize edilebilir.

References

- Ağgül Yalçın, F. & Bayrakçeken, S. (2010). 5E öğrenme modelinin fen bilgisi öğretmen adaylarının asit-baz konusu başarılarına etkisi. *International Online Journal of Educational Sciences*, 2(2), 508-531.
- Akbulut, H.H., (2010). *Sıvıların kaldırma kuvveti ve yüzme kavramlarına yönelik probleme dayalı öğrenme uygulaması ve değerlendirilmesi*, Yayınlanmamış Yüksek Lisans Tezi, KTU Fen Bilimleri Enstitüsü, Trabzon.
- Akbulut, H. İ. & Çepni, S. (2013). Bir üniteye yönelik başarı testi nasıl geliştirilir? İlköğretim 7. sınıf kuvvet ve hareket ünitesi. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 2(1), 18-44.
- Aksoy, M. (2010). *Ortaöğretim kimya dersindeki çözünürlük konusunun kavram haritaları ile öğretilmesinin öğrencilerin başarı ve tutumlarına etkisi*. Yüksek Lisans Tezi, Selçuk Üniversitesi, Konya.
- Atılğan, H. (Ed.) (2009). *Eğitimde ölçme ve değerlendirme* (4. Baskı). Ankara: Anı Yayıncılık.
- Atılğan, H. (Ed.) (2013). *Eğitimde ölçme ve değerlendirme* (6. Baskı). Ankara: Anı Yayıncılık.
- Ayvacı, H. Ş. & Durmuş, A. (2016). Bir başarı testi geliştirme çalışması: Isı ve sıcaklık başarı testi geçerlik ve güvenilirlik araştırması. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 35 (1), 87-103.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması [Measurement in education and psychology: Classical test theory and application]*. Ankara: Öğrenci Seçme ve Yerleştirme Merkezi Yayınları.
- Çakır, M. & Aldemir B. (2011). İki aşamalı genetik kavramlar tanı testi geliştirme ve geçerlik çalışması. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 8(16), 335-353.
- Çalık, M. (2006). *Bütünleştirici öğrenme kuramına göre lise 1 çözeltiler konusunda materyal geliştirilmesi ve uygulanması*. Yayınlanmamış doktora tezi. Karadeniz Teknik Üniversitesi, Trabzon
- Çalık, M. & Ayas, A. (2003). Çözeltilerde kavram başarı testi hazırlama ve uygulama. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 2 (14), 1-17.
- Demir, N., Kızılay, E. & Bektaş, O. (2016). 7. Sınıf çözeltiler konusunda başarı testi geliştirme: geçerlik ve güvenilirlik çalışması. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi (EFMED)*, 10 (1), 209-237.
- Erkuş, A. (2006). *Sınıf öğretmenleri için ölçme ve değerlendirme kavramlar ve uygulamalar*, Ankara: Ekinoks Yayınları.
- Fidan, E. (2013). *İlkokul öğrencileri için matematik dersi sayılar öğrenme alanında başarı testi geliştirilmesi*. Yayınlanmamış Yüksek lisans tezi, Ankara Üniversitesi.
- Gömlüksiz, M.& Erkan, S. (2010). *Eğitimde ölçme ve değerlendirme* (2. Baskı). Ankara: Nobel Yayın Dağıtım.
- Gönen, S., Kocakaya, S. & Kocakaya, F. (2011). Dinamik konusunda geçerliliği ve güvenilirliği sağlanmış bir başarı testi geliştirme çalışması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 8 (1), 40-57.
- Güler, N. (2017). *Eğitimde ölçme ve değerlendirme* (4. Baskı). Ankara: Pegem A Yayıncılık.
- Güler, N. (2012). *Eğitimde ölçme ve değerlendirme* (2. Baskı). Ankara: Pegem A Yayıncılık.
- Güven, E. (2013). Çevre sorunları başarı testinin geliştirilmesi ve öğretmen adaylarının bilgi düzeylerinin belirlenmesi. *Trakya Üniversitesi Eğitim Fakültesi Dergisi*, 3(2), 114-127.
- İdin, Ş. & Aydoğdu, C. (2016). Kuvvet ve hareket ünitesi başarı testi geçerlik ve güvenilirlik araştırması. *Eğitim, bilim ve teknoloji araştırmaları Dergisi*, 1(1), 14-33.
- Karip, E. (Ed.) (2012). *Ölçme ve değerlendirme*. Ankara: Pegem Akademi (5. Baskı).
- Karamustafaoğlu, S. & Tutar, M. (2016). 6. Sınıf dünyamız, ay ve yaşam kaynağımız güneş ünitesine yönelik bir başarı testi geliştirme. *Eğitim Bilimlerinde Yenilikler ve Nitelik Arayışı*, (303- 320).Ö. Demirel ve S. Dinçer (Ed.), Pegem Akademi Yayınları, Ankara.
- Kingır, S. (2011). *Using the science writing heuristic approach to promote student understanding in chemical changes and mixtures*. Doktora tezi, ODTÜ, Ankara.

- Kızılcık, H. Ş. & Tan, M. (2011). İtme ve momentum konusunda çoktan seçmeli bir test geliştirilmesi. *Kastamonu Eğitim Dergisi*, 19(1), 185-198.
- Kızıroğlu, İ. (1988). Günümüzde biyoloji dersi ve amaçları. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 3, 243-250.
- Klufa, J. (2015). Multiple Choice Question Tests –Advantages and Disadvantages. Retrieved from <http://www.inase.org/library/2015/zakynthos/bypaper/EDU/EDU-07.pdf>
- Narlı, S. & Başer, N. (2008). "Küme, Bağntı, Fonksiyon" Konularında Bir Başarı Testi Geliştirme ve Bu başarı testi ile Üniversite Matematik Bölümü 1. Sınıf Öğrencilerinin Bu Konulardak Hazırbulunuşluklarını Betimleme Üzerine Nicel Bir Araştırma. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi*, 24, 147-158
- Özçelik, D.A. (1998). *Ölçme ve Değerlendirme*. Ankara: ÖSYM Yayınları.
- Özçelik, D. A. (2011). *Ölçme ve Değerlendirme* (4. Baskı).Ankara: Pegem Akademi.
- Saraç, H. (2018). Fen Bilimleri Dersi 'Maddenin Değişimi' Ünitesi ile İlgili Başarı Testi Geliştirme: Geçerlik ve Güvenirlik Çalışması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 18 (1), 416-445.
- Saylan Kırmızıgül, A. & Kaya, H. (2019). 'Aynalarda Yansıma ve Işığın Soğrulması' Konusunda Geçerliliği ve Güvenilirliği Sağlanmış Başarı Testi Geliştirme Çalışması. *Kuramsal Eğitimbilim Dergisi*, 12 (82), 474 – 493.
- Singh, C. ve Rosengrant, D. (2003). Multiple-Choice Test of Energy and Momentum Concepts. *American Journal of Physics*, 71(6), 607-617.
- Sontay, G. ve Karamustafaoğlu, S. (2017) 5. Sınıf Fen Bilimleri Dersi "Yer Kabuğunun Gizemi" Ünitesine Yönelik Başarı Testi Geliştirme. *Fen Bilimleri Öğretimi Dergisi*, 5(1), 62-86.
- Sözen, M., & Bolat, M. (2016). Developing an Achievement Test for the Subject of Sound in Science Education. *Journal of Education and Learning*, 5(2), 149.
- Şahin, Z. & Keşan, C. (2017). Bir başarı testi geliştirme çalışması: beşinci sınıf öğrencilerinin geometri kazanımlarını ölçmeye yönelik bir başarı testi geçerlik ve güvenirliliğin araştırılması. *International Journal of New Trends In Arts, Sports & Science Education (Ijtase)*, 6 (3), 47-57.
- Şen, H. C. & Eryılmaz, A. (2011). Bir başarı testi geliştirme çalışması: Basit elektrik devreleri başarı testi geçerlik ve güvenirlik araştırması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 8 (1), 1-39.
- Tekin, H. (2010). *Eğitimde ölçme ve değerlendirme* (20. Baskı). Ankara: Yargı Yayınevi.
- Temizkan, M. & Sallabaş, M. (2015). Okuduğunu anlama becerisinin değerlendirilmesinde çoktan seçmeli testlerle açık uçlu yazılı yoklamaların karşılaştırılması. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (30), 207-220.
- Tosun, C. & Taşkesenligil, Y. (2011). Revize edilmiş Bloom'un taksonomisine göre çözeltiler ve fiziksel özellikleri konusunda başarı testinin geliştirilmesi: Geçerlik ve güvenirlik çalışması. *Kastamonu Eğitim Dergisi*, 19(2), 499-522.

Appendix 1

HBT Testi



1-Yandaki şekilde verilen organelin ismi hangi seçenekte doğru gösterilmiştir?
a-Çekirdek b-Mitokondri c-Lizozom d-Koful

2-Hücredeki temel yaşamsal olayların (boşaltım, solunum vb) gerçekleştiği kısım aşağıdakilerden hangisidir?

a-Çekirdek b-Hücre çeperi c-Sitoplazma d-Hücre Zarı

3-Aşağıdakilerden hangisi hücre zarının görevleri arasında yer almaz?

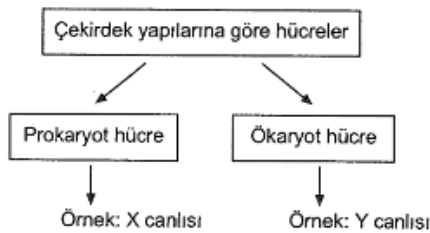
a-Esnek ve seçici geçirgendir
b-Hücreyi dış etkilerden korur
c-Hücreye şekil verir.
d-Sert bir yapısı vardır

4- I. Çoğunluğu sudan oluşan, şeffaf ve akıcı bir sıvıdır.
II. Yaşamsal faaliyetleri gerçekleştiren organelleri bulundurur.
III. Hücreye şekil verir.
Yukarıda verilen özelliklerden hangileri hücre sitoplazmasına aittir?

a-Yalnız I b-Yalnız II c-I ve II d-I, II ve III

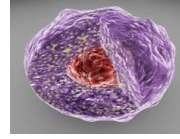
5-Aşağıda bilim insanlarının hücreyle ilgili yaptıkları çalışmalar verilmiştir. Buna göre verilen eşleştirmelerden hangisi yanlıştır.
a-1668 yılı Robert Hooke-ölü bir mantar hücresinin hücre duvarını görmeyi başardı
b-1674 yılı Antonie Van Leeuwenhoek-canlı hücre gözlemlemeyi başardı
c-1898 yılı Camilio Golgi-Mitokondri organelini gözlemledi
d-1833 yılı Brown-Orkide hücrelerinde orkide çekirdeğini gözlemledi

6-



Yukarıdaki şemada çekirdek yapılarına göre hücreler gruplandırılmıştır. Buna göre X ve Y hücreleri hangi seçenekte belirtilen canlılara ait olamaz?

a-Bakteri-Baykuş
b-Şapkalı Mantar-Yeşil Alg
c-Mavi Alg-Çilek
d-Bakteri-Domates



7-Bir deri hücresinin çekirdeği çıkartıldığında, hücre bir süre sonra ölüyor. Aşağıdakilerden hangisi bu durumun sebeplerinden biri değildir?

a-Çekirdeğin hücreyi yönetimden sorumlu olması
b-Hücrenin metabolizmasının durması
c-Hücrenin çoğalamaması
d-Çekirdeğin hücrenin onarımı ile ilgili bilgileri bulundurması

8-Yukarıda şekli verilen organelin ismi hangi seçenekte doğru verilmiştir?

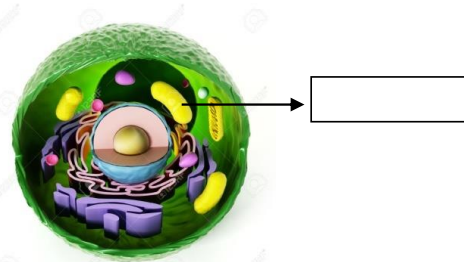
a-Lizozom b-Golgi c-Ribozom d-Çekirdek

9-Hücre zarından, küçük moleküller büyük moleküllere göre, yağda çözünen maddeler çözünmeyenlere göre daha kolay geçer. Yukarıda verilen ifade hücre zarının hangi özelliği ile ilgilidir?

a-Canlı olması
b-Esnek olması
c-Seçici geçirgen olması,
d-Saydam yapıda olması



10-Yukarıda şekli verilen organelin görevi hangi seçenekte doğru verilmiştir?
a-Enerji üretmek
b-Sindirim yapmak
c-Salgi üretmek
d-Bölünme yapmak



11-Yukarıdaki şekilde boş kutucuğa hücrenin kısımlarından hangisi gelmelidir?

a-Çekirdek
b-Kloroplast
c-Lizozom
d-Hücre zarı

Hücresel yapı	Bitki hücresi	Hayvan hücresi
K	+	+
L	+	-
M	-	+
N	+	+

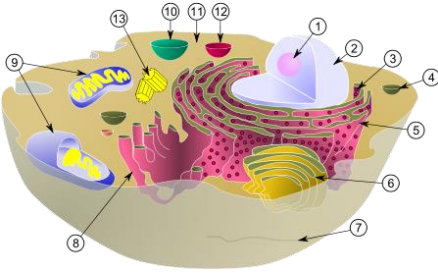
12- Tabloda bitki ve hayvan hücresinde K,L,M ve N harfleri ile gösterilen hücresel yapılardan bulunanlar +, bulunmayanlar - işaretiyle gösterilmiştir. Buna göre aşağıdakilerden hangisi söylenemez?

- a-K, hücre zarıdır
b-L, hücre çeperidir
c-M, kloroplasttır
d-N, çekirdektir.



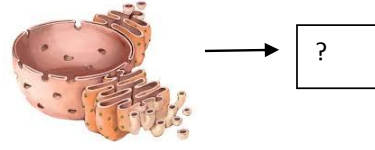
13- Yukarıda şekli gösterilen hücre modeli aşağıdaki canlılardan hangisine ait olabilir?

- a-Kedi b-Gül c-Lale d-Çimen



14) Yukarıda hayvan hücresine ait bir resim verilmiştir. 3 numaralı gösterilen bölüme hayvan hücresine ait organellerden hangisi gelmelidir?

- a-) Ribozom
b-) Mitokondri
c-) Hücre Zarı
d-) Hücre çekirdeği



15- Yukarıdaki şekilde okla gösterilen kısım hücrenin hangi organeline aittir?

- a- Çekirdek
b- Hücre Zarı
c- Koful
d- Endoplazmik retikulum

16- Aşağıdaki seçeneklerde verilen bilgilerden hangisi yanlıştır?

- a- Lizozom hücre içi sindirimde görevlidir
b- Ribozom protein sentezinde görevlidir
c- Golgi hücreye şekil verir.
d- Çekirdek hücrenin yönetim merkezidir.



17- Yukarıda şekilde gösterilen yapının ismi seçeneklerden hangisinde doğru verilmiştir?

- a- Mitokondri
b- Lizozom
c- Nükleotit
d- Hücre Zarı

18- Buldukları hücrelerin görevlerine göre, bazı organeller sayıca farklılık gösterebilir. Buna göre süt bezi, gözyaşı bezi ve tükürük bezi gibi yapılarda hangi organel sayıca diğerlerinden fazladır?

- a- Golgi cisimciği b- Lizozom c- Ribozom d- Sentrozom

19- Aşağıdaki organellerden hangisinin görevi yanlış verilmiştir?

- a- Endoplazmi Retikulum – Madde iletimi
b- Lizozom – Hücre Yönetimi
c- Mitokondri - Enerji Üretimi
d- Ribozom – Protein yapımı

20- Bir öğrenci kitapta gördüğü bir hücre resminin, bitki ya da hayvan hücresi olduğuna karar vermek istiyor. Buna göre öğrenci,

- I. Hücre duvarı
II. Mitokondri
III. Ribozom
IV. Kloroplast

yapılarının hücrede bulunma durumlarından hangilerine bakarak bu hücrenin ait olduğu canlı sınıfına karar verebilir?

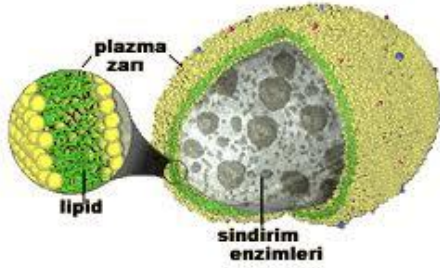
- a- I ve II. b- I ve IV. c- II ve III. d- III ve IV.

- 21- 1.Hücre
2.Organizma
3.Sistem
4.Doku
5.Organ

Yukarıda verilen kavramların basitten karmaşığa doğru sıralanması nasıl olmalıdır?

- a- 1 – 2 – 3 – 4 – 5 b- 1 – 4 – 5 – 3 – 2
c- 2 – 3 – 5 – 4 – 1 d- 4 – 1 – 2 – 5 – 3

22-



Yukarıdaki şekilde verilen organelin ismi aşağıdaki seçeneklerin hangisinde doğru verilmiştir?

- a-Lizozom
b-Mitokondri
c-Koful
d-Hücre zarı

23- Aşağıdaki organellerden hangisi hayvan hücresinde bulunmaz?

- a-Sentrozom b-Lizozom c-Koful d-Kromoplast

24- Bir sporcu, olimpiyatlara hazırlanmak için düzenli ve çok yoğun bir antrenman programı uygulamaktadır. Buna göre, bu sporcunun özellikle kas hücrelerinde bulunan organellerden hangilerinin diğerlerine göre çok daha fazla çalışması gerekmektedir?

- a-Koful b-Mitokondri c-Çekirdek d-Sentrozom

25-Hücre içerisinde protein üretmekle görevli organel hangisidir?

- a-Plastit b-Mitokondri c-Ribozom d-Lizozom