

ÖLÇME VE DEĞERLENDİRMEDE YENİ BİR ANLAYIŞ: BİLGİSAYAR ORTAMINDA BİREYSELLEŞTİRİLMİŞ TESTLER VE KISA CEVAPLI MADDELER

Serdar SAFALI (*)
Bürhan AKPUNAR (**)
Ender ÖZEREN (***)

Öz

Bu çalışmanın amacı, ölçme ve değerlendirmede bireyselliğe verdiği önemle daha çok kullanılmaya başlanılan “bilgisayar ortamında bireyselleştirilmiş testler”in (BOB testlerin) ve Türkiye’de ÖSYM ve Anadolu Üniversitesinin sınavlarda sormaya başladığı “kısa cevaplı sorular”ın temellerini incelemek ve karşılaştırmaktır. Çalışmada literatür tarama yöntemi kullanılmış olup, veriler ilgili yerli ve yabancı literatürden elde edilmiştir. Öncelikle, programın en önemli unsurlarından bir tanesi olan “ölçme ve değerlendirme” değerlendirilmiş, günümüzde halen kullanılmakta olan kâğıt kalem testlerinin kullanımıyla birlikte getirmiş olduğu dezavantajlar irdelenmiş, teknolojinin gelişmesiyle birlikte ölçme ve değerlendirme sürecinde bireyselliğin öneminin artmasıyla birlikte, bilgisayar ve teknolojiyi de içeren geliştirilmiş alternatif ölçme araçlarından bahsedilerek, bilgisayar ve teknolojinin de kullanıldığı bilgisayar destekli alternatif testler ve bilgisayar ortamında bireyselleştirilmiş testlerde kısa cevaplı maddelerin kullanılabilirliği tartışılmıştır. Çalışma sonunda, gelişmiş teknolojinin ve buna bağlı olarak bilgisayarların eğitimin ölçme ve değerlendirme aşamasında kullanım amacı belirtilmiş, eğitimde ölçme ve değerlendirmede yeni arayışların sonucu olan, güvenilirliği ve sınav uygulayıcısına sağladığı avantajlardan dolayı tercih edilebilen kısa cevaplı maddelerin önemi vurgulanmıştır. Yapılan çalışmayla oluşan alan yazındaki boşluğun doldurulması ve bilgisayar ortamında bireyselleştirilmiş testlerde kısa cevaplı sınav sistemiyle ilgili alan yazına katkı sağlanması hedeflenmiştir.

Anahtar Kelimeler: *Bilgisayar Ortamında Bireyselleştirilmiş Testler, Eğitim Programı, Eğitimde Ölçme ve Değerlendirme, Kısa Cevaplı Maddeler, Madde Tepki Kuramı.*

- *) Arş. Gör., Ağrı İbrahim Çeçen Üniversitesi Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme Değerlendirme Ana Bilim Dalı
(e-posta: serdar_safali@hotmail.com), ORCID ID: <https://orcid.org/0000-0001-5916-1106>
- **) Prof. Dr., Harran Üniversitesi Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitim Programları ve Öğretimi Anabilim Dalı
(e-posta: bakpinar23@gmail.com), ORCID ID: <https://orcid.org/0000-0003-3509-0475>
- ***) Okutman, Dicle Üniversitesi, Çermik Meslek Yüksekokulu
(e-posta: enderozeren@gmail.com), ORCID ID: <https://orcid.org/0000-0003-4605-8120>

A New Approach to Measurement and Evaluation: Individualized Tests and Short-Responsible Items in Computer

Abstract

The purpose of this study is; to examine and compare the bases of "computerized individualized tests" (BOB tests) which began to be used more with the importance given to individuality by measurement and evaluation and "short answer questions" which begin to ask the examination of OSYM and Anadolu University in Turkey. Literature review model was used in the study and the data were obtained from related domestic and foreign literature. First of all, one of the most important elements of the program, "measurement and evaluation" was assessed and the disadvantages brought about by the use of paper pencil tests which are still in use today were investigated. With the development of technology, besides the increasing importance of individualism in the measurement and evaluation process, computer aided alternative tests used in this computer and technology, and the availability of short answer items in computerized individualized tests are discussed. At the end of the study, the purpose of using the advanced technology and accordingly the education of the computers in the measurement and evaluation phase is stated and the importance of short answer materials which can be preferred because of the advantages that are provided by the reliability and the practitioner, which are the result of new quests in measurement and evaluation in education was emphasized. It is aimed to fill the gap in the field created by the work done and to contribute to the writing of the field related to short answer test system in computerized individualized tests.

Keywords: *Computerized Adaptive Tests, Curriculum, Short Answered Questions, Substance Response Theory, Measurement and Evaluation.*

Giriş

Bug n eđitimin ve eđitimi bir insanın sosyal ve ekonomik b y menin itici g c  olduđu bir s re yařanmaktadır (Ornstein ve Hunkins, 2009). Eđitim toplumun sahip olduđu deđerlerin ve k lt r n gelecek nesillere aktarımını sađlayan, bireyleri ve buna bađlı olarak toplumu geleceđe hazırlayan bir s retir. Birey bu s rete sahip olduđu bilgilere yenilerini ekler, yeteneklerini keřfederek kendini geliřtirir; yaratıcılık, eleřtirel d ř nme, diđer bireylerle sosyal etkileřime girebilme, iřbirliđi ile  đrenme gibi becerileri kazanır (Kandemir, 2016). Eđitim s recinde bu beceri ve davranıřların kazanılmasında eđitim programların  nemi ok b y kt r.

Posner (1995) ve Ert rk (2013) eđitim programını, tasarlanmış durumların gerekleřtirilmesi iin yapılması gereken eđitim faaliyetlerinin sıra ve zamanını belirten  đretim ve deđerlendirme s recine karar vermeyi sađlayan bir s re řeklinde tanımlarken, Uřun'a g re (2012, s.2) ise eđitim programı; belirli bir taksonomi ierisinde d zenlenen amalar, bu amaları gerekleřtirmek iin gerekli řartlar, uygulama s recindeki metotlar, s re sonucunda deđerlendirmeyi kapsayan bir faaliyetler b t n d r. Hunkins ve Ornstein'e g re

(2014, s.10) eğitim programı istenilen hedeflere veya amaçlara ulaşmak için stratejiler içeren yazılı bir belge, eğitilecek kişiler için öğrenme fırsatları içeren eylem planıdır.

Eğitim programı sürekli olarak birbirileri ile etkileşim içinde olan aktif dört yapıdan oluşmaktadır. Bunlar hedef, içerik, uygulama ve ölçme-değerlendirme basamaklarıdır (Demirel, 2012; Kandemir, 2016). Eğitim programının en son ögesi ölçme ve değerlendirmedir. Ölçme ve değerlendirme, hedefler doğrultusunda bireylerde meydana gelmesi istenilen istendik davranışların kazanılıp kazanılmadığını, kazanılma düzeyini, kazanılmadıysa nedenlerini ve kazandırabilmek için eğitimde yapılması gerekenleri belirleyen doğrudan veya dolaylı sınav durumları ve yargıya varma işlemleridir (Ornstein ve Hunkins, 2009, s.235; Sönmez, 2012, s.451). Bu süreçte çeşitli ölçme araçları kullanılarak öğrenenin hedeflere ulaşma dereceleri ve öğretmenin eğitim sürecinde uyguladığı ders işleme sürecinin etkililiği hakkında dönüt sağlanır, bu dönütlere anlam verilir, bu dönütlere istendik hedefleri karşılama ve belirli bir anlamda olup olmama bakımından yorumlanır (Marsh, 2004; Özçelik, 2010, s.231). Bir etkinlik olarak ölçme; bir durumun açıklaması, belirlenen durumun planlanan ve planlanmamış etkilerinin değerlendirilmesinden saptanan olayların sarmal bir dizisi olarak kavramsallaştırılabilir (Preston ve Aikman, 2005). Eğitim programının son ögesi olan ölçme ve değerlendirme, eğitim sürecinde “ne kadar öğrettik” sorusunun cevabını vermektedir. Ölçme ve değerlendirme düzey belirleme odaklı, süreç odaklı veya sonuç odaklı olarak sadece öğrenenlerin hedefe ulaşma düzeyleri hakkında öğreticiye bilgi vermez, öğretim programının geliştirilmesinde de önemli bir yere sahiptir (Koç, 2015).

Ölçme ve değerlendirme etkinlikleri eğitim programının eksiklerini belirlemede ve gidermede en önemli tamamlayıcı unsurlarından birisidir. Ancak hem öğrenenin başarısını yükseltmek hem de eğitim programının eksiklerini belirleyip düzeltmek için ölçme-değerlendirme çalışmalarının doğru yapılması gerekmektedir (Physical Education Curriculum Guide: Primary/Elementary, 2011). Buna göre seçilecek olan doğru ölçme aracının programda belirtilmiş olan ve öğrenenin kazanması istenilen hedefler ve davranışları ölçer nitelikte olması gerekmektedir. Ancak bu durum sağlandığı zaman sürece uygun ve doğru bir ölçme-değerlendirme yapılmış olur (Koç, 2015).

Programın Sürekliliğini Sağlayan Unsur “Değerlendirme”

Bireyler zekâ, tutum, yetenek, beceri, kişilik yapısı ve öğrenme biçimi gibi temel özelliklerde farklılık gösterirler. Bu nedenle günümüzde bireysel eğitime önem verilmekte, her bireyin farklı olması eğitimde çeşitliliği ortaya çıkarmaktadır. Eğitimde program geliştirme ve program geliştirmenin bir aşaması olan değerlendirmenin başarılı olabilmesi için her bir öğrencinin bireysel farklılığı göz önünde bulundurulmalıdır. Günümüzde teknoloji ve buna bağlı olarak bilgisayar kullanımı her bir öğrencinin sahip olduğu farklı özelliklere göre göz değerlendirme yapılabilmesini mümkün kılmaktadır.

Değerlendirme, kökenini latince “valor” (güçlü olmak, değer olmak) kelimesinden almakta ve İngilizcedeki “evaluation” kelimesinin karşılığı olan “durum tesbiti”, “değer

verme”, “h k m verme”, “not verme” ve derecelendirme anlamlarına gelmektedir (Wotowa ve Thierau, 1990, akt.  ltanır, 2016). Reischmann’a g re (2006’dan akt.  ltanır, 2016) deęerlendirme d ş ncesi, “y ntemsel kavrayıř, eęitim alanında etki kontrol , y netim ve yansıtımlarla uygulama tedbirlerinin daha iyi anlařılması ve biimlendirilmesi iin s re ve sonuların deęerlendirilmesidir” Eęitim programının deęerlendirilmesinin temel amacı, programın uygulanmasında g r len aksaklıkların kaynaęı belirlenerek giderilmesi, dersin daha etkin hale getirilmesi,  ğrenenin ulařılması istenen hedeflere ulařma d zeyini belirleme, okul sistemi ve  ğretmenler hakkında karar verme, dięer bir deyiřle kalite kontrol  iin gerekli verileri toplama ve bu veriler doęrultusunda kararlar almadır (Akpınar, 2015).

Ert rk’e g re (1998, akt. Akpınar, 2015), eęitim programını deęerlendirme odak noktasına g re farklılık g stermektedir. Deęerlendirmede odak noktasına g re incelenen program deęerlendirme yaklařımları; program tasarısına bakarak deęerlendirme, bařarıya bakarak deęerlendirme, eriřiye bakarak deęerlendirme,  ğrenmeye bakarak deęerlendirme ve  rt ne bakarak deęerlendirmedir.

Deęerlendirme, eęitmenlerin geri bildirim bilgilerini toplamalarının t m yollarını belirten genel bir terimdir. Genel olarak, deęerlendirme prosed rleri sınıfta eřitli iřlevlere hizmet ederler; oęunlukla biimlendirici deęerlendirme,  zet deęerlendirme ve  z deęerlendirme iin kullanılır (Seng, Parsons, Hinson ve Sardo-Brown, 2003). Biimlendirici deęerlendirme eęitmenler ve  ğrenciler iin geici bilgi saęlar (Wang, 2007; Crisp ve Ward, 2008). Buna ek olarak,  ğretmenler aynı zamanda  ğretmenin ne kadar iyi  ğrettięini ve  ğrencilerin  ğretim sonunda  ğrendiklerini (Black ve Wiliam, 1998) gerekleřtirmek iin  zet deęerlendirme kullanmaktadır. Son olarak,  ğrenciler kendi yeteneklerini anlamaya alıřtıklarında, kendilerini de deęerlendirebilmek iin  z deęerlendirmeler yapabilirler.

Sadece s re sonunda yapılan sınavlara odaklanan “bařarıya odaklı deęerlendirme” veya programa giriřteki  ğrenci bařarıları ile programdan ıkıřtaki  ğrenci bařarıları arasındaki farka odaklanan “eriřiye bakarak deęerlendirme”, deęerlendirme yapılabilmesi iin bir ara kullanılarak veri toplanılmasını gerektirmektedir. G n m zde deęerlendirme yapabilmek iin kullanılan aralardan bir tanesi de birok bilgi ve yetenek alanlarında uygulanabilen, biliřsel  ğrenme kademelerinin her bir d zeyini  lebilen her eęitim d zeyine uygun, kısa cevaplı maddelerdir (Turgut ve Baykul, 2010).

Programın Uygulanma S recinde  lme ve Deęerlendirme İhtiyacı

Eęitim, eęitimin uygulanmasında izlenecek yol olan eęitim programları ve buna baęlı olarak  ğretim programları g n m z yařam Őartlarını yakalamak ve d nyadaki geliřmelere ayak uydurabilmek iin devamlı yenilenen dinamik bir yapıya sahiptir. Eęitimin canlılıęını koruma, yařam standartlarını yakalayabilme ancak programın deęerlendirilmesi ile m mk n olur (Ornstein ve Hunkins, 2009).

Ölçme ve değerlendirmeye ihtiyaç duyulmasının en önemli nedenlerinden bir tanesi de öğretici ve öğrenenin eğitim öğretim sürecinde işlenen ders sürecini, ders sürecinde uygulanan öğrenme etkinliklerini inceleme ve eksiklerini görebilmelerini sağlamasıdır (Black, Harrison, Lee, Marshall ve William, 2003). Ölçme ve değerlendirme sonucunda elde edilen verilere göre, öğretim etkinliklerinin amaçlarının gerçekleşme düzeyi belirlenir, okul içinde veya okul dışında olsun eğitim sürecinin işlevselliği, ders sürecinde uygulanan etkinlikler, ders anlatımı, öğrencilerin etkinliği, öğretmenin anlatış biçimi, eğitim verilen yerin sahip olduğu fiziki şartlar gözden geçirilir, eksiklikler giderilir (Semerci, 2008).

Ölçme değerlendirmeye ihtiyaç duyulmasının diğer bir nedeni ise eğitim sürecinde meydana gelen aksaklıkların giderilmesi ve öğrenenin öğrenmede yetersiz kaldığı noktaların tespit edilmesidir. Ölçme değerlendirme kullanılarak eğitim sürecinde öğrenenin öğrenmede başarısız olduğu yerler tespit edilir ve eksikler giderilebilir.

Ölçme ve değerlendirme sadece eksikliklerin belirlenmesinde değil, öğrenenin hazır-bulunmuşluk düzeyini belirlemede, ilgi ve yeteneklerinin saptanmasında, başarısını izleme, belirleme ve gerekli yerlerle paylaşma için de büyük bir ihtiyaçtır. Bireyin sahip olduğu yeteneklerin belirlenmesinde, sahip olduğu potansiyelin farkına varmasında ve o yeteneklere göre hedeflerini belirlemesinde kullanılan en önemli araçlar ölçme araçlarıdır.

Günümüzde eğitimin canlılığını korumasında, yaşam standartlarını yakalayabilmede, bireyin sahip olduğu özelliklerinin farkına varabilmesinde önemli yeri olan ölçme ve değerlendirme sürecinde farklı sorunlarla karşılaşmaktadır.

Çağdaş Program Bağlamında Ölçme Değerlendirme Sorunları

Ekonomi, teknoloji, siyaset ve kültür gibi temel alanlarda önemli gelişmeler görülmesiyle birlikte, eğitim kurumları ve bu eğitim kurumlarında uygulanan eğitim programları da bu gelişimden etkilenmektedir (Gelen ve Beyazid, 2007). Ölçme ve değerlendirme; eğitim programının canlılığını koruyan, günlük yaşamdaki gelişmelerin takip edilmesi ve teknolojinin gerisinde kalmamasını sağlayan en önemli yapılardan biri olduğu için önemi büyüktür. Öğretmenlerin eğitim sürecinde karşılaştıkları sorunları daha kolay çözebilmesi için eğitim sürecinde hiçbir öğrencinin geriye bırakılmadan ve her bir öğrencinin farkında olunarak, eğitim programımızda değişiklikler yapılmaktadır. Ancak genellikle atılan bu adımlardan olumlu dönütler alınamamaktadır. Bu önemli süreçte karşılaşılabilecek problemler; eğitim programının işlevselliği, ulaşılması gereken hedeflere ulaşılma düzeyi ve öğrenenin öğretilmek istenen davranışı ne kadar öğrendiği hakkında yanlış bilgilere sahip olmadır (Thorndike ve Thorndike-Christ, 2010). Karşılaşılabilecek bu problemler; öğretici kaynaklı, kullanılan ölçme aracı kaynaklı ve ölçme değerlendirmenin uygulandığı ortam kaynaklı olabilir.

Gelen ve Beyazid'e göre (2007) meydana gelebilecek problemlerin nedenlerinden bir tanesi; program geliştirme çalışmalarının kısa vadeli çalışmalar olması ve eğitim sorunlarının geleneksel yaklaşımlarla çözülmeye çalışılmasıdır. Türk eğitim sisteminde

yerleşik eğitim kültürünün oluşmaması, program geliřtirmenin masa başında yapılması veya geliřtirilmeye çalışılması, çalışmaların geleneksel yöntemlerle yapılması, dinamik bir yapıya sahip olan eğitim sistemi ve eğitim sürecindeki sorunları gidermede yetersiz kalmıştır. Eğitim sorunlarının çözümünün en geçerli ve güvenilir yolu bu konuyla ilgili yapılan akademik çalışmalar ve arařtırmalardır (Karakaya, 2004). Kandemir'e göre (2016), eğitim programı geliřtirmede ve buna baęlı olarak ölçme ve deęerlendirmede karşılaşılan dięer bir problem ise eğitim programının deęiřtirilme zamanı ve uygulama sürecidir. Türkiye'de program deęiřiklikleri ön deneme yapılmadan uygulama için gerekli altyapı oluşturulmadan uygulanmaktadır. Bu durum eğitim sürecinde karışıklığa neden olmakta, ölçme ve deęerlendirmede güvenilirliği düşürmektedir.

Ölçme ve deęerlendirme sürecinde meydana gelebilecek olası problemlerden bir tanesi de ölçme uygulayıcısından kaynaklanan sorunlardır. Eğitim sürecinde ölçme aracını uygulayıcı görevini üstlenen öğretmenler, öğrencilerin hazırbulunuşluk düzeyini belirlemek, öğrencilerin sahip olduęu yeteneklerin farkına varmasını sağlamak, öğretim sürecinde meydana gelen eksikleri belirlemek ve bu eksikleri gidermek için ölçme deęerlendirmeye başvurur; sonuçları öğrencilerle, velilerle ve okul yönetimiyle paylaşırlar. Bu nedenle öğretmenlere eğitim hedeflerinin ulaşması ve eğitim programının uygulanabilmesi için önemli işler düşmektedir (Semerci, 2008). Öğretmenler maksimum etkinliği elde etmek için öğrencilerin öğrenme ortamını incelerken ölçme ve deęerlendirmeyi nasıl uygun bir şekilde yorumlayacaklarını ve kullanacaklarını bilmelidir (Daniel ve King, 1998). Ancak günümüzdeki öğretmenlerin büyük kısmı yukarıda sayılan vasıfların bir kısmına sahip değillerdir.

Öğretmenin hata payının en aza indirildięi, geçerliği ve güvenilirliği yüksek bir sınav yapabilmesi için o alanda yeterli bilgiye sahip, alanında uzman ve deneyimli olması gerekmektedir. Öğretmenin sahip olduęu yeterlilik seviyesi ve deneyim ile eğitim sürecinin kalitesi arttırılır, bu sürece baęlı olarak uygulamış olduęu ölçme-deęerlendirme süreci de amaca hizmet eder (Sabancı ve Yazıcı, 2017). Ölçme deęerlendirme ile sadece eğitimin eksik yönleri deęil kalitesi de ölçülmüş olur. Bu nedenle alanında uzman olma ve yeterlilik öğretmenin öğretim sürecini, ölçme ve deęerlendirmesini ve öğrencilerin hedefe ulaşma düzeylerini etkileyen faktördür. Öğretmenin sahip olduęu alan hakkında yeterli olma ve deneyim ise ders sürecinde sınıftaki etkililiğini belirleyen en önemli motivasyonel etkendir (Tschannen-Moran ve Hoy, 2001; Pendergast, Garvis ve Keogh, 2011). Tschannen-Moran ve Hoy (2001) öğretmenlerin test ölçüm bilgilerinin çoğunun deneme yanılma yöntemiyle iş deneyimi üzerinden elde ettiğini belirtmektedir. Öğretmen adayları üniversitedeki eğitim sürecinde ölçme ve deęerlendirme üzerinde yeterli eğitimi alamamakta, öğretmen olarak atandıkları zaman öğretmen vasfını taşıyacak gerekli yetilere sahip olarak göreve başlayamamaktadır. Bu durum ölçme ve deęerlendirme sürecinde güvenilirlik yapısını olumsuz yönde etkilemektedir.

Ölçme ve deęerlendirme sürecinde karşılaşılan sorunların kaynaklarından bir tanesi de ölçme aracı kullanımınıdır. Kullanılan ölçme aracında bulunması gereken özellikler güvenilirlik, geçerlik ve kullanılabilirliktir.

Güvenilir ölçme aracı, ölçme sonuçlarının hatalardan arınık olması anlamına gelmektedir (The American Educational Research Association, The National Council on Measurement in Education, 1997). Güvenilir ölçme aracı, kendi içerisinde tutarlı, kararlı ve duyarlı olmalı, ölçüm sonucu dışardan bir müdahale olmadığı sürece aynı sonucu vermeli ve ölçmek istenilen özellik hakkında gerçek durumu vermelidir (Devellis, 2014).

Ölçme aracında aranılan diğer bir özellik ise geçerliktir. Geçerlik bir ölçme aracının ölçmek istediği özelliği ölçebilme derecesidir (Tekin, 2010). Güvenirlik geçerlik için ön şarttır. Bir ölçme aracı güvenilir olmadığı sürece geçerliğinden bahsedilemez (Thorndike ve Thorndike-Christ, 2010).

Ölçme aracında aranan diğer bir özellik ise kullanılılıktır. Kullanışlı bir test, ekonomik, puanlaması kolay, anlaşılır ve kullanılmalıdır. Bu özelliklere bakılırken amaca hizmet etmesi de göz ardı edilmemelidir (Thorndike ve Thorndike-Christ, 2010).

Eğitimde Ölçme ve Değerlendirme Alanında Uygulanan Yeni Arayışlar

Dünyada bilim, teknoloji, siyaset gibi alanlar hızlı gelişim göstermekte, ülkeler bu gelişimi yakalayabilmek için mevcut düzende değişikliklere gitmektedirler. Bu gelişim sürecini yakalayabilmenin en önemli şartlarından biri de eğitim ve eğitim sistemini değiştirmek ya da güncellemek için yeni arayışlara yönelmektir. Türkiye’de PISA, TIMMS gibi sınavlarda alınan kötü sonuçlar nedeniyle eğitim programı ve programın bir parçası olan ölçme ve değerlendirmede değişim kaçınılmaz hale gelmiştir (Arslan ve Özpinar, 2008).

Türkiye’de yapılandırmacı yaklaşıma göre oluşturulan yeni programda, değerlendirme; öğrencilerin fizyolojik ve psikolojik gelişimine destek olması amacıyla eksiklerin belirlenmesi ve giderilmesi olarak belirtilmektedir. Değerlendirmenin sadece ürün odaklı olmaması gerektiği, sürecin de dikkate alınması gerektiği vurgulanmaktadır. Bu da ancak birden fazla ölçme aracının birlikte kullanımı ile sağlanabilir.

2004 yılı öncesi Türkiye’de uygulanan eğitim programı, temelini idealizm ve realizm-den almakta ve merkezde öğretmen bulunmaktaydı. Bu modelde, öğrenci sadece kayıt cihazı görevini üstlenmekte, öğretmenin verdiği bilgilerle yetinmekte ve bu bilgileri günlük hayatta kullanımda sorunlar yaşamaktaydı (Akçadağ, 2010). 2005 yılında uygulamaya konulan yeni eğitim programı, yapılandırmacı eğitim sistemini temel almış olup ilerlemecilik felsefesine dayanmakta, merkeze öğrenciyi almakta, öğrencinin öğrenmesi istenilen davranışları yaşayarak öğrenmesi amaçlanmakta ve öğretmene sadece aracı veya rehber rolü verilmektedir (Tuncer ve Özeren, 2015). Öğrenciye soran, araştıran ve sorgulayan rolü verildiği için, ölçme ve değerlendirme de eğitim sürecinde önem kazanmaktadır. Bu nedenle yeni eğitim programında alternatif ölçme araçlarının kullanımının önemi daha da artmış durumdadır.

Sadece Türkiye değil, diğer ülkelerde de program kalitesini artırma, öğrencilerin kazanılması istenilen hedeflere ulaşma düzeyini yükseltme ve başarısını arttırmak için ça-

liřmalar yapmakta, uluslararası d zeyini belirleyebilmek iin sınavlar d zenlemektedir. D zenlenen bu sınavlar  lkelerin eđitim alanında hangi d zeyde olduklarını g rmelerini, giderilmesi gereken eksikliklerin ve alınması gereken tedbirlerin belirlenmesini sađlamaktadır. D nya apında bilinen en pop ler iki sınav PISA ve TIMSS'dir (Kelly, 2002; PISA T rkiye Resmi Web Sitesi, 2017). TIMSS sınavı  đrencilerin ne bildiđini  lmeye, PISA ise  đrencilerin sahip olduđu bilgileri g ncel hayatta uygulamalarını  lmeye odaklıdır (Berber, 2015).

PISA en az 7 yıl eđitim g rm ş olması Őartıyla 15 yařındaki  đrencilere matematik ve fen alanında uygulanır. PISA'nın amacı bir yandan  lkelerin eđitim sistemlerinin durumunu ortaya ıkarmak diđer yandan ise  lkelerin uluslararası d zeyde kaliteli insan g c  olarak konumunu belirlemek ve  đrencilerin okulda edindikleri bilgileri kullanarak bilinmeyen hakkında tahminde bulunma ve bilgilerini okul dıřında uygulayıp uygulamadığını arařtırmaktır (PISA T rkiye Resmi Web Sitesi, 2017).

TIMSS (Trends in International Mathematics and Science Study) ise uluslararası d zeyde 4 yılda bir yapılan ve 4. ve 8. sınıf  đrencilerinin matematik ve fen alanlarında kazandıkları bilgi ve becerileri deđerlendiren bir sınavdır. Ayrıca  lkeler bazında;  lkelerin diđer  lkelere karřı durumlarını belirler (TIMSS 2015 T rkiye, 2017).

Yukarıda da bahsedildiđi gibi eđitim s recinde bařarıya ulařılabilmek iin d nya genelinde yeni y nelimler ve alıřmalar yapılmaktadır. Yapılan bu alıřmaların hepsinin temelinde birey ve bireyin sahip olduđu kiřisel  zellikler bulmaktadır Artık eđitim s recinde gerek anlamda g venilir bir  lme ve deđerlendirme yapılması iin bireyin sahip olduđu kiřisel  zellikler g z  n nde bulundurulmalıdır. Bu g venirliđi ve geerliđi y ksek bireysel farklılıđı g z  n nde bulunduran  lme ve deđerlendirme, teknolojinin de yardımıyla bilgisayar ortamında bireyselleřtirilmiř testler (Computerized Adaptive Test, CAT) uygulanmaktadır (Thompson ve Weiss, 2011).

Bireyselliđe  nem Veren Testler: Bilgisayar Ortamında Bireyselleřtirilmiř Testler (BOB)

G n m zde halen kullanılmakta olan ve bireyin belirli bir alanda sahip olduđu ilgi ve yeteneklerini belirlemek iin kullanılan kâđıt-kalem testlerinin yerini bilgisayar ortamında hazırlanmıř olan testler almaya bařlamıřtır.  lkemizde halen bir ok eđitim kurumunda kullanılmakta olan kađıt-kalem testlerinde bireylere  lilmek istenen  zellikle ilgili, aynı sayıda sorular sorulur ve belirli bir s rede bu soruların cevaplandırılması beklenir. Daha sonra toplanan sınav kâđıtlarında, bireyin sahip olduđu ilgi ve beceri d zeylerini belirlemeden ve dođru cevaplandırılan soruların niteliklerine bakılmadan sadece verdiđi dođru cevap sayısına g re puanlama yapılır. Bu sınavlarda b t n  đrenciler aynı seviyede olduđu g z  n nde bulundurularak aynı sorulara tabii olurlar. Kâđıt-kalem testlerinde bireye  zg  sınavlar hazırlanamadıđı iin bu testlerle bireyin sahip olduđu  zellikleri belirlemek ok zordur. Kađıt –kalem testlerinde oluřturulan maddeler genellikle orta d zeye sahip  đrencilerin seviyelerine g re hazırlanmıř testlerdir. Bu nedenle oluřturulan

bu maddeleri kullanıldığı ölçmede yüksek veya düşük seviyeli bir öğrencinin ölçüm güvenilirliği ve geçerliği de düşük olacaktır. Kağıt-kalem kullanılarak yapılan testlerin diğer bir zayıf noktası ise süre kısıtlamasıdır. Öğrencilerden verilen soruları belirli sürede cevaplaması beklenilmektedir ki bu durum, cevaplama hızı yavaş olan bir öğrenci için yetenek düzeyinin ve sahip olduğu bilginin asıl değerinden daha düşük çıkmasına neden olacaktır (Weiss, 2004; Thompson ve Weis 2011).

Geleneksel kağıt-kalem testlerinin hazırlanması, uygulanması, geçerliği ve güvenilirliği düşük olması nedeniyle, eğitimde ölçme ve değerlendirme alanında teknolojinin de gelişmesiyle birlikte bilgisayardan faydalanılmaya çalışılmaktadır. Bu amaçla bireyin sahip olduğu özelliklerini, yeteneklerini ve bilgi düzeyini belirlemek için bilgisayar ve teknolojiyi de içeren alternatif ölçme araçları geliştirilmiştir. Geleneksel yöntemlerle hazırlanmış olan kağıt-kalem testlerinin yerini bilgisayar destekli alternatif testler (Computer-Based Tests, CBT) ve bireyin sahip olduğu zeka, yetenek ve ilgilerini de göz önünde bulunduran bilgisayar ortamında bireyselleştirilmiş testler-BOB testleri (Computerized Adaptive Test, CAT) almıştır (McBride ve Sz Martin, 1983).

Bilgisayar teknolojisinin eğitimle bütünleştirilmesi, 21. yüzyılda öğretme ve öğrenmeyi büyük ölçüde değiştirmiştir. Teknoloji, kullanan okul ortamlarının sayısı gittikçe artan bir şekilde hem öğrencilere güç kazandırmak hem de öğrenme motivasyonunu artırmak için muazzam bir potansiyel sunmaktadır (Gusev ve Armenski, 2014). Modern teknolojiler, öğrencilerin akademik performansını daha iyi izlemek ve daha geçerli bir şekilde değerlendirmek için öğretmenlerin sınıf testlerini etkin bir şekilde yönetebilmelerini sağlayan yeni kanallar da sağlamaktadır (Mayrath, Clarke-Midura ve Robinson, 2012; Wu, Kuo, Jen ve Hsu, 2015). Öğrencilerin akademik performansını değerlendirmek için teknolojiden yararlanmanın yolu, bilgisayarlı testlerin ve uyarlanabilir testlerin entegrasyonunu sembolize eden bilgisayarlı uyarlanabilir test BOB testleridir (Chang, 2015).

Günümüzde de kullanılan ve ilk defa Binet tarafından geliştirilen IQ testi dünyada kullanılan ilk bireysel test kabul edilmektedir. Bunun en büyük nedeni test madde tepki kuramına göre (MTK) hazırlanmış olmasıdır. Madde Tepki Kuramı; bireyde ölçülmek istenen bilgi, yetenek veya ilginin düzeyi ile sorulara verilen cevaplar arasında bir ilişki olduğunu göstermekte ve bu ilişkiyi matematiksel bir fonksiyon ile açıklamaktadır (DeMars, 2010; Embretson ve Reise, 2000). Binet'in bu testi oluşturmasının temel amacı, oluşturulan testin bireyselleştirilerek uygulanmasıdır. Bu amaçla BOB testleri Binet'in oluşturduğu IQ testlerine benzetilebilir. Bu iki test arasındaki fark; Binet'in oluşturduğu testleri insanlar uygularken, BOB testlerini bilgisayarlar uygulamaktadır (Thompson ve Weiss, 2011). Bilgisayar öğrencinin soruları doğru ya da yanlış cevaplayışına göre, önceden oluşturulmuş soru havuzundan soruları seçer. Bunun nedeni öğrencinin sahip olduğu yetenek veya bilgi düzeyinden daha zor ya da daha kolay sorular sorulmasının öğrencinin sahip olduğu yetenek düzeyi veya bilgisi hakkında yeterli bilgi vermeyişi ve öğrencinin seviyesine uygun sorular sorulmaması nedeniyle sınavdan sıkılmasına neden olmasıdır (Özdemir, 2015).

 ğrenciyi tanımak, biçimlendirmek ve karar vermek iin kullanılan bilgisayar ortamında bireyselleştirilmiş testler,  ğrenen ve  ğreten iin birok avantaj saėlamaktadır. En  nemli avantajlarından bir tanesi de daha az soru sorularak daha g venilir sonulara ulařabilmedir (Wainer, 1993). Bunun nedeni, bilgisayarın sınav s recinde, kolay ya da zor soru yerine  ğrencinin sahip olduėu bilgi d zeyini  lebilecek soruları belirleyip sormasıdır. BOB testlerinin saėladığı diėer bir avantaj ise zamandan kazanım olması ve sınav bittiėi zaman  ğrenciye hemen sınav hakkında d n t verebilmesidir (Lin, 2012). BOB testlerinin diėer bir faydası ise, bir test maddesine grafik, ses veya video gibi eřitli multimedia eklenebilmesi ve test maddesini kolayca tekrar kullanabilmesidir. Test  ğelerinden gelen farklı bilgiler ve  ğrencilerden gelen test verileri, bilgisayarlar tarafından zahmetsizce ve derhal toplanıp kaydedilebilir. Bilgisayar tabanlı  ğrenmenin yanında mobil  ğrenme (m- ğrenme)  nemli bir geliřmedir. Genel olarak, tařınabilir aygıtların esneklik, tařınabilirlik, d ř k maliyet, kullanım kolaylığı ve zamanında uygulama gibi birok avantajı vardır (Huang, Lin ve Cheng, 2009). Bu avantajlar  ğretmenlere ve  ğrencilere uzaktan eėitime m- ğrenme ortamlarıyla katılmak iin daha kolay bir yaklařım saėlar.

Weiss ve Thompson'a g re (2011) BOB testinin uygulanması beř ařama ile saėlanmış olur. Bu ařamalar;

1. BOB testinin uygulanacağı konunun belirlenmesi
2. Madde havuzunun oluřturulması (30 maddelik bir test iin 200-300 arası soru oluřturma)
3. Test s recinde madde seim kriterlerinin belirlenmesi
4. Puanlama algoritmanın belirlenmesi
5. Analizi durdurma kuralının belirlenmesidir.

BOB testleri sistemleri beř ana bileřenden oluřur. Bu bileřenler; konunun belirlenmesi, madde havuzu, madde seim kriterlerinin belirlenmesi, madde seme algoritması ve sonlandırma kriteridir. BOB'un temel konsepti, bir sınava giren kiřinin yetenek seviyesiyle bir test maddesinin en iyi eřleřmesidir. Madde zorluėu ve en uygun eřdeėerliėi, adayın yetenek seviyesine iliřkin en  st d zeyde bilgi saėlar. Herhangi bir BOB sisteminin bařarısı b y k oranda kaliteli  r n havuzuna baėlıdır. Nihai  ėe havuzu, bařvuru sahibine sunulacak  ğelerin sayısından yaklařık 5 ila 10 kat daha fazla  ėe ierir. Buna g re 30 maddelik bir test iin bir madde havuzuna 150-300 madde eklenmesi  nerilmektedir (Georgiadou, Triantafillou, ve Economides, 2006). Maksimum Bilgi Seimi (MBS) y ntemi,  ğrencinin yetenek seviyesine en uygun soruları seebilmesi iin kullanılabilir. Bu y ntem, adayın yetenek seviyesi hakkında en fazla bilgi saėlayan soruyu seer. Bařlangı sorusu genellikle, katılımcının orta d zeyde bir yetenek seviyesine sahip olduėu varsayılarak seilir (Boyd, 2003). Yetenek deėerlendirmesi, her cevaptan sonra g ncellenir ve yeni soru, yeni tahmin tarafından tespit edilen doėru  zelliklere g re seilir. Literat rde, arařtırmacının yetenek d zeyini tahmin etmeye y nelik eřitli y ntemler mevcuttur ancak Maksimum Olabilirlik Tahmini (Maximum Likelihood Estimation - MLE) en sık kullanılan y ntemdir. Sınav sahibi tarafından verilen cevaplara dayalı olarak bir cevap

dizisinin oluşma ihtimalini ortaya koyan yetenek seviyesi değerini tanımlar. İstatistiksel olarak kabul edilebilir bir hassasiyetle veya maksimum test sayısı gibi bir sınıra erişene kadar (Tian Miao, Zhu X. ve Gong, 2007), araştırmacının yetenek tahmini değerini yeniden değerlendirir.

Bugün, BOB testleri dünyada üniversiteler tarafından geçerliliği kabul edilmiş olan Yabancı Dil Olarak İngilizce Testi (TOEFL), Lisansüstü Kayıt Sınavları (GRE), Lisansüstü Yönetim Kabul Testi (GMAT), TIMSS (Trends in International Mathematics and Science Study) gibi bir dizi yaygın testte, en azından kısmen uyarlanabilir şekilde uygulanmakta olup, Uluslararası Öğrenci Değerlendirmesi Programı (PISA) kapsamında öğrenci yeterliklerinin değerlendirilmesi için pilot uygulama planlanmaktadır (Frey ve Seitz, 2009).

Ölçmedeki Yeni Arayışın Şimdilik Tezahürü Olarak Kısa Cevaplı Maddeler

Cevabı sınava girenler tarafından bulunup yazılan ölçme araçlarından bir tanesi olan kısa cevaplı sorular, açık uçlu sorulardan biri olan klasik yazılı sınavlara benzese de, cevaplarının kısa olması nedeniyle öğrenci yorumunun olmaması özelliği ile klasik yazılı sınavlardan farklılık göstermektedir. İyi hazırlanması halinde bilişsel öğrenme kademelelerinin her bir düzeyini ölçebilme yetisine sahip, hazırlanması, uygulaması ve puanlaması kolay olan, cevabı bir kelime, kısa sözcük grubu veya sayı ya da sembollerden oluşan ölçme araçlarıdır (Yavuz ve Bilgeç, 2016). Kısa cevap maddeleri; soru sorma, tamamlama ve ilişki olmak üzere 3 grup olarak sınıflandırılır. Kısa cevaplı maddelerde sorular dolaysız olarak sorulur ve cevap nettir, öğrenciler kısa cevap verirler. Kısa cevaplı maddelerden tamamlama sorularında, öğrenciye tamamlanmamış bir cümle verilir ve öğrencinin bir veya birkaç kelime ekleyerek cümleyi tamamlaması istenir. Kısa cevaplı maddelerden ilişki çeşidi sorularında, öğrencilerin verilen kelimeyle sembolleşen sayı, etiket veya sembolleri yazmaları, resimleri çizmeleri istenir (Gredler, 1999; Wesman, 1971, akt. Nıtko, ve Brookhart, 2016).

Kısa cevaplı sınavlar eğitimde önde gelen İngiltere, ABD, Fransa, Avustralya gibi ülkelerin düzenlemiş olduğu, GRE (Graduate Record Examination), AP (Advanced Placement), TOEFL (Test of English as a Foreign Language) ve IELTS (International Language Testing System) gibi sınavlarda öğrencilerin kritik düşünme ve bilgilerini analiz etme gibi üst düzey bilişsel becerilerini ölçmek için kullanılmaktadır (Yavuz ve Bilgeç, 2016; GRE, 2017; TOEFL, 2017).

Ülkemizde de çoktan seçmeli sınavların öğrencilerin üst düzey bilişsel becerilerini ölçmede yetersiz olduğu belirlenmiş, alternatif bir sınav olarak 2017 LYS sınavında kısa cevaplı sorular sorulmaya başlanmıştır. 2017 yılında yapılan LYS sınavında fizik, kimya, biyoloji, coğrafya, tarih, coğrafya ve felsefe grubu ile din kültürü ve ahlak bilgisi testinde 1'er, Türk dili ve edebiyatı testinde 2, matematik ve yabancı dil testlerinde 3'er tane kısa cevaplı sorular sorulmuş, ileriki sınavlarda bu soruların sayılarının arttırılacağı belirtilmiştir (ÖSYM, 2017).

Nitko ve Brookhart (2016) sınavlarda kısa cevaplı maddelerin kullanımının avantajlarını  oye belirtmiŐlerdir:

- Kısa cevaplı maddeler  ğrencilerin bilgiyi kavrama ve hatırlama gibi alt d zey becerilerini belirlemede kullanılabilir.
- Verilerin yorumlanması, kuralların uygulanması, fen ve matematik gibi sayısal iŐlem gerektiren derslerde sembolleri kullanabilme ve problemlerde d rt iŐlem yapabilme, g rsel sorularda verilen resimleri yorumlayabilme ve  ıkarımda bulunabilme gibi  ğrencilerin  st d zey becerilerinin  l lmesinde kullanılabilir.
- Soruların oluŐturulması kolaydır ve puanlaması s zl  ve klasik yazılı sınavlara g re daha objektiftir.
- Sınava girenlerin, kısa cevaplı maddeleri Őansa dayalı dođru cevaplama ihtimali  ok d Ő kt r ve kısa cevaplı sınavlarda daha  ok soru sorma olanađına sahip olduđu i in sınava girenin hangi soruların sorulacađını tahmin etme Őansını azaltır.
- Kısa cevaplı maddelerin; dođru-yanlıŐ,  oktan se meli ve eŐleŐtirmeli testlere nazaran puanlaması daha zordur.
-  ğrencilerin verdiđi cevaplar kısmen dođru cevabı i erebilir,  ğrenci cevap verirken imla ve c mle kurma hataları yapabilir ki bu da puanlamayı etkileyebilir.
-  ğrenciler kendi d Ő ncelerini organize ederek anlatmalarına olanak sađlamaz.
-  ğrencilerin vermiŐ oldukları cevaplarda kendi d Ő ncelerini organize ederek tanıma imk nı yoktur (Turgut ve Baykul, 2010;G l, 2011; Nitko ve Brookhart, 2016).

Kısa Cevaplı Maddelerde Ge erlik-G venirlik ve KullanıŐlılık

Kısa cevaplı maddelerin cevaplanması kısa olduđu i in belirlenen bir sınav s resinde  ok sayıda soru sorulabilir. Soru sayısının artırılması ile bir ok hedef davranıŐının  l lebilmesi sađlanacađı i in g venirliđi ve kapsam ge erliđi y ksektir. Soru cevaplama da Őans fakt r  olduk a d Ő kt r. Bu nedenle g venirlik, g venirliđe bađlı olarak da ge erlik y ksektir.

Kısa cevaplı maddelerde istenilen cevaplar kısa olduđu i in  ğrenenin sorulara verdiđi cevaplar kısıtlıdır. Bu durum  ğrenenin yorum yapma durumunu kısıtladıđı i in puanlayıcı diđer a ık u lu sınavlara g re daha objektif puan verir. Ama sonu ta puanlama tam anlamıyla nesnel olamaz. Bu nedenle a ık u lu sorulara g re puanlama g venirliđi daha y ksektir.

Bu testler kısa cevaplar gerektirdiđi i in, hazırlanması, uygulanması ve puanlaması fazla zaman almaz. Bu nedenle hazırlama, uygulama ve puanlama y n nden kullanıŐlı testlerdir (Thorndike ve Thorndike-Christ, 2010).

Sonuç

Teknoloji alanında meydana gelen hızlı gelişmeler, teknolojiyi hayatımızın ayrılmaz bir parçası durumuna getirmiş, eğitim başta olmak üzere birçok alanı etkilemiş ve bilgi toplumuna geçişi gerekli kılmıştır. Teknolojinin etkilemiş olduğu unsurlardan bir tanesi de programın, dolayısıyla da eğitimin canlılığını koruyan ve devamlılığını sağlayan unsur olan “ölçme ve değerlendirmedir”.

Ölçme değerlendirmenin eğitimin temel unsurlarından biri olduğu söylenebilir. Zira eğitimde ölçme araçlarının çeşitlenmesi ve çeşitlenmeyle beraber daha kaliteli hale gelmesi büyük önem oluşturmaktadır. Ölçme araçlarının daha kaliteli hale gelmesi alanında yapılan incelemelerde birçok yeni yönelim olduğu göze çarpmaktadır.

Günümüzde gelişmiş teknolojinin ve buna bağlı olarak bilgisayarların eğitimin ölçme ve değerlendirme aşamasında kullanım amacı, ölçmeyi uygulayan ve öğrenenlerden veri toplamak için kullanılan ölçme araçlarından ölçme sonucunda daha güvenilir ve geçerli sonuçlar elde etmektir (Weis, 2004). Günümüzde eğitimde ileri düzeyde görülen ülkelerin uyguladığı ve sonuçlarını kabul ettiği GRE (Graduate Record Examination), AP (Advanced Placement), TOEFL (Test of English as a Foreign Language) ve IELTS (International Language Testing System) sınavlarda bireysel farklılık göz önünde bulundurulduğu bilgisayar kullanımı etkin olarak görülmektedir (Frey ve Seitz, 2009; Wu, Kuo, Jen ve Hsu, 2015).

Verilen eğitimin kalıcılığı ve öğrencilere aktarımı konusunda geri dönüt alınması bakımından yapılan sınavlar büyük önem taşımaktadır. Günümüzde eğitim kalitesinin ölçüm araçları olan birçok sınav yöntemi geliştirilmiştir. Bunlardan biri de kısa cevaplı çoktan seçmeli yöntemidir. Kısa cevaplı çoktan seçmeli sınav türü bireylerin belirli bir konu veya problemle ilgili gerekli bilişsel işlemlerin yapılmasından sonra verilen cevaplar içerisinde doğru olanı işaretlemesi sistemine dayanmaktadır (Yavuz ve Bilgeç, 2016).

Kısa cevaplı sınav sisteminin yazılı ve çoktan seçmeli sınav sistemlerine göre daha avantajlı olduğu söylenebilir. Çünkü seçmeli sınav sistemi öğrencilerin ezbere daha çok yönelmelerine neden olmakta ve bireylerin analitik düşünme becerilerinin gelişimine engel olabilmektedir. Yazılı sınav sistemleri ise zaman kullanımı açısından ve öğrencilere yöneltilen soru sayısında ciddi düşümlere neden olduğu için öğrencilerin öğrenim yaşantılarının tam olarak ölçülememesine neden olmaktadır. Ancak kısa cevaplı sınav sistemi, çoktan seçmeli ve yazılı sınav sisteminin barındırmış olduğu avantajları (zaman tasarrufu, analitik çözümlenme, öğrenme yaşantılarının ölçümüne ilişkin birden çok sorunun sorulabilmesi vb.) bünyesinde toplamaktadır. Kısa cevaplı sınav sistemi bireylerin problem çözme becerilerini geliştirmeleri yanında zaman ekonomisine yaptığı katkı bağlamında oldukça dikkat çekicidir (Nitko ve Brookhart, 2016; Gredler, 1999) .

Kısa cevaplı sınav sisteminin yeni nesil bir sistem olması, sistemin ülkemizde henüz etkin kullanılmıyor olması ve sistemin tanıtılması açısından yapılan çalışmanın önemli olduğu söylenebilir. Bunun yanı sıra yapılan alan taraması neticesinde kısa cevaplı sınav

sistemleri ve  l m kurallarıyla ilgili ok fazla alıřmaya rastlanmamıřtır. Yapılan alıřmayla oluřan alanyazın bořluęunun doldurulması ve bilgisayar ortamında bireyselleřtirilmiř testlerde kısa cevaplı sınav sistemiyle ilgili alan yazınına katkı saęlanmıřtır.

Kaynaka

- Akadaę, T. (2010).  ğretmenlerin ilköęretim programındaki y ntem teknik  lme ve deęerlendirme konularına iliřkin eęitim ihtiyaları. *T rk D nyası Sosyal Bilimler Dergisi*, 53, 29-50.
- Akpınar, B. (2015). *Eęitimde program geliřtirme*. Ankara: Data Yayıncılık.
- Arslan, S. ve  zpinar, İ. (2008).  ğretmen nitelikleri: ilköęretim programının beklentileri ve eęitim fak ltelerinin kazandırdıkları. *Necatibey Eęitim Fak ltesi Fen ve Matematik Eęitim Dergisi*, 2(1), 38-63.
- Berber, C. N. (2015). T rkiye ve Hong Kong fizik  ğretim programlarının karřılařtırılması. *Necati Bey Eęitim Fak ltesi Elektronik Fen ve Matematik Dergisi*, 9(2), 61-84.
- Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.
- Black, P., Harrison, C., Lee, C., Marshall, B. and William, D. (2003). *Assessment for learning: Putting it into practice*. New York, NY: Open University Press.
- Boyd, A.M. (2003). *Strategies for controlling testlet exposure rates in computerized adaptive testing systems*. Unpublished doctoral dissertation. Texas University.
- Chang, H. H. (2015). Psychometrics behind computerized adaptive testing. *Psychometrika*, 80, 120.
- Crisp, V., and Ward, C. (2008). The development of a formative scenario-based computer assisted assessment tool in psychology for teachers: The PePCAA project. *Computers & Education*, 50(4), 1509–1526.
- Daniel, L. G., and King, D. A. (1998). Knowledge and use of testing and measurement literacy of elementary and secondary teachers. *The Journal of Educational Research*, 91(6), 331-344.
- DeMars, C. E. (2010). Type I error inflation for detecting DIF in the presence of impact. *Educational and Psychological Measurement*, 70, 961-972.
- Demirel,  . (2012). *Eęitimde program geliřtirme*. Ankara: PegemA Yayıncılık.
- DeVellis, R. F. (2014). *Scale development theory and applications*. (T. Totan ve A. ř. Avřar, ev.) Ankara: Nobel Akademik Yayıncılık.
- Embretson, S. E., and Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Ert rk, S. (2013). *Eęitimde program geliřtirme*. Ankara: Edge Akademi.

- Frey, A., and Seitz, N. (2009). *Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges*, *Studies in Educational Evaluation*, 35, 89-94.
- Gelen, İ. ve Beyazıt, N. (2007). Eski ve Yeni ilköğretim programları ile ilgili çeşitli görüşlerin karşılaştırılması, *Educational Administration: Theory and Practice*, 51, 457-476.
- Georgiadou, E., Triantafillou, E., and Economides, A.A. (2006). Evaluation parameters for computer-adaptive testing. *British Journal of Educational Technology*, 37(2), 261-278.
- GRE Türkiye Resmi Web Sitesi (2017). <https://www.ets.org/gre> adresinden 12.06.2017 tarihinde alınmıştır.
- Gredler, M. E. (1999). *Classroom assessment and learning*. Newyork: Longman.
- Gusev, M., and Armenski, G. (2014). E-assessment systems and online learning with adaptive testing. In (M. Ivanovic, and L. C. Jain (Eds.), *E-Learning Paradigms And Applications*. Heidelberg: Springer.
- Gül, E. (2011). *İlköğretim öğretmen adaylarının ölçme-değerlendirme okuryazarlığı ve ölçme-değerlendirmeye ilişkin tutumlarının belirlenmesi*, Yayınlanmış Yüksek Lisans Tezi. Fırat Üniversitesi, Fen Bilimleri Enstitüsü.
- Huang, Y., Lin, Y., and Cheng, S. (2009). An adaptive testing system for supporting versatile educational assessment. *Computers and Education*, 52, 53-67.
- Hunkins, P.C., and Ornstein C.A. (2014). *Eğitim programı, temeller, ilkeler ve sorunlar*. (A. Arı, Çev.). Konya: Eğitim Yayınevi.
- Kandemir, A. (2016). *İlkokul 2. sınıf İngilizce öğretim programının katılımcı odaklı program değerlendirme yaklaşımıyla değerlendirilmesi*, Yayınlanmış Yüksek Lisans Tezi. Pamukkale Üniversitesi, Eğitim Bilimleri Enstitüsü
- Karakaya, Ş. (2004). *Eğitimde program geliştirme çalışmaları ve yeni yönelimler*. Ankara: Asil Yayın Dağıtım.
- Kelly, D.L. (2002). The TIMSS 1995 international benchmarks of mathematics and science achievement: profiles of world class performance at fourth and eighth grades. *Educational Research and Evaluation*, 8 (1), 41-54.
- Koç, E. S. (2015). Oyun ve fiziki etkinlikler dersi öğretim programının karşılaştırmalı olarak incelenmesi (kanada örneği). *Kastamonu Eğitim Dergisi*, 21(1), 263-280.
- Lin, H. (2012). *Item selection methods in multidimensional computerized adaptive testingadopting polytomously-scored items under multidimensional generalized partial creditmodel*. Unpublisjed Doctoral Dissertation. University of Illinois at Urbana-Champaign.

- Marsh, C. J. (2004). *Key concepts for understanding curriculum (3rd Edition)*. Taylor & Francis e-Library.
- Mayrath, M. C., Clarke-Midura, J., and Robinson, D. H. (2012). *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. Charlotte, NC: Information Age Publishing.
- McBride, J.R., and Martin, J.T. (1983). Reliability and validity of adaptive ability tests in a military setting. in Weiss D.J. (Ed.) *New Horizons in Testing*. New York: Academic Press.
- Nitko, A. J., and Brookhart, S. M. (2016). * ğrencilerin eđisel deęerlendirmesi*. (B. Bıçak, M. Bahar ve S.  zel,  ev.). Ankara: Nobel Akademik Yayıncılık.
- Ornstein, A. C., and Hunkins, F. P. (2009). *Curriculum: foundations, principles, and issues (5th Edition)*, Boston: Allyn & Bacon.
-  SYM Resmi Web Sitesi. (2017). <http://www.osym.gov.tr/TR,12909/2017-lisans-yerlestirme-sinavlari-2017-lys-acik-uclu-sorular-hakkinda-bilgilendirme-ve-acik-uclu-soru-orneklere-05012017.html> 12.06.2017 tarihinde alınmıřtır.
-  zcelik, D. A. (2010). *Eđitim programları ve  ğretim*. Ankara: Pegem Akademi.
-  zdemir, B. (2015). *Madde d zeyinde boyutluluk modellerinin bilgisayar ortamında bireyselleřtirilmiř test y ntemleri  zerindeki etkisinin incelenmesi*. Yayınlanmış Doktora Tezi. Hacettepe  niversitesi, Eđitim Bilimleri Anabilim Dalı.
- Pendergast, D., Garvis, S., and Keogh, J. (2011). Pre-service student-teacher self-efficacy beliefs: An insight into the making of teachers, *Australian Journal of Teacher Education*, 36(12), 46–58.
- Physical Education Curriculum Guide: Primary/Elementary. (2011). <http://www.ed.gov.nl.ca/edu/k12/curriculum/guides/physed/index.html#primary/> adresinden 21.05.2017 tarihinde alınmıřtır.
- PISA T rkiye Resmi Web Sitesi. (2017). http://pisa.meb.gov.tr/?page_id=18 adresinden 25.05.2017 tarihinde alınmıřtır.
- Posner, G. J. (1995). *Analyzing the curriculum*. New York: McGraw Hill.
- Practical Assessment, Research & Evaluation, 16(1). Available online: <http://pareonline.net/getvn.asp?v=16&n=1>.
- Preston, R., and Aikman, S. (2005). Setting the scene: measurement and learning beyond the mainstream. *International Journal of Educational Development*, 25, 99-110.
- Sabancı, O. ve Yazıcı, K. (2017).  ğretmen adayların  lme ve deęerlendirmeye y nelik yeterli algılarının incelenmesi. *Trakya  niversitesi, Eđitim Fak ltesi Dergisi*, 7(7), 128-153.
- Semerci,  . (2008). * lme ve deęerlendirme*. Ankara: Pegem Akademi.

- Seng, T. O., Parsons, R. D., Hinson, S. L., and Sardo-Brown, D. (2003). Educational psychology: A practitioner–researcher approach. *Thomson Learning*, 250–251.
- Sönmez, V. (2012). *Program geliştirmede öğretmen el kitabı (17. Baskı)*. Ankara: Anı Yayıncılık.
- Tekin, H. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı Yayınevi.
- The American Educational Research Association, The American Psychological Association, The National Council on Measurement in Education. (1997). Eğitimde ve psikolojide ölçme standartları. (S. Hovardaoğlu ve N. Sezgin, Çev.) Ankara: Türk Psikologlar Derneği.
- Thompson, Nathan A., and Weiss, David A. (2011). *A framework for the development of computerized adaptive tests*.
- Thorndike, R. M., and Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education*. Pearson.
- TIMMS 2015, Türkiye. (2017). <http://timss.meb.gov.tr> adresinden 25.05.2017 tarihinde alınmıştır.
- Tian, J., Miao D., Zhu X., and Gong, J. (2007). An introduction to the computerized adaptive testing. *US-China Education Review*, 4(1), 72-81.
- TOEFL Resmi Web Sitesi. (2017). <https://www.ets.org/toefl> adresinden 12.06.2017 tarihinde alınmıştır.
- Tschannen-Moran, M. And Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct, *Teaching and Teacher Education*, 17(7), 783–805.
- Tuncer, M. ve Özeren, E. (2015). Alternatif ölçme ve değerlendirme araçları üzerine bir meta analiz çalışması. *Eğitim ve Öğretim Araştırmaları Dergisi*, 4(2), 374-386.
- Turgut, M. F. ve Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- Uşun, S. (2012). *Eğitimde program değerlendirme süreçler yaklaşımlar ve modeller*. Ankara: Anı Yayıncılık.
- Wainer, H. (1993). Some practical considerations when converting a linearly administrated test to an adaptive format. *Educational Measurement: Issues and Practices*, 12, 15-20.
- Wang, T. H. (2007). What strategies are effective for formative assessment in an e-learning environment? *Journal of Computer Assisted Learning*, 23(3), 171–186.
- Weiss, J. D. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, 37(2), 70-84.
- Wesman, A. G. (1971). *Educational measurement. Writing the test item*. Wahington DC: American Council on Education.

- Wu, H. K., Kuo, C. Y., Jen, T. H., and Hsu, Y. S. (2015). What makes an item more difficult? effects of modality and type of visual information in a computer-based assessment of scientific inquiry abilities. *Computers & Education*, 85, 35-48.
- Yavuz, İ. ve Bilgeç, İ. (2016). Aık ulu sorularla yapılan matematik sınavlarının  lme ve deęerlendirilmesinin incelenmesi. *Eęitim ve  ęretim Arařtırmaları Dergisi*, 5(3), 183-193.
-  ltanır, G. (2016). *Program deęerlendirme*. Ankara: Nobel Yayınevi.