4.Sınıf Geometrik Şekiller Konusu Başarı Testi Geçerlik ve Güvenirlik Çalışması

Cumali ÖKSÜZ, Aydın Adnan Menderes Üniversitesi, ORCID ID: 0000-0002-3255-2542
Galip GENÇ, Aydın Adnan Menderes Üniversitesi, ORCID ID: 0000-0003-2447-4844

Öz

Bu çalışmada 4. sınıf matematik dersi geometrik şekiller konusu ile ilgili başarı testinin geliştirilmesinde geçerlik ve güvenirlik süreci açıklanmaya çalışılmıştır. Çalışma kapsamında, ilkokul dördüncü sınıf matematik ders programı (2018) incelenerek "Geometrik Şekiller" ile ilgili kazanım listesi çıkarılarak, belirtke tablosu hazırlanmış ve bu tabloya dayalı olarak, 21 çoktan seçmeli, 5 de açık uçlu denemelik maddeler geliştirilmiştir. Uzman görüşü alındıktan sonra, soru maddeleri zorluk düzeyi ve testin görünüş geçerliğine göre teste yerleştirilmiştir. Testin uygulanması için 40 dakikalık bir süre belirlenmiştir. Testin anlaşılabilirliğinin ortaya konulabilmesi bakımından konuyu görmüş olan 4. sınıftan 8 öğrenciye birebir uygulanarak, sesli şekilde soruları cevaplamaları istenmiş ve buna bağlı olarak anlaşılmayan soruların düzenlemeleri yapılmıştır. Testin asıl uygulaması için 2017-2018 eğitim öğretim yılında Balıkesir ili Altıeylül ilçesindeki merkez ilkokullarında (4.sınıf) okuyan 203 öğrenci seçkisiz örnekleme alma yöntemi ile belirlenmiştir. Testin sonuçları R programı ile incelenmiştir. Testin madde güçlüğü, madde ayırt edicilik indeksi, güvenirlik ve yapı geçerliği anlamında tetrakorik faktör analizi ile yapılmıştır. Madde güçlük indeksi 0,19'dan küçük olan 5 soru, madde ayırt edicilik indeksi sıfırın altında olan 2 soru testten çıkarılmıştır. Böylece çalışmada kullanılacak 16 tane çoktan seçmeli, 3 tane açık uçlu sorudan oluşan 19 soruluk nihai test oluşturulmuştur. Testin güvenirlik analizi sonucu KR-20 değeri 0,68 olarak bulunmuştur. Testin cevaplama süresi 30 dakika olarak belirlenmiştir. Yapılan geçerlik ve güvenirlik analizleri sonucunda, test, 4. Sınıf Matematik dersi geometrik şekiller konusunda geçerli ve güvenilir bir test olarak kullanılabilir.

Anahtar kelimeler: Başarı testi, geçerlik ve güvenirlik, geometrik şekiller.

İnönü Üniversitesi
Eğitim Fakültesi Dergisi
Cilt 22, Sayı 3, 2021
ss. 2232-2258
DOI:10.17679/inuefd.879853

Makale Türü
Araştırma Makalesi

Gönderim Tarihi
13.02.2021

Kabul Tarihi
05.12.2021

Önerilen Atıf

Öksüz, C. ve Genç, G. (2021). 4.sınıf geometrik şekiller konusu başarı testi geçerlik ve güvenirlik çalışması. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, *22*(3), 2232-2258. DOI:10.17679/inuefd.879853

Genişletilmiş Özet

Giriş

Okullarda öğrencilere bilişsel (zihinsel), duyuşsal ve devinimsel becerilerin kazandırılması amaçlanmaktadır (Özçelik, 2013). Bu sebeple öğrencilerin hedeflenen düzeyde başarılı olup olmadığını tespit etmek için öğrencilerin kazanması beklenen becerilerinin ölçülmesi gerekmektedir. Bu durumda eğitim sürecinde daha çok bilişsel becerilerin ölçüldüğü başarı testleri kullanılmaktadır. Eğitimde başarı testleri, öğrencilerin öğrenme seviyelerini görmede, konuların ne kadarının ne düzeyde kazanıldığının belirlenmesinde ve olası öğrenme güçlüklerinin saptanmasında kullanılan ölçme araçlarıdır (Gül, 2014).

Özellikle ilkokul döneminde testlerin geçerli ve güvenilir olması çok önemlidir. İlkokul, eğitimin ilk seviyesi olduğu için, her eğitim sisteminde çok önemlidir; bu seviyedeki herhangi bir hata, eğitim sisteminin diğer seviyelerine de nüfuz edecektir. Toplumdaki eğitim sisteminin her bir düzeyinde matematik öğretiminin hedeflerine ulaşmak için, eğitim süreçlerinin ve ürünlerinin kalitesinin izlenmesi ve sürdürülmesi gerekir. Okullardaki matematik öğretimi ve öğreniminin kalitesini ve standartlarını izlemenin en önemli yollarından birisi, öğrencilerin öğrenme çıktılarının nitelikli bir test ile değerlendirilmesidir. Bu sebeple çalışmada, yenilenen "İlkokul 2018 Matematik Öğretim Programı" kapsamında 4. sınıf geometrik şekiller konusu ile ilgili kazanımlara yönelik başarı testinin geliştirilme süreci detaylı olarak anlatılmıştır.

Amaç ve Önem

Geometri matematiğin en önemli alt öğrenme alanlarından biridir ve tarihçesi en az sayılar kadar eskidir (Akkaya, 2018). Geometri konusu hem şekiller hem de cisimler konularını içermesinden dolayı, öğrencilerin yaşadıkları çevreyi ve dünyayı daha iyi anlayıp tanımalarını sağlar (Pesen, 2003). Bireyler, geometrik şekillerin birbirine benzeyen ve farkı olan özelliklerini anlamaya çalışırlar ve süreç içinde şekillerin özelliklerini keşfetmeye çalışırlar (Van De Walle, Karp ve Bay-Williams, 2012). Bu sebeple öğrencilerin daha iyi bir öğretim yöntem ve tekniği ile anlatılan matematik dersleri öğrencilerin güçlüklerini azaltabilir. Aynı zamanda her geçen gün geometri öğretiminde uygulanacak yeni yöntem ve tekniklerin etkili olup olamayacağını anlamak için geçerli ve güvenilir testlere ihtiyaç vardır. Geliştirilen bu testle yeni yöntem ve tekniklerin etkililiği test edilebilecek ve alan yazına olduğu gibi uygulamaya da katkıda bulunabilecektir.

Yöntem

Çalışmanın Deseni

Test geliştirmenin ilk adımı olarak testin amacı belirlenmiştir. Bu araştırmanın amacı, matematik programına bağlı olarak (2018) ilkokul 4. sınıf matematik dersi “Geometri” öğrenme alanındaki “Geometrik şekiller” konusuna yönelik bir başarı testi geliştirmektir. Testin geliştirilmesi aşamasında literatür taraması yapılmış ve İlkokul 4. Sınıf Matematik Programında (2018) “Geometrik Şekiller” konusu ile ilgili kazanımlar listelenmiştir. Bu kazanımlar doğrultusunda Bloom taksonomisine göre hangi seviyede soruların çoğunlukta olmasının gerekliliğine ilişkin bir belirtke tablosu oluşturulmuştur. Sonraki aşamada araştırmacılar tarafından belirtke tablosuna göre denemelik maddeler yazılmıştır. Denemelik maddeler uzmanlara (Türkçe, Matematik ve Ölçme Değerlendirme uzmanı) gönderilerek geçerlilik konusunda dönütler alınmıştır ve uzmanlardan gelen dönütlere göre test formu tekrar düzenlenmiştir. Test daha önce konuyu görmüş olan 4. Sınıftan 8 öğrenciye birebir uygulanmış ve uygulama sürecinde hatalı ve anlaşılamadığı düşünülen sorular tekrar düzenlenmiştir. En son aşamada seçkisiz örneklem alma yöntemi ile 4. Sınıftan 203 öğrenciye uygulanmıştır. Bu uygulamadan sonra test maddeleri analiz edilerek testin son hali oluşturulmuştur.

Katılımcılar

Başarı testi geçerlik ve güvenirlik çalışmasının katılımcıları 2017- 2018 eğitim- öğretim yılı ilkokul 4. sınıf "Geometrik Şekiller" konusunu görmüş Balıkesir ili Altıeylül ilçesine bağlı merkez ilkokullardaki 4. sınıfta öğrenim gören 203 öğrenci oluşturmaktadır. Bu öğrencilerden 113'ü erkek, 90'ı kız öğrencidir.

Veri Toplama Araçları

İlkokul 4. sınıf matematik programına (2018) dayanılarak "Geometrik Şekiller" konusu ile ilgili kazanımlar tespit edilmiş ve Bloom taksonomisine göre hangi seviyede soruların çoğunlukta olmasının gerekliliğine ilişkin bir belirtke tablosu oluşturulmuştur. Belirtke tablosu doğrultusunda kazanımlara uygun 19 madde araştırmacılar tarafından geliştirilmiş, ancak 7 madde için farklı tarzda soruların artırılması için bazı kaynaklar taranarak oradaki sorulara benzer maddeler de türetilmiştir.

Verilerin Analizi

Başarı testinin maddelerinin kontrolü için uzmanlara (Bir Matematik uzmanı, Bir Türkçe alan uzmanı, Bir Ölçme Değerlendirme uzmanı) başvurulmuştur. Buna göre uzmanların görüşü doğrultusunda testte gerekli düzenlemeler yapılmıştır. Bu aşamadan sonra, hazırlanan soru maddeleri zorluk derecesi ve kazanımların sırasına göre uygun şekilde teste dağıtılmıştır. Teste öğrenciler için gerekli açıklama ve yönergeler de yazılmış olup uygun punto büyüklüğünde test formunun yazımı araştırmacılar tarafından gerçekleşmiştir. Sonraki aşamada testin uygulanması için cevaplama süresi 40 dakika belirlenmiştir.

Başarı testinin geçerlik ve güvenirlik çalışması kapsamında R programında tetrakorik faktör analizi yapılmıştır. Bu kapsamda R programında "CTT" ve "psych" paketleri kullanılmıştır. Öğrencilerin her soruya hangi cevabı verdiği 0 ve 1 olarak kodlanmış R programına girilmiş ve böylece her maddenin seçenek analizi de yapılmıştır. Buna göre başarı testinden öğrenciler en yüksek puan "26" ve en düşük puan "0" puan alabilmektedirler.

Bulgular ve Yorumlar

Çalışmada madde güçlüğü, madde ayırt edicilik indeksi, güvenirlik ve yapı geçerliği anlamında tetrakorik faktör analizi yapılmıştır. Soru maddelerinden 20., 5., 14., 21. ve 17. maddelerin madde güçlüklerinin 0.19'un altında olması ve bu maddelerin çok zor maddeler olması sebebiyle bu maddeler araştırmacılar tarafından testten çıkarılmıştır. 14, 10, 20, 8, 5, 21, 17, 3, 12, 19, 6, 26 ve 24. maddelerin madde ayırıcılıklarının 0.19'un altında olduğu göze çarpmaktadır. Fazla sayıda madde çıkartarak temel bilgileri yoklayan soruların testten çıkartılmasını önleme ve çok fazla madde çıkartarak testin güvenirliğini düşürmeme anlamında madde ayırt edicilik gücü sıfırın altında olan 10. ve 8. maddeler testten çıkartılmıştır (14. ve 20. madde halihazırda madde güçlüğü sebebiyle çıkartılmıştı). Dolayısıyla madde ayırt ediciliği 0 ile 0.19 arasında olan 3., 12., 19., 6., 26.ve 24. maddeler düzenlenerek testte tutulmuştur. Madde çıkartma işlemi gerçekleştirmeden önce KR-20 değeri incelendiğinde bu değerin 0,623 olduğu göze çarpmaktadır. Madde çıkartma işlemi gerçekleştirdikten sonra KR-20 değeri incelendiğinde bu değerin 0,68 olduğu göze çarpmaktadır.

Tartışma ve Öneriler

Araştırmada ilkokul 4. Sınıf geometrik şekiller ile ilgili testin yapılan güvenirlik analizi sonucunda KR-20 değeri 0,688 olarak bulunmuştur. Açık uçlu maddelerden oluşan testler tipik olarak 0,65 ve 0,80 aralığındadır (Nitko & Brookhart, 2016). Bu anlamda testten elde edilen güvenirlik katsayısının (KR-20=0,688) bu aralıkta olması ölçme aracının güvenirliği için yeterli görülmüştür. Bu testin analizlerinin R programı ve tetrakorik faktör analizi ile yapılması da yapılan çalışmalardan (Akgün (2011), Tutak ve Birgin (2008), Kurt (2015), Çilingir (2015), Genç (2016) ve Pandra, Sugiman and Marda (2017) daha farklı analiz yöntemleri kullanıldığını da göstermektedir.

Araştırmada elde edilen sonuçlar doğrultusunda; öğrencilerin başarılarını ölçebilmek amacıyla çoktan seçmeli ve açık uçlu testler yanında daha farklı soru tarzlarında (eşleştirmeli, tanılayıcı ağaç, yapılandırılmış grid) testler de geliştirilebilir. Öğrencilerin daha üst düzey bilişsel becerilerini ölçebilmek amacıyla daha fazla açık uçlu sorular sorulabilir. İlkokul matematik programının yenilemesi ile daha farklı öğrenme ve alt öğrenme alanlarında başarı testleri geliştirilebilir.

Validity and Reliability Study of the Geometric Shapes Achievement Test for 4th Grade

Cumali ÖKSÜZ, Aydın Adnan Menderes Üniversitesi, ORCID ID: 0000-0002-3255-2542
Galip GENÇ, Aydın Adnan Menderes Üniversitesi, ORCID ID: 0000-0003-2447-4844

Abstract

The validity and reliability analyses of developing a test for the subject of geometric shapes in the fourth-grade mathematics curriculum were described in this study. For this purpose, the primary school fourth-grade mathematics curriculum (2018) was reviewed, learning outcomes related to "Geometric Shapes" were listed, and a table of specification was prepared. Based on this table, 21 multiple-choice and five open-ended items were created. After getting the experts' opinions, the items were inserted according to difficulty level and test's face validity. A period of 40 minutes has been set for the implementation of the test. To demonstrate the understandability of the test, eight students from the 4th grade who had seen the subject were interviewed one-on-one, and they were asked to answer the questions aloud, and the arrangements of the incomprehensible questions were made. The students were asked to answer the questions by thinking aloud, and the test was revised accordingly. For the test's actual implementation, 203 students attending primary schools (4th grade) in the Altıeylül district of Balıkesir province in the 2017-2018 academic year were selected by random sampling method. The results of the test were analyzed with the R program. Item difficulty, item discrimination index, internal consistency reliability, and construct validity analysis were performed with tetrachoric factor analysis. Five items with an item difficulty index less than 0.19 and two items with an item discrimination index below zero were removed from the test. As a result, the 19-question final test consisting of 16 multiple-choice and three open-ended questions was created. The KR-20 reliability value was found to be 0.68. The answering time of the test was set as 30 minutes. The validity and reliability analysis results showed that the test could be used as a valid and reliable instrument on the 4th Grade Mathematics course's geometric shapes subject.

Keywords: Achievement test, Validity and Reliability, Geometric Shapes

Inonu University
Journal of the Faculty of Education
Vol 22, No 3, 2021
pp. 2232-2258
DOI:10.17679/inuefd.879853

Article Type
Research Article

Received
13.02.2021

Accepted
05.12.2021

Suggested Citation
Öksüz, C. and Genç, G. (2021). Validity and reliability study of the geometric shapes achievement test for 4th grade, *Inonu University Journal of the Faculty of Education*, *22*(3), 2232-2258. DOI:10.17679/inuefd.879853

Validity and Reliability Study of the Geometric Shapes Achievement Test for 4th Grade

In today's understanding of education, the concept of "school" can be defined as indoor and outdoor spaces where activities are carried out to provide students with cognitive (mental), affective, and motor skills (Özçelik, 2013). To determine whether these activities are successful at the targeted level, the skills that students are expected to gain should be measured. According to these measurement results, it may be necessary to determine the educational activities that did not achieve the desired success, the reasons for the failure, and, if necessary, reorganize the training activities or educational environments. Based on this, the definition of measurement in education, according to Turgut (1982), is to observe a feature and express its results with numbers or different signs (cited in Atılgan, Kan & Doğan, 2018). Therefore, accurate measurement of the education process is closely related to the diversity of measurement tools used. In this way, the levels of students' cognitive and motor skills can be measured. Most of the measurement tools used in the education process are achievement tests in which cognitive skills are measured. Achievement tests in education are widely used measurement tools to determine students' learning levels, identify how much the subject is gained, determine possible learning difficulties, and obtain concrete evidence about students' learning status at different grades (Gül, 2014). These tests should possess three features that must be scientifically present in a measurement tool: validity, reliability, and practicality.

According to Nitko & Brookhart (2016), the validity of a measurement tool indicates the consistency of students' assessment results and the use of these results. Baykul (2015) defines validity as the scale determining the variable's level to be reached with only that scale, thus separating it from other variables. According to Nitko & Brookhart (2016), reliability, another feature that a measurement tool should have, is defined as the degree to which the scores obtained by the students at different times from the same test are the same. Özçelik (2013) describes reliability as the fact that the elements whose characteristics are measured always occur within the framework of the same criteria as long as there is no difference in the measured quality of any asset or event. Another feature that should be included in the measurement tool is usability. According to Özçelik (2013), practicality is defined as the ease of applying a measurement tool.

There are many different forms of achievement tests applied in schools based on these three features. But the most used measurement tool in schools and national examinations are multiple-choice achievement tests. According to Pressley et al. (1997), in many countries, multiple-choice tests are used to place students in a higher institution and measure their cognitive levels (cited in Akbulut & Çepni, 2013). The most important advantages of scoring in these tests, such as being objective, easy to apply and scoring, and applicability to large groups and all levels of education, make multiple-choice tests more prominent. (Atılgan, Kan & Doğan, 2018). The items in these multiple-choice tests are the type of questions in which a question is given to the student with the prepared distractors. However, one of the important criticisms in multiple-choice tests is that the student chooses only one of the given options as the correct answer (Atılgan, Kan & Doğan, 2018). Therefore, using different item types for multiple-choice tests makes the tests more valid and reliable. In addition, the use of "open-ended" questions, which provide more precise information about how students understand the concepts, contributes significantly to the measurement tool's validity and reliability (Koyuncu, 2017). Because in open-ended questions, there is almost no chance to find the correct answer randomly. In such questions, after reading the question, the student is required to write the answer he/she found in the allocated space (Özçelik, 2010). According to Airasian and Russell (2012), teachers and experts should diversify item types in the measurement tools to increase validity and reliability (cited in Gül, 2014).

Developing an accurate achievement test is a systematic process because the features of the test to be developed are predetermined. In other words, the preparation of a test whose features are predetermined makes it necessary to follow systematic process steps (Atılgan, Kan & Doğan, 2018). The motivation for test development often stems from a practical concern: can we help children learn, can we identify effective managers, can we identify those at risk of mental distress. The formal starting point for all test development is to generate a construct definition, which broadly is a definition of what is to be measured. An initial construct definition should be as clear as possible (Irwing and Hughes, 2018; 5). Therefore, to administer a standard achievement test in schools and evaluate students' achievements through its results, it is necessary to consider the target learning achievements set by the Ministry of National Education (MoNE) for each course and grade level (Çakan, 2003).

It is essential that the tests are valid and reliable, especially at the primary school level. The primary school level is very important in any educational system because it is the first level of education; any fault would affect other levels of the educational system. To realize the objectives of teaching mathematics at any level of the educational system in the society, there is a need to monitor and maintain the quality of the educational processes and products. One of the most important ways of monitoring the quality and standards of the teaching and learning of mathematics in schools is to evaluate students' learning outcomes with a qualified test. Tests and other assessment tools are used during the instructional process to guide, direct, and monitor students' learning progress towards attaining the course objectives (Alonge, 2004; Kolawole, 2010). This monitoring of learning achievements in mathematics involves the processes of testing, measurement, assessment, and evaluation (Margaret and Anthonia, 2017).

For early childhood, the domain of geometry and spatial reasoning is an important area of mathematics learning (NCTM 1991, 2006). Geometry can serve as a core-relating science and mathematics. Two of the most prominent physicists of the last 100 years attributed their advancements to geometry. As a boy, Einstein was fascinated by a compass, leading him to think about geometry and mathematics. He taught himself extensively about geometry by age 12 (Clements and Sarama, 2011). Therefore, geometry has an essential place in the mathematics course.

Mathematics and geometry are among the first courses in which achievement tests should be prepared with precision. More or less geometry subjects are included in the curriculum of all countries in basic education (Duatepe and Ersoy, 2001). There are three achievements for 4th-grade students in primary school concerning geometrical shapes subject in "2018 Primary School Mathematics Curriculum" geometry sub-learning area (MoNE). Geometric shapes sub-learning area in the curriculum has achievements in all grades of primary school starting from the first grade. There is one achievement in the 1st grade of primary school, three in the 2nd grade, two in the 3rd grade, and three in the 4th grade regarding the geometric shapes learning area. Research in geometry has shown that students have misconceptions and confuse the concepts. In the study conducted with 4th-grade students, Akkaya (2018) concluded that students have misconceptions regarding geometrical shapes such as triangle, square, rectangle, circle, and round concepts. In addition, in the studies, they conducted Başışık (2010), Dağlı (2010), and Özkan (2015) have identified that students have misconceptions about geometrical shapes. For this reason, in the study within the scope of the renewed "Primary School 2018 Mathematics Curriculum", the development process of the achievement test for the achievements related to the subject of geometric shapes in the 4th grade is explained in detail.

Purpose and Importance

We are faced with geometry, which is a part of our daily life, with its dimension concerning everyone consciously or unconsciously. All objects and matters found around appear as geometric structures (Öksüz, 2010). Thus, geometry is included in the curriculums starting from primary education (Baykul, 2009). In this regard, geometry should be seen not only as a learning area of mathematics but also as a tool to recognize and make sense of the world we live in (NCTM, 2000). Although geometry is included in both daily life and curriculums, studies show that students are less successful in geometry compared to the other fields of mathematics (Clements & Battista, 1992; Ubuz, 1999; Başışık, 2010; Öksüz, 2010).

Comparing our country's students with other countries' students in the international exams conducted in recent years shows that our education system lags behind world standards. For example, regarding the Trends in International Mathematics and Science Study (TIMSS) 2015, 49 countries have participated in the study at the 4^{th}-grade level and 39 countries at the 8^{th}-grade level. Turkey has participated in TIMSS at both levels. Turkey's 4^{th}-grade students' average mathematics achievement score was 483, below the TIMSS's standard score of 500. Turkey ranks 36^{th} among 49 countries. The 4^{th}-grade average geometry score was 475, and Turkey was below the overall average of 483. The average score in numbers was 489, and it was 476 in data. Besides, our country's 4^{th}-grade geometry average score was 448 in TIMMS 2011, and it increased by approximately 30 points in TIMSS 2015 (Yıldırım et al., 2016). Despite this, our country's success, which ranks 36^{th} among 49 countries, was not at the desired level.

This situation raises the question why the achievement tests applied in our country cannot achieve international success. To overcome this problem, it is necessary to measure students' knowledge within the framework of the skills required by the age. However, when the national studies conducted at the primary school level in our country are examined, it is seen that the development of measurement tools in mathematics is insufficient; the number of developed valid and reliable measurement tools is very scarce. The purpose of this test development study to be carried out is to increase the number of valid and reliable tests and to use the developed measurement tool in studies and examine cognitive levels.

Although geometry is one of the twin formal pillars of mathematics (Atiyah 2002), the teaching of geometry at all school levels for the past few decades has been in decline (Barbin and Rogers 2016; Mammana and Villani 1998; Olkun et al. 2017; Seah and Horne, 2020). A learning progression/trajectory approach to researching geometry education offers a solution to reverse this trend. We consider that there is a hierarchical progression in the development of geometric reasoning. This can be charted accordingly, and in turn, instructions be designed to target specific levels (Battista 2007; Clements and Sarama 2011; Clements et al. 2004). Valid and reliable tests to be prepared for the geometry learning area for these learning situations will be at an important point in determining the success of primary school students. With these tests, the effect of new methods on geometry can be examined.

Although geometry, which is learned from early childhood, is one of the most important sub-learning areas of mathematics, its history is at least as old as numbers (Akkaya, 2018). Geometry helps students comprehend better and know the environment they live in as it covers both shapes and objects (Pesen,2003). Individuals try to understand the similarities and differences of the geometrical shapes and discover the features of the shapes during this process (Van De Walle, Karp & Bay-Williams, 2012). The geometry creates a complex structure based on human thoughts and some axioms (propositions). As these structures cannot be explained with the meaning of daily life, students experience difficulty in learning these subjects (Mullis et., al., 2000; Fidan,2019). For this reason, mathematics lessons taught with a better teaching method and technique can reduce difficulties experienced by students. At the same time, there is a need for valid and reliable tests to evaluate the efficiency of the new methods and techniques to be used in teaching geometry. With the developed test efficiency of the new method and techniques will be tested, and it would contribute to the practice as well as literature.

Related Studies

Akgül (2011) developed an achievement test with 15 questions, with a Cronbach Alpha value of 0.83 for the angles subject of 4^{th}-grade. Tutak and Birgin (2008) developed an achievement test consisting of 20 multiple-choice questions, covering “Triangle, Square, and Rectangle,” with a KR-20 reliability coefficient of 0.84. Fidan (2013) developed an achievement test for "Numbers." This test's KR-20 reliability coefficient was 0.80 for primary school 1^{st} graders, 0.92 for 2^{nd} graders, 0.93 for 3^{rd} graders, and 0.95 for 4^{th} graders. The final form of the tests consists of 13 items for 1^{st} graders, 15 items for 2^{nd} graders, 16 items for 3^{rd} graders, and 24 items for 4^{th} graders. Kurt (2015) developed an achievement test consisting of 34 questions for 4^{th}-grade students, with a Cronbach alpha value of 0.84. Çilingir (2015) developed an achievement test consisting of 17 questions with a KR-20 value of 0.80 using the "geometry" questions of the "geometric shapes and measurement" part of the TIMSS 2007 4^{th} grade mathematics questionnaire. Genç (2016) developed a 30-item achievement test on decimal fractions for 4^{th}-grade students with a KR-20 value of 0.86. Pandra, Sugiman and Mardapi (2017) developed a mathematics achievement test with a reliability coefficient of 0.783.

Few studies have developed achievement tests targeting primary school mathematics education in Turkey. This study will contribute to filling the gap in the test development area.

Methodology

Design of the Study

The steps on how to develop an achievement test were followed in the research. Baykul (2015), Atılgan, Kan, and Doğan (2018) recommended to follow the steps below in test development:

1. Identifying the purpose for which the test will be used,
2. Identifying the behaviors to be measured by the test and creating a table of specifications,

3. Writing and reviewing the items,
4. Preparation of pilot test questionnaire,
5. Administration of the pilot test,
6. Scoring pilot questionnaire's answers, performing item analysis and selecting items,
7. Creating the final test and estimating the statistics.

Accordingly, first, the purpose of the test was identified as told in the first step. The purpose of this study is to develop an achievement test for "Geometric shapes" covered in the primary school 4^{th}-grade mathematics course's "Geometry" learning area, according to the mathematics curriculum (2018). A literature review was performed during the test development; the learning outcomes related to "Geometric Shapes" covered in the Primary School 4^{th} Grade Mathematics Curriculum (2018) were listed. A table specification was created in line with the learning outcomes shown in Table 1 using Bloom's taxonomy to determine question levels. In the next step, researchers wrote draft items according to the table of specifications. The draft items were sent to experts (Turkish, Mathematics, and Measurement and Assessment experts), feedback on their validity was received, and the questionnaire was revised according to experts' opinions. The test was administered to 8 students from the 4^{th}-grade who had studied the subject before. The questions that were thought to be inaccurate and incomprehensible were revised. In the last stage, the test was applied to 203 4^{th}-grade students selected by random cluster sampling. Then, the test items were analyzed, and the test's final version was formed.

Participants

The participants of the achievement test's validity and reliability study consisted of 203 fourth-grade students attending primary schools of Altıeylül district of Balıkesir province, who studied "Geometric Shapes" in the 4^{th} grade of primary school in the 2017-2018 academic year. 113 of these students were male, and 90 of them were female. According to Nunnally (1967), a sample of at least five times the number of items in the test is needed to analyze the items in test development. According to that, the number of participants was considered sufficient for the reliability and validity study of the test.

Data Collection Tools

The achievement test developed by the researchers was used as the data collection tool in the validity and reliability study. This test, consisting of 21 multiple choice questions and five open-ended questions, was administered to 203 students, and the validity and reliability analyses were performed.

Based on the literature and the revised latest primary school 4^{th}-grade mathematics curriculum (2018), the "Geometric Shapes" subject's learning outcomes were identified, and a table of specifications has been created regarding which level of questions should be the majority according to Bloom's taxonomy:

Table 1

Achievement Test Table of Specifications

Geometric Shapes Achievement Test *Table of Specifications*								
Learning Area (Geometric shapes)	Cognitive Domain (Bloom)							
Learning outcomes	Recall	Understanding	Application	Analysis	Evaluation	Creation	Total	Percent
1. Names the sides and corners of the triangle, square, and rectangle.	1	2	1				4	15.3%
2. Knows the side properties of the square and rectangle.	1	4	3	2	2	2	14	53.8%
3. Classifies triangles according to their side lengths.	1	2	1	2	1	1	8	30.9%
Total	3	8	5	4	3	3	26	100%
Percentage	11.5%	30.8%	19.2%	15.5%	11.5%	11.5%	100%	

In line with the *table of specifications*, 19 items related to the learning outcomes were developed by the researchers. The sources were reviewed to include different types of questions, and seven items similar to the questions in these sources were derived. Cited sources are given in the table below:

Table 2

Test Questions, Learning outcomes, and References

Learning outcomes	Item No.	References
1. Names the sides and corners of the triangle, square, and rectangle.	1, 2, 3, 18	İnan, Ş. (2016) Küçükaydın, A. (2017) Karadağ, S., Balcı, M., Abdik, E. Ve Demiralp, A. (2017)
2. Identifies the side properties of the square and rectangle.	7, 8, 9, 11, 12, 13, 14, 16, 19, 20, 21, 22, 25, 26	
3. Classifies triangles according to their side lengths.	4, 5, 6, 10, 15, 17, 23, 24	

In the draft item writing phase of the test, open-ended questions and multiple-choice questions were used, allowing students to express their high-level skills, such as decision-making, and explain the answers using their natural language. According to Turgut and Baykul (2014), there is a common belief that open-ended questions can be used to check simple information within the recall step's scope. However, high-level cognitive skills can also be examined with open-ended questions (Atılgan, Kan, & Doğan, 2018). Besides, in open-ended questions, the student is required to think, design, edit, and write the answers by himself (Turgut & Baykul, 2014). High-level skills, such as mathematical symbols usage, are also examined in open-ended questions (Nitko & Brookhart, 2016). For this reason, achievement-related open-ended questions were added to the *table of specifications* in addition to the multiple-choice questions.

Data Analysis

After the draft test items are created, they should be checked from different aspects, their defective features and deficiencies should be detected and corrected (Atılgan, Kan, & Doğan, 2018). Experts' opinions (Mathematics, Turkish, and Assessment and Evaluation experts) were taken to check the achievement test's items. Experts reviewed the test within the validity, accuracy, and suitability criteria, and the test was revised according to their suggestions. The test form is examined with the experts face to face. The validity of the test and whether it is correct scientifically was discussed with mathematic and statistical experts. Spelling and language mistakes were revised with a Turkish expert. Necessary revisions were made in accordance with the feedback from experts. The prepared items were then distributed to the test according to their degree of difficulty and sequence of the learning achievements. The necessary explanations and instructions were also written in the test for the students. The researchers then wrote the questionnaire using understandable language and the appropriate font size considering students' developmental characteristics. In the next step, the response time for the test is determined. Baykul (2015) suggested 60- 70 seconds per question for mathematic tests. As there are open-ended questions in this test, considering the difficulty level of the questions, response time is determined as 40 minutes. In the next stage, the test was administered one-to-one to eight 4th grade students who had already studied the subject. The students were asked to think aloud while answering the questions, and the incomprehensible questions were re-evaluated regarding students' gestures and mimics. After the one-to-one application, 203 fourth grade students in the Altıeylül district of Balıkesir province were selected by random cluster sampling. The 26-item achievement test, consisting of 21 multiple-choice and five open-ended questions, was administered to these students within the given time.

Tetrachoric factor analysis was performed in the R program for the achievement test's validity and reliability study. In this context, "CTT" and "psych" packages were used in the R program. The students' answers were coded as 0 and 1 for each item and entered into the R program, and each item's option analysis was made. The options of twenty-one multiple-choice questions were examined as four options. According to the given answers, open-ended questions were transformed into multiple-choice with four options (one correct, three incorrect) to be analyzed in the R program. In open-ended questions, the values of students' incorrect answers were considered, and these values were entered into the system as an option and subjected to analysis. While identifying these options, researchers worked with assessment and evaluation expert and mathematic expert, and the answers to the questions were transformed into one correct three incorrect options. For example, a square was drawn, and the student was asked to name the square with appropriate letters and write the notation correctly in the reserved space below the figure. The answer is coded as A for those who named and wrote it correctly; as B for those who named it correctly but failed to write; as C for those who named it correctly but wrote the name in reverse order; and as D for those who named it correctly but named it by writing the corners of the shape diagonally. For the test results, “1” point was given to each correct answer, and “0” to each wrong answer. The total score that a student gets from the test is the number of correct answers. Accordingly, the highest score that students can get from the achievement test is 26, and the lowest score is “0”.

Ethical permission was obtained from Adnan Menderes University Educational Research Ethics Committee (09.02.2021-2328) for this study.

Findings and Comments

This part includes tetrachoric factor analysis results related to item difficulty, item discrimination index, reliability, and construct validity. The analyzes were carried out in the R program using "CTT" and "psych" packages. While calculating the Kuder Richardson-20 (KR-20) statistics, a manual calculation was performed using the R program. Item analysis results of the achievement test in the R program and KR-20 value, which indicates the reliability of the achievement test, are given in the table below:

Table 3

Item Difficulty Index and Item Discrimination Index of the Geometric Shapes Achievement Test

Questions	Item Difficulty Index	Item Discrimination Index	KR-20
1	0.60	0.24	0.62
2	0.28	0.33	
3	0.58	0.10	
4	0.44	0.25	
5	0.04	0.07	
6	0.32	0.15	
7	0.52	0.26	
8	0.24	-0.01	
9	0.73	0.26	
10	0.33	-0.07	
11	0.70	0.35	
12	0.34	0.11	
13	0.81	0.23	
14	0.12	-0.10	
15	0.73	0.29	
16	0.67	0.35	
17	0.18	0.09	
18	0.19	0.20	
19	0.40	0.14	
20	0.01	-0.02	

21	0.16	0.09
22	0.46	0.11
23	0.59	0.41
24	0.55	0.17
25	0.28	0.29
26	0.55	0.16

Table 3 includes item difficulty and item discrimination values and the reliability value, which indicates internal consistency. According to Sözbilir (2010), item difficulty should be interpreted as follows; 0.80 and above - very easy; between 0.65 and 0.79 – easy; between 0.35 and 0.64 – moderate; between 0.20 and 0.34 – difficult; 0.19 and below - very difficult.

Figure 1

Item Difficulty Values

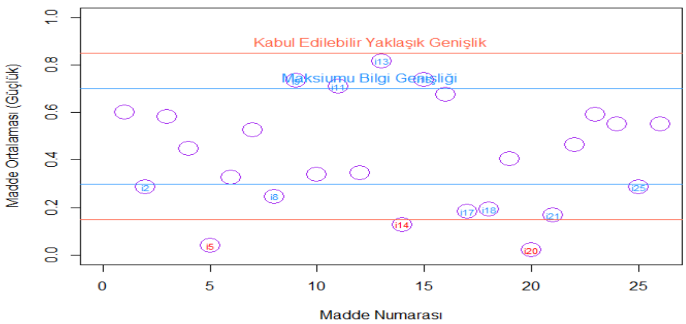
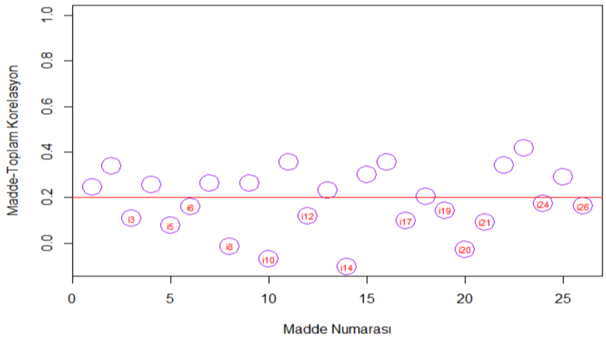


Figure 1 shows the item difficulty range that the items fall within. The orange lines in Figure 1 represent the critical points; only the items positioned above/below it were taken into account. The area between the orange lines in Figure 1 shows the acceptable zone in terms of item difficulty; the region between the blue and the orange line at the top shows the items with complete information.

The review of the item difficulties in Table 3 and the positions of the items in Figure 1 regarding their difficulty in the two-dimensional analytical plane together shows that the item difficulties of item 20, 5, 14, 21, and 17 are below 0.19. The fact that the item difficulty of the relevant items is below 0.19 indicates that these items are very difficult. Therefore, they were excluded from the test by the researchers.

Figure 2

Item Discrimination Values



In the test development process, test items were also evaluated according to their discrimination index. The resulting item discrimination value should be interpreted as follows: Items with a value of 0.19 or less should never be included in the test or should be revised completely. Items with values between 0.20 and 0.29 are borderline items and should be included in the test after a revision. Items with values between 0.30 and 0.39 can be included without any modification or with minor revisions. Items with a value of 0.40 and higher are the ones that work very well; they can be taken as they are (Atılgan, Kan & Doğan, 2018).

Figure 2 shows the discriminatory power of the items. The orange line in Figure 2 represents the critical point in terms of item discrimination. The review of the item discrimination in Table 3 and the positions of the items in Figure 2 regarding their discrimination in the two-dimensional analytical plane together shows that the discriminatory power of the items 14, 10, 20, 8, 5, 21, 17, 3, 12, 19, 6, 26 and 24 are below 0.19. To prevent excluding the questions involving basic knowledge and decreasing the test's reliability by removing too many items, items 10 and 8, with a discriminatory power below zero, were removed from the test (items 14 and 20 have already been removed due to item difficulty). Therefore, items 3, 12, 19, 6, 26, and 24, whose item discrimination is between 0 and 0.19, were revised and kept in the test.

The KR-20 value, which indicates the reliability, was 0.623 before removing the items. After removing the items, the KR-20 value became 0.68. According to Nitko & Brookhart (2016), the tests' reliability value having open-ended items is expected to be between 0.65 and 0.80. Considering this range and the increase in the KR-20 value, the test's reliability was concluded to be sufficient.

The results of the tetrachoric factor analysis in the R program are as follows:

Table 4

Descriptive Statistics of the Items

Variable	Mean	Confidence Interval (95%)	Variance	Skewness	Kurtosis
M1	0.598	(0.51 0.69)	0.240	-0.202	-1.834
M2	0.284	(0.20 0.37)	0.203	0.961	-1.076

M 3	0.578	(0.49 0.67)	0.244	-0.119	-1.894
M 4	0.446	(0.36 0.54)	0.247	0.218	-1.948
M 5	0.324	(0.24 0.41)	0.219	0.258	-1.423
M 6	0.525	(0.43 0.61)	0.249	-0.099	-1.985
M 7	0.730	(0.65 0.81)	0.197	-1.043	-0.912
M 8	0.706	(0.62 0.79)	0.208	-0.208	-1.174
M 9	0.343	(0.26 0.43)	0.225	0.664	-1.556
M 10	**0.814**	(0.74 0.88)	**0.152**	-1.620	0.615
M 11	0.735	(0.66 0.81)	0.195	-1.072	-0.852
M 12	0.672	(0.59 0.76)	0.221	-0.734	-1.459
M 13	**0.191**	(0.12 0.26)	0.155	1.578	0.484
M 14	0.402	(0.31 0.49)	0.240	0.402	-1.834
M 15	0.461	(0.37 0.55)	**0.248**	0.158	-1.970
M 16	0.588	(0.50 0.68)	0.242	-0.360	-1.866
M 17	0.549	(0.46 0.64)	**0.248**	-0.198	-1.956
M 18	0.284	(0.20 0.37)	0.203	0.961	-1.076
M 19	0.549	(0.46 0.64)	0.248	-0.198	-1.956

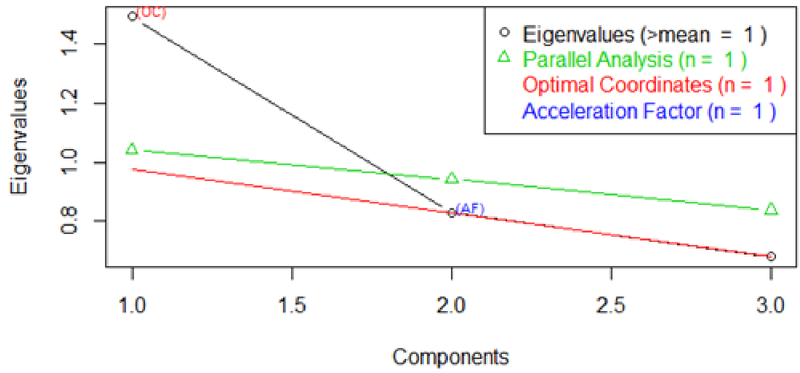
Regarding Table 4, which includes descriptive statistics of the items in the achievement test, the average values of the questions vary between 0.19 and 0.81, and the variances vary between 0.15 and 0.24. The variances are very close to each other, and the kurtosis-skewness coefficients are between the -3 to +3 range.

The Kaiser Meyer Olkin (KMO) test and Bartlett's Test of Sphericity were used for checking whether factor analysis can be applied to the achievement test items. KMO value (0.82) and Bartlett test were statistically significant at $\alpha = 0.001$ level. Thus H_0 hypothesis that "universe correlation matrix is identity matrix" was rejected. The sample size was deemed sufficient to apply factor analysis to the achievement test. Besides, considering that the number of participants forming the workgroup was 203 and that 200 participants are adequate for factor analysis (Kline, 1994), the workgroup's size was concluded to be sufficient.

In deciding the number of factors, the factors with an eigenvalue greater than one are taken into account; it is seen that the test has one factor. As the last step in determining the ideal number of factors, the Eigenvalue-Component Graph in Figure 3, which allows visual and numerical comparison of various methods, was used. According to Figure 3, the ideal number of factors is 1, considering the acceleration factor, parallel analysis, and optimal coordinates. As a result of the holistic evaluation of the Kaiser method, acceleration factor, Horn's parallel analysis, and optimal coordinates, it was concluded that the achievement test has a single factor.

Figure 3.

Eigen Value-Component Plot for Ideal Number of Factors



As the achievement test has a single factor, in other words, the only factor is not associated with another factor; it was decided to use the Varimax vertical rotation technique in the tetrachoric factor analysis. Regarding the total variance explained in the tetrachoric factor analysis, the single factor with an eigenvalue greater than 1 (eigenvalue = 6.64) explains 34% of the variance. It is sufficient for the explained variance to be 30% or more in single-factor measurement tools (Kline, 1994). Besides, regarding the factor loads related to the items that constitute the achievement test, they vary between 0.51 and 0.70. Factor loads of the items should be at least 0.32 to be interpreted within the scope of Factor analysis (Comrey & Lee, 1992); therefore, factor loads of the items are found to be sufficient.

Conclusion, Discussion and Suggestions

In this study, an achievement test involving geometric shapes covered in the 4^{th}-grade of primary school was developed, and the validity and reliability study of the test was discussed. As a result of the tetrachoric factor analysis conducted within the study's scope, the model's goodness of fit statistics was calculated. The goodness of fit indices is a measure of the variance and covariance explained by the model. They can be interpreted as the coefficient of determination (R2) of multiple regression. The closer the goodness of fit indexes to 1, the more compatible the model with the data. Model fit indices are found to be within the ranges of numerical criteria indicating model fit (RMSEA = 0.06, CFI = 0.91, GFI = 0.90, AGFI = 0.90, NFI = 0.90). According to the literature, RMSEA values below 0.10 are acceptable (Cole, 1987); Absolute fit indices GFI and AGFI between 0.90-0.95 indicates a near-perfect fit (Baumgartner & Hombur, 1996; Hooper, Coughlan & Mullen, 2008); and incremental fit indices NFI and CFI in the range of 0.90-0.95 indicate an acceptable fit (Bentler, 1992). As a result of the holistic review of the model fit statistics, it was concluded that the model fit of the achievement test developed within the scope of the study was ensured. Utilizing tetrachoric factor analysis in R program has shown that a different analysis method was used compared to other conducted studies; Akgül (2011), Tutak & Birgin (2008), Kurt (2015), Çilingir (2015), Genç (2016), and Pandra, Sugiman and Marda (2017).

The term validity refers to whether or not a test measures what it claims to measure. On a test with high validity, the items will be closely linked to the test's intended purpose (Larsen and Puck, 2020). In this case, the test has been prepared for which the target is tested with the intended purpose.

As a result of the test's reliability analysis, the KR-20 value was found to be 0.688. In the computation of this reliability value, each problematic item was removed from the test one by one. The reliability was calculated according to the change in the reliability coefficient. At the end of the reliability test analysis, it was found that the test has a lower reliability value compared to the following studies; Akgül (2011) Tutak & Birgin (2008), Kurt (2015), Çilingir (2015), Genç (2016) and Pandra, Sugiman and Marda (2017). Özdamar (1999) and Tavşancıl (2006) stated that a considerably reliable test has a reliability coefficient between 0.60 and 0.80. The reliability of the tests that include open-ended items is typically between 0.65 and 0.80 (Nitko & Brookhart, 2016). In this sense, as the reliability coefficient (KR-20 = 0.688) obtained from the test falls within this range, it is sufficient for the measuring tool's reliability.

The ability to describe, use, and visualize the effects of composing and decomposing geometric shapes is a major conceptual field and set of competencies in the domain of geometry (Clements, Wilson, & Sarama, 2004). Therefore, this test about geometric shapes will be an essential test to be used in the field of geometry.

As Alonge (2004) and Kolawole (2010) stated, tests are guiding and leading tools in students' progress towards achieving their lesson goals throughout the teaching process. The primary school 4th-grade geometric shapes test developed in this study can also be used as a valid and reliable test to show how much students have achieved the goals of the course in schools. In this respect, the test will also be useful for teachers.

There are suggestions for future studies in line with the results obtained in the study. In addition to multiple-choice and open-ended tests, tests with different question types (matching, diagnostic tree, structured grid) can be developed to measure students' achievement. Considering the studies at home and abroad, it is not seen that there is too much place in test development studies. Therefore, valid and reliable tests can be developed, especially in the field of mathematics. In order to measure the higher-level cognitive skills of students in mathematics education, open-ended questions can be asked within the tests to be prepared further. With the renewal of the primary school mathematics program, achievement tests can be developed in different learning and sub-learning areas.

Çıkar Çatışması Bildirimi

Yazar(lar), bu makalenin araştırılması, yazarlığı ve / veya yayınlanmasına ilişkin herhangi bir potansiyel çıkar çatışması beyan etmemiştir.

Destek/Finansman Bilgileri

Yazar(lar), bu makalenin araştırılması, yazarlığı ve / veya yayınlanması için herhangi bir finansal destek almamıştır.

Etik Kurul Kararı

Bu araştırma için Adnan Menderes Üniversitesi Eğitim Araştırmaları Etik Kurulu kurumdan (09.02.2021-2328) etik izin alınmıştır.

Kaynakça/References

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, 2013. https://dergipark.org.tr/tr/download/article-file/19600

Akgül, E. (2011). *İlköğretim 4. sınıf Öğrencilerine Matematik Dersinde "Açılar" Konusunun Öğretilmesinde Proje Tabanlı Öğretim Yönteminin Etkisinin İncelenmesi*, (Yayımlanmamış yüksek lisans tezi), Marmara Üniversitesi, İstanbul.

Akkaya, S. (2018). İlkokul dördüncü sınıf matematik dersinde geometri alt öğrenme alanlarına ilişkin kavram yanılgılarının giderilmesinde oyun temelli öğretimin etkisi, (Yayımlanmamış doktora tezi),İnönü Üniversitesi, Malatya.

Alonge, M.F.(2004). Measurement and evaluation in education and psychology. Ado-Ekiti: Adebayo Printing Press

Atılgan, H., Kan, A. ve Doğan, N.(2018). *Eğitimde Ölçme Ve Değerlendirme*, (Ed: Hakan ATILGAN) Ankara: AnıYayıncılık.

Atiyah, M. (2002). Mathematics in the 20 th century. Bulletin of the London Mathematical Society, 34(1), 1– 15. https://doi.org/10.1007/BF03033096.

Barbin, E., & Rogers, L. (2016). Geometry in the secondary school curriculum and in progression to university. Proceedings of the 9th quadrennial meeting of the History and Pedagogy of Mathematics Group (HPM), 18–22 July. Montpellier: HPM.

Başışık, H. (2010). İlköğretim 5.sınıf öğrencilerinin çokgenler ve dörtgenler konularındaki kavram yanılgılarının belirlenmesi. (Yayınlanmamış Yüksek Lisans Tezi), Adnan Menderes Üniversitesi, Aydın.

Battista, M. T. (2007). The development of geometric and spatial thinking. In F. K. Lester (Ed.), Second handbook of research on mathematics teaching and learning. Charlotte: Information Age Publishing.

Baumgartner, H., ve Hombur, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing, 13*, 139-161.

Baykul, Y. (2009). İlköğretimde matematik öğretimi 6-8. sınıflar. Ankara: Pegem Akademi Yayıncılık.

Baykul, Y. (2015). *Eğitimde ve Psikolojide Ölçme*, Ankara : Pegem Akademi Yayıncılık.

Bentler, P. M. (1992). On the fit of models to covariances and methodology to the Bulletin. *Psychological Bulletin, 112*, 400-404.

Clements, D. and Battista, M. (1992). Geometry and spatial reasoning. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning pp. 420-464, Toronto: Macmillan

Clements, D. H., Wilson, D. C., & Sarama, J. (2004). Young children's composition of geometric figures: A learning trajectory. *Mathematical Thinking and Learning*, 6, 163–184.

https://www.researchgate.net/publication/250890418_Young_Children's_Composition_of_Geometric_Figures_A_Learning_Trajectory

Clements, D. and Sarama, J. (2011). Early childhood teacher education: the case of geometry, *J Math Teacher Educ* (2011) 14:133–148 DOI 10.1007/s10857-011-9173-0

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology, 55*, 1019-1031. Doi: 10.1037/0022-006X.55.4.584

Comrey, A. L., & Lee, H. B. (1992). A First Course in Factor Analysis (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Çakan, M. (2003). Geniş Ölçekli Başarı Testlerinin Eğitimindeki Yeri ve Önemi, *Eğitim ve Bilim,* Cilt 28, sayı: 128 (19-26). http://egitimvebilim.ted.org.tr/index.php/EB/article/view/5130/1212

Çilingir, E. (2015). *Gerçekçi Matematik Eğitimi Yaklaşımının İlkokul Öğrencilerinin Görsel Matematik Okuryazarlığı Düzeyine Ve Problem Çözme Becerilerine Etkisi*. (Yayımlanmamış Yüksek lisans tezi), Çukurova Üniversitesi, Adana.

Dağlı, H. (2010). İlköğretim beşinci sınıf öğrencilerinin çevre, alan ve hacim konularına ilişkin kavram yanılgıları. (Yayımlanmamış Yüksek Lisans Tezi), Afyon Kocatepe Üniversitesi, Afyonkarahisar

Duatepe, A. and Ersoy, Y. (2001). Teknoloji destekli matematik öğretimi-I: hesap makinesi ve okullarda geometri öğretimi, Matematik Etkinlikleri 2001 Sempozyumu, Ankara, 110 119.

Fidan, E. (2013). *İlkokul Öğrencileri İçin Matematik Dersi Sayılar Öğrenme Alanında Başarı Testi Geliştirilmesi*. (Yayımlanmamış Yüksek lisans tezi), Ankara Üniversitesi, Ankara.

Fidan, B. (2019). Üç boyutlu geometrik şekillerin ortaokul öğrencilerine yağlı boya resimler ve dinamik görsellerle öğretilmesinin başarıyla tutuma etkisi. (Yayımlanmamış Yüksek lisans tezi), Bolu Abant İzzet Baysal Üniversitesi, Bolu.

Genç, G. (2016). İlkokul matematik derslerinde olumlu bir söylem ortamının etkisinin söylem analizi yöntemiyle incelenmesi, (Yayımlanmamış doktora tezi), Pamukkale Üniversitesi, Denizli.

Gül, Ç. D. (2014). *Madde Güçlüklerine Göre Farklı Sıralanan Testlerde Düşük Ve Yüksek Kaygılı Öğrencilerin Akademik Başarı Puanlarının Karşılaştırılması*. (Yayımlanmamış Yüksek lisans tezi), Ankara Üniversitesi, Ankara.

Hooper, D., Coughlan, J., ve Mullen, M.R. (2008). Structural equation modelling: Guidelines for determining model fit. *Journal of Business Research Methods, 6*, 53-60.

İnan, Ş. (2016). *4.sınıf ilkokul matematik not defteri*, Ankara: Mutlu.

Irwing, P. and Hughes, D. J. (2018). *Test Development, The Wiley Handbook of Psychometric Testing: A Multidisciplinary Approach to Survey, Scale and Test Development* Editors: Paul Irwing, Tom Booth, David J.Hughes, https://www.researchgate.net/publication/309761309

Karadağ, S., Balcı, M., Abdik, E. Ve Demiralp, A. (2017). *İlkokul 4 Matematik Öğretmen Kılavuz Kitabı*, Ankara: Yakın Çağ Yayınları.

Kline, P. (1994). *An Easy Guide to Factor Analysis*. UK: Routledge.

Koyuncu, E.(2017). *Temel Eğitimden Ortaöğretime Geçiş (Teog) Sınavında Açık Uçlu Soruların Kullanılmasına İlişkin Öğrenci Görüşlerinin İncelenmesi*. (Yayımlanmamış Yüksek lisans tezi), Gaziantep Üniversitesi, Gaziantep.

Kolawole, E.B.(2010). Principles of test construction and administration. Lagos: Bolabay Publications.

Kurt, E.S.(2015). *Gerçekçi Matematik Eğitiminin Uzunluk Ölçme Konusunda Başarı Ve Kalıcılığa Etkisi*. (Yayımlanmamış Yüksek lisans tezi), Ondokuz Mayıs Üniversitesi, Samsun.

Küçükaydın, A.(2017).*4.sınıf Matematik*, Ankara: Mavi Deniz Yayınları.

Larsen, D. M. and Puck, M. R. (2020). Developing a Validated Test to Measure Students' Progression in Mathematical Reasoning in Primary School, *International Journal on Social and Education Sciences* Volume 2, Issue 1, 2020 ISSN: 2688-7061 (Online)

Mammana, C., & Villani, V. (1998). Perspectives on the teaching of geometry for the 21st century. Dordrecht: Kluwer Academic Publishers.

Margaret, I. and Anthonia, M.A.E. (2017). Development and validation of mathematics achievement test for primary school pupils, *British Journal of Education,* Vol.5, No.7, pp. 47-57, July 2017

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrotowski, S. J., and Smith, T. A. (2000). Findings from IES's repeat of third international Mathematics and Science study at the eight grade: International Mathematics Report. Boston College: MA.

MONE. (2018). *Matematik Dersi Öğretim Programı (İlkokul ve Ortaokul 1, 2, 3, 4, 5, 6, 7 ve 8. Sınıflar)*.Ankara: MONE.

NCTM. (1991). Professional standards for teaching mathematics. Reston, VA: National Council of Teachers of Mathematics.

NCTM (2000). Principles and Standards for School Mathematics. Reston, Va. NCTM.

NCTM. (2006). Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence. Reston, VA: National Council of Teachers of Mathematics.

Nitko, A. J. ve Brookhart, S. M. (2016). Değerlendirme ve öğretimin entegrasyonu için planlama (ss. 105-129, çeviren: S. Özel), Öğrencilerin eğitsel değerlendirilmesi (6. Basımdan çev. ed. B. Bıçak, M. Bahar, S. Özel). Ankara: Nobel Akademik Yayıncılık.

Nunnally, J. C. (1967). Psychometric theory. New York, NY: McGraw-Hill Book Company

Olkun, S., Swoboda, E., Vighi, P., Yuan, Y., & Wolring, B. (2017). Topic study group no. 12: teaching and learning of geometry (primary level). In G. Kaiser (Ed.), Proceedings of the 13th International Congress on Mathematical Education Monographs (pp. 429–434). Hamburg: Springer.

Öksüz, C. (2010). İlköğretim Yedinci Sınıf Üstün Yetenekli Öğrencilerin "Nokta, Doğru ve Düzlem" Konularındaki Kavram Yanılgıları*, İlköğretim Online*, 9(2), 508-525, 2010. https://dergipark.org.tr/tr/pub/ilkonline/issue/8595/106917

Özçelik, D.A. (2010). *Okullarda Ölçme ve Değerlendirme El Kitabı,* Ankara :Pegem Akademi Yayıncılık.

Özçelik, D.A. (2013). *Test Hazırlama Kılavuzu,* Ankara :Pegem Akademi Yayıncılık.

Özdamar, K. (1999). *Paket Programlar İle İstatistiksel Veri Analizi I,* Eskişehir :Kaan Kitabevi *(2. Baskı).*

Özkan, M. (2015). 7. sınıf öğrencilerinin çokgenlerde ve özel dörtgenlerde yaptıkları kavram yanılgılarının incelenmesi. (Yayımlanmamış Yüksek Lisans Tezi), Çukurova Üniversitesi, Adana.

Pandra, V., Sugiman and Mardapi, D. (2017). Development of Mathematics Achievement Test for Third Grade Students at Elementary School in Indonesia. *Internatıonal Electronıc Journal Of Mathematıcs Educatıon* e-ISSN: 1306-3030. 2017, VOL. 12, NO. 3, 769-77

Pesen, C. (2003). Eğitim fakülteleri ve sınıf öğretmenleri için matematik öğretimi. Ankara: Nobel

Seah, R. and Horne, M. (2020). The construction and validation of a geometric reasoning test item to support the development of learning progression, *Mathematics Education Research Journal* (2020) 32:607–628.

Sözbilir, M. (2010). Madde Analizi ve Test Geliştirme. 02/12/2020 tarihinde http://olcmevedegerlendirme.wordpress.com/about/ adresinden alınmıştır.

Tavşancıl, E. (2006). *Tutumların ölçülmesi ve spss ile veri analizi,* Ankara: Nobel Yayın Dağıtım.

Turgut, M. F.,& Baykul, Y. (2014). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Pegem Akademi.

Tutak, T. ve Birgin, O. (2008). Geometri öğretiminde bilgisayar destekli öğretimin öğrenci başarısına etkisi, IETC - May 06-08 2008 Eskişehir, TURKEY

Van de Walle, J. A., Karp, K. S. and Bay-Williams, J. M. (2012). İlkokul ve ortaokul matematiği (S. Durmuş, Çev.). Ankara: Nobel Akademik Yayıncılık

Ubuz, B. (1999). 10. ve 11. Sınıf öğrencilerinin geometride kavram yanılgıları ve cinsiyet farklılıkları, *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi*. Özel Sayı:11, 179–184.

Yıldırım, A., Özgürlük, B., Parlak, B., Gönen, E. & Polat, M. (2016). Timss 2015 ulusal mathematics and science preliminary report 4th and 8th grades. http://timss.MoNE.gov.tr/wp-content/uploads/TIMSS_2015_Ulusal_Rapor pdf

İletişim/Correspondence

Prof. Dr. Cumali ÖKSÜZ. cumalioksuz@adu.edu.tr

Dr. Öğr. Üyesi Galip GENÇ. galipgencc@gmail.com

EK:

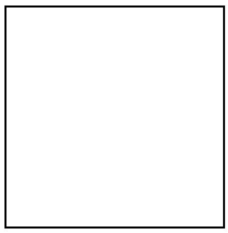
Primary School 4th Grade

Geometric Shapes Achievement test

Explanation

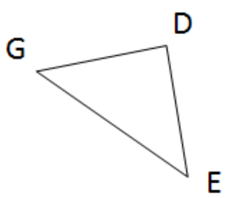
1. This test includes 16 multiple choice and three open-ended questions.
2. Multiple-choice questions have only one answer. You can use spaces to solve questions. You can also answer open-ended questions using the space provided below the question.
3. The time for answering the test is 30 minutes.
4. Only correct answers will be taken into account in the evaluation; wrong answers will not be considered.
5.

1. **Name the geometric figure below with appropriate letters and write the naming correctly in the space below the figure.**



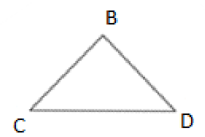
………………………………………………

2. **Write the name of the triangle below in the space below the figure using the triangle symbol.**



………………………………………

3. How many different ways can the triangle on the right be named?



A) 1 B) 2
C) 3 D) 6

4. Imagine an isosceles triangle with one side length 9 cm and the other side 5 cm. How many values can the third side of this triangle take?

A) 1 B) 2
C) 3 D) 4

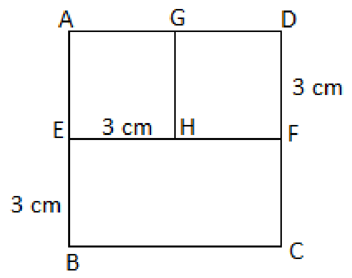
5. I. Acute triangle
II. Equilateral triangle
III. Right triangle
IV. Isosceles triangle
V. Obtuse triangle
Which of the above are types of triangles concerning their sides?

A) I and IV B) Only III
C) II and IV D) I, II, III, IV and V

6. Which of the following <u>cannot be</u> the edge of a rectangle called BEFG?

A) |BF| B) |FE|
C) |BG| D) |GF|

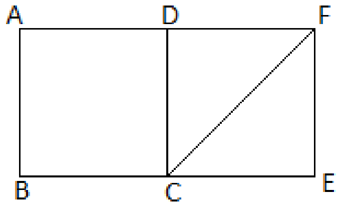
7.



How many squares are there in the above figure?

A) 2 B) 3
C) 4 D) 5

8.



ABCD and CDFE in the figure are square.
EFC is an isosceles triangle.

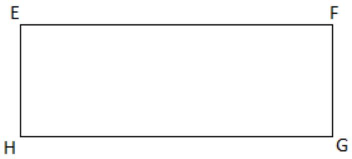
Accordingly, which of the following is true?

A) |EF| = |FB| B) | BC| = |EF|
C) |AF| = |AB| D) | EC| = |EB|

9. How many separate triangles should be drawn to get the number of sides of the three squares separated from each other?

A) 2 B) 4
C) 5 D) 9

10.



In the EHGF rectangle, EH = 4 cm, HG = 7 cm, how many centimeters is FG + FE?

A) 8 B) 11
C) 12 D) 14

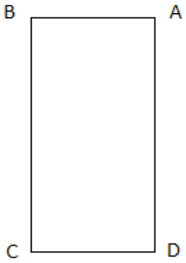
11.

If the line segments join the marked dots on the paper, what type of triangle will be formed?

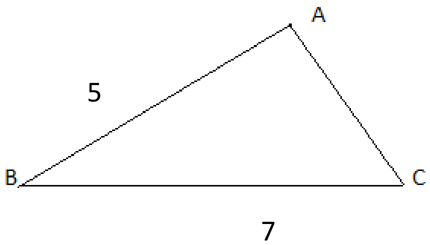
A) Scalene triangle
B) Isosceles triangle
C) Obtuse triangle
D) Equilateral triangle

12. Which option is <u>wrong</u> about the rectangle on the right<u>?</u>

A) |AB|=|DC|
B) |AD|=|BC|
C) |BC|=|BA|
D) |CB|=|DA|



13. Mr. Ahmet placed huts in each corner of the triangular garden and named these huts A, B, and C. Write how to show mathematically the lengths between each hut in the map made by Mr. Ahmet in the space under the triangle below.



…………………………………………………………….

14. I. It has four sides.

II. Opposing side lengths are equal to each other.

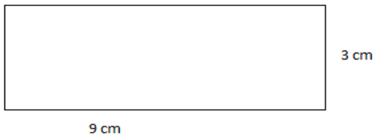
III. All side lengths are equal.

Which of the above are common features of squares and rectangles?

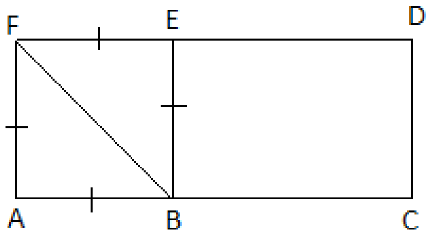
A) Only I B) I and II
C) II and III D) I, II and III

15. How many of the rectangles above can be superimposed to form a square?

A) 1 B) 2
C) 3 D) 4



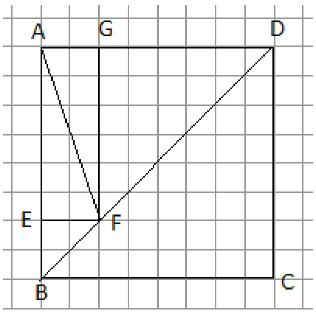
16. .



The names of the triangles, squares, and rectangles shown in the figure above are given. Accordingly, which of the following is <u>wrong</u>?

	Geometric shape	Name
A)	Isosceles triangle	FEB
B)	Triangle	ABF
C)	Square	BCDE
D)	Rectangle	ACDF

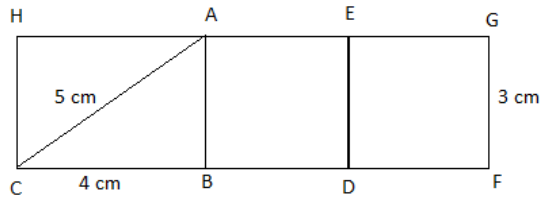
17.



Which of the following figures is not present in the drawing above?

A) Isosceles triangle
B) Scalene triangle
C) Square
D) Equilateral triangle

18.



In the above figure, ABDE and EDFG are squares. IGFI = 3 cm; ICAI = 5 cm and ICBI = 4 cm. Accordingly, what is the length of the HG side?

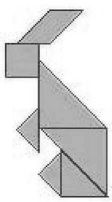
Yukarıdaki şekilde ABDE ve EDFG birer karedir. IGFI =3 cm; ICAI = 5 cm ve ICBI = 4 cm'dir. Buna göre HG kenarı kaç cm'dir?

A) 9 B) 10
C) 11 D) 12

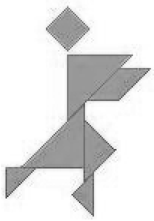
19.

Which of the following shapes <u>cannot be made</u> using the tangram pieces above?

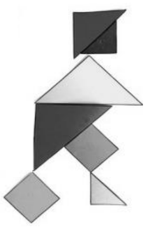
A)



B)



C)



D)