



Achievement Test Development Study Towards Fundamentals of Computer Programming

Gönül ALTAY ¹, Tarık KIŞLA ²

¹ Computer and Instructional Technologies Education, Faculty of Education, Ege University, gonulaltay@gmail.com, <http://orcid.org/0000-0002-7251-4487>

² Doç. Dr., Computer and Instructional Technologies Education, Faculty of Education, Ege University, tarik.kisla@gmail.com, <http://orcid.org/0000-0001-9007-7455>

Received : 06.02.2019

Accepted : 18.11.2019

Doi: [10.17522/balikesirnef.523595](https://doi.org/10.17522/balikesirnef.523595)

Abstract – The aim of this study is to develop a achievement test that will provide reliable and valid results for Vocational and Technical Education's Information Technologies Departure 10th Grade Programming Fundamentals course. For this purpose, firstly, Vocational and Technical Anatolian High School Information Technologies 10th Grade Programming Fundamentals course was taken into consideration, variables and constants, operators, decision structures and loops issues have been determined and after the validity of the test, a 36-question test was prepared. The test was applied to 145 students in the 11th grade of Vocational and Technical Anatolian High School Information Technologies Departure. As a result of item analysis, Programming Fundamentals Achievement Test consisting of 20 multiple choice questions was created. The average item difficulty of the test was found to be 0.53 and the mean item discrimination index was 0.66. The internal consistency value of the test obtained using KR-20 method was 0.88. As a result, a measurement tool has been developed that can be used both in the curriculum and in other scientific studies, in which teachers can use in the classroom assessments and obtain reliable and valid measurement results.

Key words: Achievement test, fundamentals of computer programming, test development, validity, reliability.

Corresponding author: Doç. Dr. Tarık KIŞLA, Computer and Instructional Technologyies Education, Faculty of Education, Ege Universit. This study was made out within the scope of 17-EĞF-002 Scientific Research Project supported by Ege University. In Addition, this study was presented as an oral presentation at the International Computer and Instructional Technologies Symposium (ICITS2018) held on May 02-04, 2018

Summary

The common objectives of all countries in the world are to make the education system functional, to see the impact of education policies on the student, to increase the students' level of success and to improve the quality of education. In order to eliminate the deficiencies in our education system, the Head Council of Education and Morality has made many innovations in both primary and secondary education programs. One of the most interesting of these innovations is that the software and coding courses are given to the students starting from the primary level. Because, a student who can have programming skills will have high mathematical skills, problem solving, critical thinking, and productive skills such as being productive (Akpınar and Altun, 2014; Jenkins, 2002). For this reason, many countries have begun to teach the students beginning from at primary school (Gülmez, 2009). While programming courses in primary curricula focus on algorithm learning, programming courses at high schools and universities focus on learning the commands of one of the programming languages. The programming language is the syntax of what the programmer wants to do from the computer (Van-Roy and Haridi, 2004).

Education of programming using a programming language is a complex process and therefore it is important that students are exposed to different activities on programming. Because an experienced programmer should have many skills and experience (Bennedsen and Caspersen, 2008). A programmer should be able to use the computer effectively, create the program in a file, compile the program, test the produced program, find and correct the errors (Jenkins, 2002). Therefore we need to teach students a programming language by an integrated development environment intended for the development of programs, as well as the development process. Many experimental studies have been conducted to teach students the programming skills in the literature (Eben B. Witherspoon et al., 2018; Chang, 2014; Crescenzi et al., 2012; Marcos Roman-Gonzalez, 2017; Velasquez et al. 2014) and also have developed the achievement tests to explain the relationship between the experimental and control groups (Chiu, 2015; Demir, 2015; Gulmez, 2009; Kalelioglu, 2015; Lye ve Koh, 2015; Tekerek and Altan 2014; Vobornik, 2011; Yilmaz, 2012; Yiğit, 2016). Because to have the knowledge about the individuals' factors of success of, and to make the measurements about their cognitive processes will be a clue in the organization of teaching (Mazman, 2013).

The aim of the achievement test is to determine the degree to which students learn the programming subjects according to the cognitive process dimensions of bloom taxonomy, and to determine on which subjects unsuccessful students are insufficient about (Demir, 2015).

Achievement tests are needed to determine the weaknesses of individuals, to provide the teaching of programming appropriate to their characteristics, and to measure the success of the training (Saraç, 2018). Therefore, developing a valid and reliable measurement tool is important to obtain accurate information about the quality of education.

Many success tests have been developed to measure students' programming skills in the literature review. However, there is no improved achievement test for C # Object-Oriented Programming language console applications for high school students. With this study, it is aimed to develop a valid and reliable achievement test to measure the success of the students in order to realize the targets and behaviors determined on C # Object Oriented Programming language learning for the 10th Grade Programming Basics Course of the Vocational and Technical Education Programming.

75 questions were prepared in accordance with the outcomes determined by the Ministry of National Education for the 10th Grade Programming Basics Course in the testing and development steps of the Vocational and Technical Anatolian High School in order to gain the competencies included in the Frame Teaching Programs applied in Vocational and Technical Education Schools and Institutions. While questions were formed according to the cognitive level of Bloom Taxonomy, numbers of questions according to the subject matters were equal in each cognitive level. For the scope validity of the test, 3 experts were consulted, and some questions were corrected and some questions were removed and the number of questions was decreased to 36. This test was applied to 145 students studying in the 11th Grade of Information Technologies Departure of Vocational and Technical High School. The reason for choosing 11th grade students as the sample group is that they have taken the 10th Grade Programming Basics course and have knowledge about programming subjects. According to the results of the application, the item discrimination index and Item difficulty index of each item have been performed separately.

Item difficulty index (P) shows the correct response rate for each item and can take values between "0" and "1". if the value is approaches 0, the item is the harder, if the value is approaches 1, the item is easy. Generally, it is desirable that item difficulty index is around 0.50 in success tests. The item discrimination index (D) is the degree to which a substance distinguishes students with a high level of achievement with low-achievement students. The item discrimination index can take values between "-1" and "+1". When the item discriminant index approaches 0, the difference between the upper and lower groups of the item is low, and the approach to +1 is high. As a result of the item analysis conducted in this study, 16 items were excluded from the test and 20 Fundamentals of Success Test has been formed consisting

of multiple choice questions. The mean item difficulty index (P) of the test is 0.53, and the mean item discrimination index (D) is 0.66, KR-20 reliability coefficient of the test is also 0.88.

When the item analysis is examined, it is seen that degree of discriminating of the test for students who have low, medium and high success has a high. So, it can be said that successful students who use programming language better understand the programming structures and concepts, and studies should be done for the shortcomings seen in other students.

As a result, it can be said that a valid and reliable achievement test has been developed in order to measure students' success in terms of variables and constants, operators, decision structures and loops.

Programlama Temelleri Dersine Yönelik Başarı Testi Geliştirme Çalışması

Gönül ALTAY ¹, Tarık KIŞLA ²

¹ Bilgisayar ve Öğretim Teknolojileri Eğitimi, Eğitim Fakültesi, Ege Üniversitesi, gonulaltay@gmail.com, <http://orcid.org/0000-0002-7251-4487>

² Doç. Dr., Bilgisayar ve Öğretim Teknolojileri Eğitimi, Eğitim Fakültesi, Ege Üniversitesi, tarik.kisla@gmail.com, <http://orcid.org/0000-0001-9007-7455>

Gönderme Tarihi: 06.02.2019

Kabul Tarihi: 18.11.2019

Doi: [10.17522/balikesirnef.523595](https://doi.org/10.17522/balikesirnef.523595)

Özet – Bu çalışmanın amacı, Mesleki ve Teknik Eğitim Ortaöğretim Programı Bilişim Teknolojileri Alanı 10. Sınıf Programlama temelleri dersi için güvenilir ve geçerli sonuçlar elde etmemizi sağlayacak bir başarı testi geliştirmektir. Bunun için öncelikle Mesleki Ve Teknik Anadolu Lisesi Bilişim Teknolojileri Alanı 10. Sınıf Programlama Temelleri dersi öğretim programı dikkate alınarak değişkenler ve sabitler, operatörler, karar yapıları ve döngüler konularında kazanımlar belirlenmiş ve testin kapsam geçerliliği yapıldıktan sonra 36 soruluk bir test hazırlanmıştır. Oluşturulan bu test, Mesleki Ve Teknik Anadolu Lisesi Bilişim Teknolojileri Alanı 11. Sınıfta öğrenim gören 145 öğrenciye uygulanmıştır. Madde analizi sonucunda 20 adet çoktan seçmeli sorulardan oluşan Programlama Temelleri Başarı Testi oluşturulmuştur. Testin ortalama madde gücü 0.53, ortalama madde ayırt edicilik indeksi ise 0.66 olarak bulunmuştur. KR-20 yöntemi uygulanarak elde edilen testin güvenilirlik iç tutarlılık değeri ise 0.88 olarak elde edilmiştir. Sonuç olarak, hem müfredata uygun hem de diğer bilimsel çalışmalarda öğretmenlerin bu çalışmada yer alan konularda sınıf içindeki değerlendirmelerde kullanabilecekleri, güvenilir ve geçerli ölçüm sonuçları elde edilebilecekleri bir ölçüm aracı geliştirilmiştir.

Anahtar kelimeler: Başarı testi, programlama temelleri, test geliştirme, geçerlik, güvenilirlik.

Sorumlu yazar: Doç Dr. Tarık KIŞLA, Bilgisayar ve Öğretim Teknolojileri Eğitimi, Eğitim Fakültesi, Ege Üniversitesi. Bu çalışma Ege Üniversitesi tarafından desteklenen 17-EĞF-002 Bilimsel Araştırma Projesi kapsamında gerçekleştirilmiştir. Ayrıca bu çalışma, 02-04 Mayıs 2018 tarihinde düzenlenen 12. Uluslararası Bilgisayar ve Öğretim Teknolojileri Sempozyumu' nda (ICITS2018) sözlü bildiri olarak sunulmuştur.

Giriş

21. yüzyılda, insanların tüketici olmalarının yanında üretici de olmaları beklenmektedir. Bu ihtiyaç doğrultusunda, son yıllarda bilgisayar bilimlerinin K-12 ortamlarına entegre edilmesine yönelik artan bir vurgu vardır (Chen ve diğerleri, 2017). Öğrencilere bilgisayar bilimlerinde yetkinliklerini arttırmak amacıyla alanın temel becerilerinden olan programlama becerilerine yönelik etkinlikler düzenlenmektedir (Kalelioğlu, 2015). Alan yazında yer alan

birçok çalışmada, K-12 için bilgisayar programlamanın öğrencilerin problem çözme becerilerini öğrenmek için analitik akıl yürütme stratejilerini geliştirdiği (Liu, Cheng ve Huang, 2011), Gelecekteki kariyer seçimlerinde kalıcılığı etkilediği (Margolis, 2010), onlara sistematik düşünme yeteneği kazandırdığı (Kafai ve Burke, 2013) ortaya konmuştur. Bu nedenle programlama eğitiminin önemi artmış ve birçok ülke programlama derslerini ilk ve ortaöğretim müfredatlarına eklemiştir (Gülmez, 2009).

İlköğretimdeki programlama ders müfredatı, öğrencilerin algoritma öğrenmesine odaklanırken, lise ve üniversitelerdeki programlama dersleri, öğrencilerin programlama dillerinden birinin komutlarını öğrenmesine odaklanır. Programlama dili programcının bilgisayardan ne yapmasını istediğini belirten sözdizimleridir (Van-Roy ve Haridi, 2004). Diğer bir deyişle soyutlama, hata ayıklama ve yineleme gibi bilgisayar bilimi kavramlarının kullanılmasını içeren programlamanın (Lye ve Koh, 2014), problemlerin çözümünde kullanıcı ile bilgisayar arasındaki iletişimi gerçekleştirmek kullandığı, bilgisayarın anlayabileceği bir dildir. Bu Günümüzde en çok kullanılan programlama dilleri Java, C, C++ (Moons ve De Backer, 2013) ve görevlerin yerine getirilmesi için hangi nesnelere ve mesajların aralarında aktarıldığının daha kolay anlaşılmasını sağlayan diğer Nesne Yönelimli Programlama dilleridir (Corral ve diğerleri, 2014).

Bir programlama dilinin öğrenim sürecinde programcının kod yazarken birçok bilgi ve beceriye sahip olması gerekmektedir. İlk olarak seçilen programlama dilinin yazım biçiminin öğrenilmesi gerekmektedir. Her programlama dilinde farklı semboller kullanılmış ve bu sembolere farklı işlevler yüklenmiştir. İkinci olarak kavramsal bilgiye ihtiyaç vardır. Örneğin, “if else” komutu karar yapılarında kullanılmaktadır. Son olarak programcı, problemin çözümüne yönelik bir algoritma geliştirmek zorundadır. Programcılar bu üç tip bilgi ve beceriyi birlikte kullanarak anlamlı ve tam bir program haline getirebilmelidirler. Bu süreç ile ilgili yapılan araştırmalar bireylerin bu bilgileri bir araya getirmekte zorlandıklarını ortaya koymuştur (Jenkins, 2002; Krpan ve diğerleri, 2015). Mazman (2013), programlama dilinin kod okuma, yazım kurallarını bilme ve hataları düzeltme gibi birçok sözel beceri gerektirmesi, algoritma akışını takip etme, döngüler arası geçişleri ve karar yapıları arasındaki ilişkiyi kavrama gibi uzamsal süreçler içermesi sebebiyle bireylerin çalışma belleği kapasitelerinin zorlandığını belirtmiştir. Programlamanın, hem sözel çalışma belleğini hem de görsel-uzamsal çalışma belleğini etkili bir şekilde kullanabilme becerisi olduğunu savunan Mazman (2013), bu durumun programlama eğitimi sürecinde göz önünde bulundurulması gerektiğini vurgulamıştır.

Araştırmalara göre programlama becerileri öğrencilerin uygulama yapmasını sağlayacak yardımcı materyal kullanarak ve deneyimlerle kazandırılabilir (Arabacıoğlu, Bülbül ve Filiz, 2007; Gültekin, 2006; Traynor ve Gibson, 2004). Alanyazında geleneksel öğretim metotlarının dışına çıkarak öğrencilere programlama becerisi kazandırmak için çeşitli deneysel çalışmalar yapılmış (Chang, 2014; Crescenzi ve diğerleri, 2012; Roman-Gonzalez ve diğerleri, 2017; Witherspoon ve diğerleri, 2018; Velasquez ve diğerleri, 2014) ve deney-kontrol grupları arasındaki ilişkiyi açıklamak için de başarı testleri geliştirilmiştir (Chiu, 2015; Demir, 2015; Gülmez, 2009; Lye ve Koh, 2015; Kalelioğlu, 2015; Tekerek ve Altan 2014; Vobornik, 2011; Yılmaz, 2012; Yiğit, 2016). Çünkü bireylerin başarı etkenleri hakkında bilgi sahibi olunmasıyla ve bilişsel süreçlerine ilişkin yapılan ölçümler bireylere verilecek öğretimin düzenlenmesinde de ipucu olacaktır (Mazman, 2013).

Öğrencilerin programlama eğitiminde başarılarını belirleyebilmek amacıyla başarı testleri kullanılmaktadır. Başarı testleri, verilen eğitimin başarılı olup olmadığını ölçmek, belirlenen hedef ve davranışların gerçekleşme düzeyini saptamak ve öğrencilerde ne düzeyde bir değişim olduğunu ortaya çıkarmak için günümüzde sıklıkla kullanılan ölçme araçlarından biridir (Saraç, 2018). Başarı testinin amacı, Bloom taksonomisinin bilişsel süreç boyutlarına göre öğrencilerin işlenen konuları ne derece öğrendiklerini, başarısız öğrencilerin hangi konularda eksik olduklarını belirlemektir (Demir, 2015). Bu araştırma kapsamında geliştirilmiş olan başarı testinin amacı da, öğrencilerin programlama konularını öğrenme derecelerini ve eksik oldukları konuları belirlemektir. Başarı testleri, öğrencilerin eksik ve yanlış öğrenmelerinin ortaya çıkarılmasında sınıf ortamında kolayca kullanılabilir ve sonuçların genellenebilmesini sağlamaktadır (Kan, 2014, Saraç, 2018). Bu tanımlamalara göre de hazırlanan başarı testlerinin verilen eğitimin kazanımlara ulaşip ulaşmadığı hakkında doğru bir bilgi verebilmesi için geçerlilik ve güvenilirliklerinin yüksek olması gerekmektedir. Çünkü her başarı testi puanı bir ölçüm hatası içerir (Gardner, 1989).

Standart başarı testi puanlarının çoğunda hataların büyüklüğü bilinmez ve puanın yorumlanmasında ölçüm hatalarının büyüklük bilgisi kullanılmadığı için puanlama hatası meydana gelir (Gardner, 1989). Bu durum değerlendirme problemleri yaşanmasına sebep olur. Bu kapsamda öğretmenlerin de ölçme ve değerlendirme konularında yeterli bilgi ve beceriye sahip olması gerekmektedir (Daniel ve King, 1998; Gullickson, 1985; Mertler, 1999; Zhang ve Burry-Stock, 2003). Ölçme ve değerlendirme konusu hakkında öğretmenlerin sahip olması gereken hangi amaca hangi test türünün uygun olması gerektiği, dersin hedeflerinin doğru tespit edilebilmesi, kullanılacak testlerin ölçüm sonuçlarında geçerliliğin ve güvenirliliğin nasıl

sağlanacağı, test puanlamasının ve istatistiklerinin doğru yapılabilmesi gibi bilgi ve beceriler yapılan eğitim hakkında doğru veriler elde edilmesini sağlayacaktır (Kubiszyn ve Borich, 1996). Ancak hem yurtiçinde hem de yurt dışında yapılan çalışmalar göstermiştir ki sınıf içi ölçme ve değerlendirme konusunda öğretmenler istenilen düzeyde olmamaktadır (Bıçak ve Çakan, 2004; Daniel ve King, 1998; Güven, 2001; Temel, 1992; Yanpar, 1992). Ayrıca Çakan' ın (2004) ilk ve ortaöğretim öğretmenlerinin ölçme ve değerlendirme uygulamaları ve yeterlilik düzeyleri hakkında yaptığı çalışmada, ortaöğretim kademesindeki öğretmenlerin kendilerini daha yetersiz olarak nitelendirdikleri sonucuna ulaşılmıştır. Bu yüzden geçerlik ve güvenilirlik analizleri yapılmayan veya analiz sonuçları düşük çıkan bir ölçme aracının öğretmenler tarafından kullanılması başta öğrenci olmak üzere öğretim ortamına ve öğretim programına yönelik eksik ya da yanlış kararlar alınmasına sebep olacaktır (Çakan, 2004).

Alan yazında yapılan çalışmalar incelendiğinde programlama becerilerini ölçmeye yönelik olarak ortaokul ya da yükseköğretim düzeyinde birçok başarı testinin geliştirildiği fakat ortaöğretim alanında herhangi bir başarı testinin olmadığı görülmüştür. Benzer ve Erümit (2017) 'in 2008-2017 yılları arasında ülkemizde programlama öğretimine dayalı yapılan 21 adet yüksek lisans ve 8 adet doktora tezi üzerinde yaptıkları araştırmaya göre de başarı testleri veri toplama araçları içerisinde % 24,7 oran ile ölçekten sonra en çok kullanılan ölçme aracıdır. Fakat bu testlerin sadece bir tanesi lise öğrencilerine yöneliktir. Bununla birlikte Şimşek' e (2015) göre, öğretim etkinliklerinin planlanmasında öğrencilerin yaş ve öğrenmeye hazır bulunuşluk düzeyi gibi öğrenci özellikleri önemli bir etkidir. Bu konuya ilişkin yapılan araştırmalar, lise öğrencileri ile diğer öğretim kademesindeki öğrencilerin düzenleme, yineleme, ayrıntılandırma, zaman ve çaba yönetimi, eleştirel düşünme, akran desteği ile yardım arama gibi özelliklerinin farklı olduğunu göstermiştir (Çakmak ve diğerleri, 2008). Bu yüzden diğer öğretim kademeleri için geliştirilen başarı testlerinin lise öğrencileri üzerinde kullanılması uygun görülmemektedir. Alan yazındaki bu bilgiler dikkate alındığında eksikliği doldurmak amacıyla bu çalışmada, Mesleki ve Teknik Eğitim Ortaöğretim Programı Bilişim Teknolojileri Alanı 10. Sınıf Programlama temelleri dersi için C# Nesne Tabanlı Programlama dili öğrenimi üzerine belirlenen hedef ve davranışları gerçekleştirmek ve eğitimin kazanımlara ulaşip ulaşmadığını belirlemek amacıyla öğrencilerin başarısını ölçmeye yönelik güvenilir ve geçerli ölçüm sonuçlarının elde edilebileceği bir başarı testi geliştirmek amaçlanmıştır.

Yöntem

Bu çalışmada Mesleki ve Teknik Eğitim Ortaöğretim Programı Bilişim Teknolojileri Alanı 10. Sınıf Programlama temelleri dersinin öğrencilerin başarısını ölçmek için bir başarı testi geliştirilmiştir. Şekil 1’ de test geliştirme sürecinde kullanılan adımlar verilmiştir.



Şekil 1. Test Geliştirme Sürecinde Kullanılan Adımlar

Test geliştirme adımlarında ilk olarak Mesleki Ve Teknik Anadolu Lisesi Bilişim Teknolojileri Alanı 10. Sınıf Programlama Temelleri dersi için Milli Eğitim Bakanlığı’ nın mesleki ve teknik eğitim veren kurumlarında uygulanan Çerçeve Öğretim Programlarındaki yeterlikleri kazandırma amacıyla belirlemiş olduğu kazanımlara uygun 75 adet soru hazırlanmıştır. Sorular, Bloom Taksonomisi’ nin bilişsel düzey basamaklarına göre oluşturulurken, konular bazında da her bilişsel düzey basamağında eşit sayıda soru olmasına özen gösterilmiştir. Testin kapsam geçerliliği için bilgisayar ve öğretim teknolojileri alanında öğretim üyesi olan 3 uzmandan görüş alınmıştır. Buna göre testteki kimi maddeler düzeltilmiş, kimi maddeler de çıkartılarak madde sayısı 36’ya indirilmiştir. Oluşturulan bu test, Mesleki Ve Teknik Anadolu Lisesi Bilişim Teknolojileri Alanı 11. Sınıfta öğrenim gören 145 öğrenciye uygulanmış, çıkan sonuçlara göre madde ayırt edicilik ve madde güçlüğü analizleri yapılmıştır.

Çalışma Grubu

Çalışma grubu, 2017-2018 eğitim öğretim yılında Manisa, İstanbul, Tokat, Zonguldak, Aydın, Kocaeli, Sakarya, Antalya, Kahramanmaraş illerinde Mesleki Ve Teknik Anadolu Lisesi Bilişim Teknolojileri Alanı 11. Sınıfta öğrenim gören 145 öğrenciden oluşmaktadır. Çalışma grubunun seçiminde araştırma sonuçlarını etkileyecek herhangi bir bilgi aktarımını

önlemek için öğrencilerin farklı il ve ilçelerden seçilmesine özen gösterilmiştir. Tablo 1' de cinsiyet değişkenine göre öğrenci sayıları verilmiştir.

Tablo 1. Cinsiyet Değişkenine Göre Öğrenci Sayıları

	Kız	Erkek
Öğrenci Sayısı	43	102

Yapılan araştırmalarda örneklem büyüklüğü, doğru tahminler yapmak ve gerçek puanlara daha fazla yaklaşmak için önemli bir faktördür ve madde analizi gibi işlemlerin yapılabilmesi için madde sayısının örneklem büyüklüğünün en az yarısı olması önerilmektedir (Açıkgöz, 2015). Bu araştırmada hazırlanan 36 maddelik test, madde sayısının dört katı sayıda örnekleme uygulanmıştır. Bu durumda örneklem sayısının testin geçerlilik ve güvenilirlik analizleri için yeterli olduğu söylenebilir. Örneklem grubu olarak 11. Sınıfta öğrenim gören öğrencilerin seçilmesinin nedeni ise bu öğrencilerin 10. Sınıfta Programlama Temelleri dersini almış ve programlama konuları hakkında bilgi sahibi olmalarıdır.

Verilerin Analizi

Geliştirilen başarı testinin, ölçülmek istenen özelliği başka özelliklerle karıştırmadan ne derece ölçebildiğinin, çıkacak sonuçların hatalardan ne derece arındığının belirlenmesi gerekir. Bu yüzden bir ölçme aracının her bir test maddesinin madde gücü ve ayırt edicilik analizlerinin ayrı ayrı hesaplanması gerekmektedir.

Başarı testi geliştirme birçok aşamadan oluşan sistematik bir işlemdir (Gönen ve diğerleri, 2011). Bu aşamalar test geliştirme süreci alt başlığı altında sunulmuştur.

Test Geliştirme Süreci

Test geliştirme süreci, konu ile ilgili kazanımların belirlenerek madde havuzunun oluşturulması, belirtke tablosundan yararlanılarak testte yer alacak soruların seçilmesi, uzman görüşü alma, testin pilot uygulaması, madde analizinin ve son olarak da test ölçümlerinin güvenilirliğinin yapılması şeklinde sıralanmaktadır (Gönen ve diğerleri, 2011; Karslı ve Ayas, 2013).

1. Adım. Madde Havuzunun Oluşturulması: Test geliştirme sürecinin ilk aşaması kazanımların belirlenerek soruların hazırlanmasıdır. Bu çalışmada, Mesleki Ve Teknik Anadolu Lisesi Bilişim Teknolojileri Alanı 10. Sınıf Programlama Temelleri dersi için Milli Eğitim

Bakanlığı' nın mesleki ve teknik eğitim veren kurumlarında uygulanan Çerçeve Öğretim Programlarındaki yeterlikleri kazandırma amacıyla belirlemiş olduğu kazanımlara uygun 75 adet soru hazırlanmıştır. Test soruları, Programlama Temelleri dersi için hazırlanmış Basit Kodlar ve Kontrol Deyimleri Bireysel Öğrenme Materyalleri' nden seçilmiş ve ek olarak araştırmanın amacına yönelik araştırmacı tarafından geliştirilmiştir.

2. Adım. *Belirtke Tablosunun Hazırlanması*: Başarı testlerinin öğrenci kazanımlarının tümünü kapsayarak öğrencilerin bilişsel becerilerini ölçmesi beklenir (Krathwohl, 2002). Alanyazında öğrencilerin bilişsel becerileri 1956 yılında geliştirilen Bloom Taksonomisi ile ölçülmektedir. Çünkü Bloom Taksonomisi öğrencilerin öğrenim gördükleri ve öğrenmelerini istediğimiz şeylerin sınıflandırılması için bir çerçevedir. Benjamin Bloom ise bu çerçeveyi her biri aynı eğitim hedefini ölçen maddelerin havuzunu oluşturmak ve test öğelerinin değişimini kolaylaştırmak için bir araç olarak tasarlamıştır (Krathwohl, 2002). Bloom'un önerdiği taksonomiye göre, öğrencilerin bilişsel becerileri bilgi, kavrama, uygulama, analiz, sentez ve değerlendirme basamakları ile ölçülmektedir. Buna istinaden hazırlanan sorular belirlenen kazanımlara ve Bloom taksonomisine göre belirtke tablosunda yer almıştır.

Bu çalışmanın içerik konuları değişkenler ve sabitler, operatörler, karar yapıları ve döngülerden oluşmaktadır. Tablo 2' de soru dağılımları, konuların kazanımlarına ve bilişsel becerileri basamaklarına göre soru numaraları yazılarak verilmiştir.

Tablo 2. Uzman Görüşünden Önce Belirtke Tablosu

Konular	Kazanımlar	Bilgi	Kavrama	Uygulama	Analiz
Değişkenler ve Sabitler	Değişken nedir, Nasıl Kullanılır?	1			4
	Değişkenleri İsimlendirme Kuralları		2	33	
	Veri tiplerini tanımlayabilmek	3	34		
	Giriş - Çıkış İşlemleri		18	26	5, 24
Operatörler	Aritmetiksel operatörleri kullanabilmek	6		9	
	İlişkisel operatörleri kullanabilmek	7			
	Mantıksal operatörleri kullanabilmek	11	35, 36		
	İşlem önceliğine göre problem çözebilmek	22		8, 10	
Karar Yapıları	Karar kontrol deyimlerini kullanabilmek. (if-else yapısı)	12, 23, 52	13, 19	29, 31	14, 15, 16, 21, 25
	Karar kontrol deyimlerini ve operatörleri birlikte kullanabilmek. (iç -içe if yapısı)	32,50	17, 20	27, 28, 30	49

	Switch-case yapısı	37, 38, 51	43,46	44,45	47,48
Döngüler	For Döngüsü	41	39, 66	59, 65	60, 63, 64
	While Döngüsü	40, 53	68, 75	69, 70	57, 61
	Do -while Döngüsü	55	67, 71	72, 74	56, 62
	İç içe döngüler	42, 54	73		58

3. Adım. *Uzman Görüşü Alma ve Yazım Denetimi*: Testin kapsam geçerliliği için bilgisayar ve öğretim teknolojileri alanında 3 alan uzmanından görüş alınmıştır. Bu görüşler doğrultusunda, aynı kazanımın birden fazla soru ile ölçülmesi, bazı soruların kazanım ile örtüşmemesi gibi nedenlerden ötürü testteki kimi sorular düzeltilmiş, kimi sorular da çıkartılarak soru sayısı 36'ya düşürülmüştür. Bu süreçte, her konu için belirlenmiş kazanım sayısının fazlalığı soru sayısının çok fazla olmasına, bu durumun oluşmaması için de her kazanım için en az bir sorunun yer almasına sebep olmuştur. Ayrıca kazanımlarda yer alan soru sayısı belirlenirken dersin müfredat programında verilen konulara göre öğrenim süresindeki dağılımı da dikkate alınmıştır. Tablo 3' te uzman görüşüne göre düzenleme yapılmış belirtke tablosu verilmiştir.

Tablo 3. Uzman Görüşünden Sonra Belirtke Tablosu

Konular	Kazanımlar	Bilgi	Kavrama	Uygulama
Değişkenler ve Sabitler	Değişken nedir, Nasıl Kullanılır?	1		4
	Değişkenleri İsimlendirme Kuralları		2	
	Veri tiplerini tanımlayabilmek	3	34	
	Giriş - Çıkış İşlemleri		18	
Operatörler	Aritmetiksel operatörleri kullanabilmek	6		
	İlişkisel operatörleri kullanabilmek	7		
	Mantıksal operatörleri kullanabilmek	11	36	
	İşlem önceliğine göre problem çözebilmek	22		8
Karar Yapıları	Karar kontrol deyimlerini kullanabilmek. (if-else yapısı)		13	14, 21, 31
	Karar kontrol deyimlerini ve operatörleri birlikte kullanabilmek. (iç -içe if yapısı)		17, 50	20, 28, 30

	Switch-case yapısı	38	43	44
	For Döngüsü		39	59, 60
Döngüler	While Döngüsü			57, 61, 70
	Do -while Döngüsü			56, 62, 71, 72
	İç içe döngüler	42		58, 73

4. Adım. *Testin Pilot Uygulaması*: Geliştirilen test, 2017-2018 eğitim öğretim yılında Manisa, İstanbul, Tokat, Zonguldak, Aydın, Kocaeli, Sakarya, Antalya, Kahramanmaraş illerinde Mesleki Ve Teknik Anadolu Lisesi Bilişim Teknolojileri Alanı 11. Sınıfta öğrenim gören 145 öğrenciye uygulanmıştır.

5. Adım. *Madde Güçlük Ve Ayırt Edicilik Analizi*: Madde analizi için bütün öğrencilerin cevap kâğıtları puanlanarak büyükten küçüğe sıralanmıştır. En yüksek puanlıdan başlanarak %27' lik üst grup, en düşük puanlıdan başlanarak %27' lik alt grup olmak üzere iki grup belirlenmiştir. Bu iki grup arasında kalan cevap kâğıtları ise madde analizinde yer almamıştır. Bu durumda 38 öğrenci üst grupta, 37 öğrenci alt grupta yer almıştır. Bütün sorular için “madde ayırt edicilik indeksi (D)” ve “madde güçlük indeksi (P)” hesaplanmıştır. Madde güçlük indeksi (P), “0” ile “1” arasında değerler alabilmekte ve her bir sorunun doğru cevaplanma oranını göstermektedir. Maddenin zor olduğu bulunan değerler sıfıra yaklaşmasıyla, maddenin kolay olduğu ise bulunan değerler bire yaklaşmasıyla belirlenmektedir. Bu yüzden madde güçlük indeksinin 0.5 olduğu orta seviyedeki zorluk ve kolaylıkta olan maddeler tercih edilmekte ve testlerin güvenilirlik düzeyleri daha yüksek bulunmaktadır (Gönen ve diğerleri, 2011). Madde ayırt edicilik indeksi (D) ise, “-1” ile “+1” arasında değerler alabilmektedir ve maddelerin düşük düzeyde başarılı öğrenciler yüksek düzeyde başarılı öğrencileri ayırt etme derecesidir. Bir maddenin alt ve üst grubu ayırt ediciliğinin düşük olduğunu indeksin 0’ a yaklaşması, ayırt ediciliğin yüksek olduğunu ise +1’ e yaklaşması belirtmektedir. Madde analizi sonucunda ayırt ediciliği değerlendirirken şu kriterlere dikkat edilir (Turgut, 1992; aktaran, Ayvacı ve Durmuş, 2015):

- $\underline{D} \geq 0.40$: Madde çok iyi, düzeltilmesi gerekmez.
- $0.40 > \underline{D} \geq 0.30$: Madde iyi, yine de geliştirilebilir.
- $0.30 > \underline{D} \geq 0.20$: Maddenin düzeltilmesi ve geliştirilmesi gerekir.
- $\underline{D} < 0.20$: Madde kullanılmamalıdır veya yeniden düzenlenmelidir.
- $\underline{D} \leq 0$: Maddeler teste dâhil edilemez.

Madde güçlük ve madde ayırt edicilik indeksleri aşağıdaki formüllerle hesaplanmaktadır.

$$P = \frac{(Tü+Ta)}{(Fü+Fa)} \quad D = \frac{(Tü-Ta)}{(Fü \text{ veya } Fa)}$$

Tü: Maddeyi doğru cevaplayan üst gruptaki öğrencilerin sayısı

Ta: Maddeyi doğru cevaplayan alt gruptaki öğrencilerin sayısı

Fü: Üst grup öğrenci sayısı

Fa: Alt grup öğrenci sayısı

6. Adım: *Test Ölçümlerinin Güvenirlik Analizi*: Tüm bilimsel ölçümlerde güvenilirlik konusu, geçerlilik ile birlikte ele alınan temel bir konudur (Gaffney, 1997). Güvenirlik, bir ölçme aracından elde edilmiş ölçümlerin tesadüfi hatalardan arınık olmasını ve sonuçların tekrarlanabilirlik derecesini ifade etmektedir (Turgut, 1995). Buna göre, geliştirilen nihai ölçme aracından elde edilen sonuçların güvenilirlik değeri önem kazanmaktadır. Bu çalışmada güvenilirlik belirlemede en sık kullanılan yöntemlerden biri olan Kuder-Richardson-20 (KR-20) yöntemi kullanılmıştır. KR-20 yöntemi, ölçeğin bütünüyle ve maddelerin birbirleriyle olan iç tutarlılığını tahmin etme üzerine kurudur (Ercan ve Kan, 2004). Bu yöntem, ölçeğin tek boyutlu olduğu, tüm maddelerin aynı değişkeni ölçtüğü durumlarda kullanılır. Ayrıca ölçme aracındaki tüm maddeler aynı ağırlıklı puan ile puanlanmış olmalıdır. Yöntemde veri seti oluşturulurken doğru cevap verilen maddeler “1”, yanlış cevap verilmiş ve boş bırakılmış maddeler “0” ile ifade edilir. KR-20 değeri aşağıdaki formül ile hesaplanmaktadır (Bademci, 2011).

$$KR\ 20 = \frac{K}{K-1} \left(1 - \frac{\sum pq}{S_x^2} \right)$$

K = Ölçme aracındaki soru sayısı

$\sum pq$ = Ölçme aracındaki her maddenin ölçüm varyanslarının toplamı

S_x^2 = Ölçme aracının varyansı

Bir ölçme aracından elde edilen güvenilirlik ölçümlerinin aralığı ise şu şekilde derecelendirilmektedir (Salvucci, Walter, Conley, Fink, ve Saba, 1997).

- KR-20 < 0.5 :Güvenilirlik düşüktür.
- 0.5 < KR-20 < 0.8 :Güvenirlik orta derecedir.
- KR-20 > 0.8 :Güvenirlik yüksektir.

Bulgular ve Yorumlar

Çalışmada testteki bütün maddeler için “madde ayırt edicilik indeksi (D)” ve “madde güçlük indeksi (P)” hesaplanmıştır. Testte bulunan her bir madde için alt ve üst gruptaki öğrencilerin seçtikleri cevap şıkları ile maddelerin D ve P değerleri ise Tablo 4’ te verilmiştir.

Tablo 4. Başarı Testinin Madde Güçlük ve Ayırt Edicilik İndeksleri

SORU	GRUPLAR	SEÇENEKLER				P VE D DEĞERLERİ	
		A	B	C	D	P=	D=
S1	Üst Grup	0	38	0	0	P=	0.85
	Alt Grup	6	26	4	3	D=	0.32
S2	Üst Grup	0	0	0	38	P=	0.72
	Alt Grup	5	6	12	16	D=	0.59
S3	Üst Grup	0	31	0	7	P=	0.63
	Alt Grup	5	16	8	10	D=	0.41
S4	Üst Grup	25	9	1	3	P=	0.49
	Alt Grup	12	6	9	12	D=	0.35
S5	Üst Grup	1	3	34	0	P=	0.55
	Alt Grup	8	12	7	12	D=	0.73
S6	Üst Grup	25	0	3	10	P=	0.45
	Alt Grup	9	4	21	5	D=	0.43
S7	Üst Grup	2	23	12	1	P=	0.47
	Alt Grup	9	12	10	8	D=	0.30
S8	Üst Grup	4	33	0	1	P=	0.52
	Alt Grup	14	6	15	4	D=	0.73
S9	Üst Grup	0	0	36	2	P=	0.67
	Alt Grup	12	10	14	3	D=	0.59
S10	Üst Grup	5	0	32	1	P=	0.55
	Alt Grup	7	15	9	8	D=	0.62
S11	Üst Grup	3	31	4	0	P=	0.48
	Alt Grup	11	5	12	11	D=	0.70
S12	Üst Grup	1	0	0	37	P=	0.59
	Alt Grup	9	12	11	7	D=	0.81
S13	Üst Grup	25	1	11	1	P=	0.45
	Alt Grup	9	11	11	8	D=	0.43
S14	Üst Grup	3	5	1	29	P=	0.43
	Alt Grup	9	17	10	3	D=	0.70
S15	Üst Grup	5	1	0	32	P=	0.61
	Alt Grup	8	8	9	14	D=	0.49
S16	Üst Grup	1	7	25	5	P=	0.51
	Alt Grup	5	11	13	10	D=	0.32
S17	Üst Grup	0	1	35	2	P=	0.53
	Alt Grup	10	14	5	10	D=	0.81
S18	Üst Grup	10	11	6	11	P=	0.23
	Alt Grup	7	11	12	9	D=	0.08
S19	Üst Grup	4	0	33	1	P=	0.61
	Alt Grup	7	11	13	8	D=	0.54
S20	Üst Grup	28	8	1	1	P=	0.55
	Alt Grup	13	7	13	6	D=	0.41
S21	Üst Grup	33	5	0	0	P=	0.59
	Alt Grup	11	8	13	7	D=	0.59

S22	Üst Grup	9	22	4	3	P=	0.36
	Alt Grup	6	5	12	16	D=	0.46
S23	Üst Grup	1	1	35	1	P=	0.63
	Alt Grup	4	13	12	10	D=	0.62
S24	Üst Grup	0	0	0	38	P=	0.59
	Alt Grup	12	8	13	6	D=	0.86
S25	Üst Grup	35	2	0	1	P=	0.57
	Alt Grup	8	14	13	4	D=	0.73
S26	Üst Grup	2	0	32	4	P=	0.51
	Alt Grup	7	16	6	10	D=	0.70
S27	Üst Grup	0	1	1	36	P=	0.49
	Alt Grup	16	13	9	1	D=	0.95
S28	Üst Grup	29	5	3	1	P=	0.45
	Alt Grup	5	15	12	7	D=	0.65
S29	Üst Grup	21	15	2	0	P=	0.27
	Alt Grup	6	5	18	10	D=	0.27
S30	Üst Grup	30	5	2	1	P=	0.59
	Alt Grup	14	11	7	7	D=	0.43
S31	Üst Grup	36	1	1	0	P=	0.65
	Alt Grup	13	10	10	6	D=	0.62
S32	Üst Grup	34	1	1	2	P=	0.61
	Alt Grup	12	8	9	10	D=	0.59
S33	Üst Grup	21	4	13	0	P=	0.41
	Alt Grup	10	10	10	9	D=	0.30
S34	Üst Grup	35	0	2	1	P=	0.56
	Alt Grup	7	12	12	8	D=	0.76
S35	Üst Grup	22	4	6	6	P=	0.32
	Alt Grup	2	12	10	15	D=	0.54
S36	Üst Grup	27	5	5	1	P=	0.41
	Alt Grup	4	10	14	11	D=	0.62

Başarı testinin pilot uygulamasında öğrenciler tüm maddeleri 40 dakika içinde cevaplandırmışlardır. Tablo 4 incelendiğinde, testte bulunan maddelerin ayırt edicilik indekslerinin 0.08 ile 0.95 arasında, madde güçlük indekslerinin de 0.23 ile 0.85 arasında olduğu görülmektedir. Buna göre Tablo 5’ te başarı testinin D değerlerine göre madde sayısı yer almaktadır.

Tablo 5. Başarı Testinin D Değerlerine Göre Madde Sayısı

D değeri	Madde Sayısı
$D \geq 0.40$	29
$0.40 > D \geq 0.30$	5
$0.30 > D \geq 0.20$	1
$D < 0.20$	1

Tablo 5 incelendiğinde D değeri 0.40 ve daha büyük olan 29 madde çok iyi madde özelliği göstermektedir. Bu maddeler herhangi bir değişiklik yapılmadan testte kullanılmış ayrıca başarılı ve başarısız öğrencileri ayırt etme derecesi yüksek olan bir test oluşturmak amacıyla da tüm test maddeleri bu maddelerden seçilmiştir. Testte yer alacak madde seçiminde madde güçlük indeksinin 0.50 civarında olmasına, belirtke tablosundaki her kazanımda en az bir maddenin yer almasına da dikkat edilmiştir. Madde analizi sonucunda 16 madde testten çıkarılmış ve 20 adet çoktan seçmeli sorulardan oluşan Programlama Temelleri Başarı Testi oluşturulmuştur. Testin ortalama madde güçlüğü (P) 0.53, ortalama madde ayırt edicilik indeksi (D) ise 0.66 olarak hesaplanmıştır. Ölçme aracının KR-20 güvenirlik değeri ise 0.88 olarak hesaplanmıştır. Tablo 6’ da analiz sonuçlarından sonra testin yeni soru numaralarına göre düzenlenmiş hali verilmiştir.

Tablo 6. Testin Yeni Soru Numaralarına Göre Düzenlenmiş Hali

Konular	Kazanımlar	Bilgi	Kavrama	Uygulama
Değişkenler ve Sabitler	Değişken nedir, Nasıl Kullanılır?			4
	Değişkenleri İsimlendirme Kuralları		2	
	Veri tiplerini tanımlayabilmek		19	
	Giriş - Çıkış İşlemleri		12	
Operatörler	Aritmetiksel operatörleri kullanabilmek	5		
	İlişkisel operatörleri kullanabilmek	6		
	Mantıksal operatörleri kullanabilmek	8		
	İşlem önceliğine göre problem çözebilmek	15		
Karar Yapıları	Karar kontrol deyimlerini kullanabilmek. (if-else yapısı)			10, 14
	Karar kontrol deyimlerini ve operatörleri birlikte kullanabilmek. (iç-içe if yapısı)		11	17
	Switch-case yapısı		24	25
Döngüler	For Döngüsü		22	
	While Döngüsü			27, 31
	Do -while Döngüsü			26
	İç içe döngüler			28, 36

Tablo 7’ de başarı testi için seçilen soru maddelerinin bilişsel düzey basamaklarına göre dağılımları gösterilmiştir.

Tablo 7. Madde Numaralarının Bilişsel Düzey Basamaklarına Göre Dağılımları

	Bilgi	Kavrama	Uygulama
Madde Numaraları	5, 6, 8, 15	2, 11, 12, 19, 22, 24	4, 10, 14, 17, 26, 27, 28,31,34, 36

Sonuç ve Tartışma

Günlük hayatta karşılaşılan problemlere çözümler üretilmesini sağlayan programlara duyulan ihtiyaç, programlamanın önemini ortaya koymaktadır. Alan yazına göre programlama, öğrenciler tarafından anlaşılması ve başarılması güç bir konudur. Bu yüzden uygulayıcılar, öğrencilerin konuyu kavrayıp kayrayamadıkları hakkında bilgi edinebilmek için nitelikli bir başarı testine ihtiyaç duymaktadır. Bu çalışmada da programlama öğretimi için hem müfredata uygun hem de diğer bilimsel çalışmalarda öğretmenlerin bu çalışmada yer alan konularda sınıf içinde değerlendirme yaparken kullanabilecekleri geçerliği ve güvenilirliği yüksek bir ölçüm aracı geliştirmek amaçlanmıştır. Bu amaçla testin geliştirilme sürecinde bazı basamaklar izlenmiştir. Bu basamaklar; konu ile ilgili kazanımların belirlenerek madde havuzunun oluşturulması, belirtke tablosundan da yararlanılarak testte yer alacak soruların seçilmesi, uzman görüşü alma, testin uygulama süreci ve son olarak madde analizlerinin yapılması şeklindedir.

Test geliştirilirken cevaplanma süresinin bir ders saatini aşmamasına, maddelerin Bloom Taksonomisi' ne göre öğrencilerin bilişsel seviyelerine uygun olmasına dikkat edilmiştir. Testin öğrencilere uygulanmasından sonra her madde için madde ayırt edicilik ve güçlük indeksleri hesaplanmış, ortalama madde güçlüğü (P) 0.53, ortalama madde ayırt edicilik indeksi (D) ise 0.66 olarak bulunmuştur. Testin KR-20 güvenilirlik analizine göre içi tutarlılık değeri de 0.88 olarak hesaplanmıştır. Madde analizleri incelendiğinde testin düşük, orta ve yüksek başarı göstermiş öğrencileri ayırt etme derecesinin yüksek olduğu görülmektedir. Bunun sonucu olarak programlama dilini kullanan başarılı öğrencilerin programlama yapılarını ve kavramlarını daha iyi anladığını, diğer öğrencilerde görülen eksiklikler için çalışmalar yapılması gerektiği söylenebilir.

Bu çalışmada örneklem sayısı toplam madde sayısının dört katıdır. Fakat testin geçerlik ve güvenilirlik çalışmaları yapılırken örneklem sayısının fazla olması çalışmanın amacını daha da kuvvetlendirecektir. Ayrıca araştırmacıya önemli derecede bilgi sağlamış olsa da testin uygulama aşamasında çeşitli zorluklar yaşanmıştır. Bu zorlukların bazılarının üstesinden gelinmiş, bazıları ise kontrol edilmeye çalışılmıştır. İç geçerliliği tehdit eden bu durumlar öğrenci özellikleri, zaman, yer ve uygulayıcının etkisidir.

Sonuç olarak, öğretmenlere Bloom taksonomisine göre düşünme becerilerini temel alarak değerlendirme yapmalarını sağlayan, böylelikle öğrencilerin eleştirel düşünme eğilimlerini ve becerilerini kazanmalarına yardımcı olacak, Mesleki ve Teknik Eğitim Ortaöğretim Programı Bilişim Teknolojileri Alanı 10. Sınıf Programlama temelleri dersinin değişkenler ve sabitler, operatörler, karar yapıları ve döngüler konularında öğrencilerin başarısını ölçmek için geçerliği ve güvenilirliği yüksek ölçme sonuçları elde edilebileceği bir başarı testi geliştirildiği söylenebilir.

Kaynakça

- Açıkgöz, M., ve Karşlı, F. (2015). Alternatif ölçme-değerlendirme yaklaşımları kullanılarak iş ve enerji konusunda geliştirilen başarı testinin geçerlilik ve güvenilirlik analizi. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 4(1), 1-25.
- Akpınar, Y., ve Altun, A. (2014). Bilgi toplumu okullarında programlama eğitimi gereksinimi. *İlköğretim Online*, 13(1).
- Arabacıoğlu, T., Bülbül, H. İ., ve Filiz, A. (2007). Bilgisayar programlama öğretiminde yeni bir yaklaşım. *Akademik bilişim*, 193-197.
- Ayvacı, H. Ş., ve Durmuş, A. (2016). Bir Başarı Testi Geliştirme Çalışması: Isı Ve Sıcaklık Başarı Testi Geçerlik Ve Güvenirlik Araştırması. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 35(1), 87-103.
- Bademci, V. (2011). Kuder-Richardson 20, Cronbach'ın Alfasi, Hoyt'un Varyans Analizi, Genellenirlik Kuramı Ve Ölçüm Güvenirliği Üzerine Bir Çalışma. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, (17), 173-193.
- Bennedsen, J., Caspersen, M. E., ve Kölling, M. (Eds.). (2008). *Reflections on the teaching of programming: Methods and implementations* (Vol. 4821).
- Benzer, A. İ., ve Erümit, A. K. (2017). Programlama Öğretimine Yönelik Lisansüstü Tezlerin İncelenmesi. *Journal of Instructional Technologies & Teacher Education* Vol, 6(3), 99-110.
- Bıçak, B. ve Çakan, M. (2004). *Lise Öğretmenlerinin Sınıf İçi Ölçme ve Değerlendirme Uygulamalarına Dönük Görüşleri*. Milli Eğitim Bakanlığı, Orta Öğretimde Yeniden Yapılanma Sempozyumunda sunulmuş bildiri, Ankara, 20-22.

- Byrne, P., ve Lyons, G. (2001). The effect of student attributes on success in programming. In *Acm sigcse bulletin*, 33(3), 49-52.
- Çakan, M. (2004). Öğretmenlerin ölçme-değerlendirme uygulamaları ve yeterlik düzeyleri: İlk ve ortaöğretim. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 37(2), 99-114.
- Çakmak, E. K., Akgün, Ö. A., Karadeniz, Ş., Büyüköztürk, Ş., ve Demirel, F. (2008). İlköğretim ikinci kademe ve lise öğrencilerinin ders ve sınıf düzeylerine göre öğrenme stratejileri ve güdülenme düzeylerinin belirlenmesi. *Uluslararası İnsan Bilimleri Dergisi*, 5(1), 1-27.
- Chang, C. K. (2014). Effects of using Alice and Scratch in an introductory programming course for corrective instruction. *Journal of Educational Computing Research*, 51(2), 185-204.
- Chen, G., Shen, J., Barth-Cohen, L., Jiang, S., Huang, X., ve Eltoukhy, M. (2017). Assessing elementary students' computational thinking in everyday reasoning and robotics programming. *Computers & Education*, 109, 162-175.
- Chiu, C. F. (2015, April). Introducing Scratch as the fundamental to study app inventor programming. In Learning and Teaching in Computing and Engineering (LaTiCE), 2015 International Conference on, 219-220.
- Corral, J. R., Morgado-Estevez, A., Molina, D. C., Perez-Pena, F., Amaya Rodríguez, C. A., ve Civit Balcells, A. A. (2016). Application of Robot Programming to the Teaching of Object-Oriented Computer Languages. *International Journal of Engineering Education*, 32(4), 1823-1832.
- Crescenzi, P., Malizia, A., Verri, M. C., Díaz, P., ve Aedo, I. (2012). Integrating algorithm visualization video into a first-year algorithm and data structure course. *Journal of Educational Technology & Society*, 15(2), 115.
- Daniel, L. G. & King, D. (1998). A knowledge and use of testing and measurement literac of elementary and secondary teachers. *Journal of Educational Research*, 91 (6), 331-344.
- Demir, F. (2015). *Programlama öğretiminde eğitsel programlama dilinin farklı kullanımlarının programlama başarısı ve kaygısına etkisi*. Doktora Tezi, Atatürk Üniversitesi.
- Ercan, İ., ve İsmet, K. A. N. (2004). Ölçeklerde güvenirlik ve geçerlik. *Uludağ Üniversitesi Tıp Fakültesi Dergisi*, 30(3), 211-216.

- Gaffney, P. V. (1997). A test reliability analysis of an abbreviated version of the pupil control ideology form. *Reports-Evaluative*, 407-422.
- Gardner, E. (1989). Five Common Misuses of Tests. ERIC Digest No. 108.
- Gomes, A., Carmo, L., Bigotte, E., ve Mendes, A. (2006, September). Mathematics and programming problem solving. In 3rd E-Learning Conference–Computer Science Education, 1-5.
- Gönen, S., Kocakaya, S., ve Kocakaya, F. (2011). Dinamik konusunda geçerliliği ve güvenilirliği sağlanmış bir başarı testi geliştirme çalışması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 8(1).
- Gülmez, I. (2009). *Programlama öğretiminde görselleştirme araçlarının kullanımının öğrenci başarı ve motivasyonuna etkisi*. Yüksek Lisans Tezi, Marmara Üniversitesi.
- Gültekin, K. (2006). *Çoklu ortamın programlama başarısı üzerindeki etkisi*. Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü.
- Güven, S. (7-9 Haziran 2001). *Sınıf öğretmenlerinin ölçme ve değerlendirmede kullandıkları yöntem ve tekniklerin belirlenmesi*. 10. Ulusal Eğitim Bilimleri Kongresinde sunulmuş bildiri, Abant İzzet Baysal Üniversitesi, Bolu.
- Jenkins, T. (2002, August). On the difficulty of learning to program. In *Proceedings of the 3rd Annual Conference of the LTSN Centre for Information and Computer Sciences*, Vol. 4, No. 2002, 53-58.
- Kafai, Y., ve Burke, Q. (2013). Computer programming goes back to school. *Phi Delta Kappan*, 95(1), 61–65.
- Kalelioğlu, F. (2015). A new way of teaching programming skills to K-12 students: Code. org. *Computers in Human Behavior*, 52, 200-210.
- Kan, A. (2014). Ölçme Aracı Geliştirme, Eğitimde Ölçme ve Değerlendirme (ed. Satılmış Tekindal), Ankara, Pegem A.
- Karşlı, F., ve Ayas, A. (2013). Fen ve teknoloji dersi öğretmen adaylarının bilimsel süreç becerilerinin ölçülmesine ilişkin bir test geliştirme çalışması. *Türk Fen Eğitimi Dergisi*, 10(2), 66-84.

- Korkmaz, Ö. (2013). Prospective CITE Teachers' Self-efficacy Perceptions on Programming. *Procedia-Social and Behavioral Sciences*, 83, 639-643.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, 41(4), 212-218.
- Krpan, D., Mladenović, S., ve Rosić, M. (2015). Undergraduate programming courses, students' perception and success. *Procedia-Social and Behavioral Sciences*, 174, 3868-3872.
- Kubiszyn, T. ve Borich, G. (1996). Educational testing and measurement: Classroom application and practice (5th ed.). New York: HarperCollins.
- Liu, C. C., Cheng, Y. B., ve Huang, C. W. (2011). The effect of simulation games on the learning of computational problem solving. *Computers & Education*, 57, 1907-1918.
- Lye, S. Y., ve Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12?. *Computers in Human Behavior*, 41, 51-61.
- Mazman, S. G. (2013). Programlama performansını etkileyen faktörlerin bilişsel tabanlı bireysel farklılıklar temelinde modellenmesi. Ankara: Hacettepe Üniversitesi.
- Margolis, J. (2010). Stuck in the shallow end: Education, race, and computing. MIT Press.
- Moons, J., ve De Backer, C. (2013). The design and pilot evaluation of an interactive learning environment for introductory programming influenced by cognitive load theory and constructivism. *Computers & Education*, 60(1), 368-384.
- Román-González, M., Pérez-González, J. C., ve Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in Human Behavior*, 72, 678-691.
- Salvucci, S., Walter, E., Conley, V., Fink, S., ve Saba, M. (1997). Measurement error studies at the National Center for Education Statistics (NCES). Washington D. C.: U. S. Department of Education, 115
- 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*.

- Şimşek, N. (2002). BİG 16 öğrenme biçemleri envanteri. *Eğitim Bilimleri ve Uygulama*,1(1), 33-47.
- Tekerek, M., ve Altan, T. (2014). The effect of scratch environment on student's achievement in teaching algorithm. *World Journal on Educational Technology*, 6(2), 132-138.
- Temel, A. (1991). “ Ortaöğretimde Ölçme ve Değerlendirme Sorunları”.*Yaşadıkça Eğitim Dergisi*, 18, 23-27.
- Traynor, D., ve Gibson, P. (2004, January). Towards the development of a cognitive model of programming: a software engineering approach. In *Proceedings of the 16th Workshop of Psychology of Programming Interest Group*.
- Turgut M.F. (1995). *Eğitimde Ölçme ve değerlendirme metodları*. Ankara: Yargıcı Matbaası
- Witherspoon, E. B., Schunn, C. D., Higashi, R. M., ve Shoop, R. (2018). Attending to structural programming features predicts differences in learning and motivation. *Journal of Computer Assisted Learning*.
- Van-Roy, P., ve Haridi, S. (2004). *Concepts, techniques, and models of computer programming*. MIT press.
- Velasquez, N. F., Fields, D. A., Olsen, D., Martin, T., Shepherd, M. C., Strommer, A., ve Kafai, Y. B. (2014, January). Novice programmers talking about projects: What automated text analysis reveals about online Scratch users' comments. *System Sciences (HICSS), 2014 47th Hawaii International Conference on*,1635-1644.
- Voborník, P. (2011). Teaching algorithms using multimedia tools. *8th International Conference on Efficiency and Responsibility in Education*, 9-10.
- Yanpar, T. (1992). *Ankara ilkokullarındaki ikinci devre öğretmenlerinin öğretmenlik mesleği ve konu alanlarıyla ilgili eğitim ihtiyaçları*. Yayımlanmamış yüksek lisans tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Yılmaz, M. (2012).*C#programlama dersinde, çoklu ortam tasarım ilkelerine göre hazırlanmış materyallerin moodle öğrenme yönetim sistemi üzerinden kullanılmasının yüksek öğrenim öğrencilerinin bilişsel yüklerine ve ders başarılarına etkisi*, Yüksek Lisans Tezi, Başkent Üniversitesi.

Yiğit, M.F. (2016). *Görsel programlama ortamı ile öğretimin öğrencilerin bilgisayar programlamayı öğrenmesine ve programlamaya karşı tutumlarına etkisinin incelenmesi*. Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi.

Yurdugül, H., ve Aşkar, P. (2013). Learning programming, problem solving and gender: A longitudinal study. *Procedia-Social and Behavioral Sciences*, 83, 605-610.