

Türk Eğitim Sistemi'nde Sınıf İçi ile Geniş Ölçekli Ölçme ve Değerlendirmeye Genel Bir Bakış

Mehmet ŞATA

Ağrı İbrahim Çeçen Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Ağrı, Türkiye

Öz

Yeni öğretim programında öğretmenlerin yeterlikleri arasında ölçme ve değerlendirme bilgisi önem kazanmıştır. Bu yüzden sınıf içi ölçme ve değerlendirme faaliyetleri geniş ölçekli ölçme ve değerlendirme faaliyetleri kadar önemli bir noktaya gelmiştir. Bu araştırmanın amacı, Türk Eğitim Sistemi'nde sınıf içi ve geniş ölçekli ölçme ve değerlendirme faaliyetlerinin nasıl yapıldığının incelenmesidir. Ulusal alanyazına bakıldığında hem sınıf içi hem de geniş ölçekli ölçme ve değerlendirme faaliyetlerinin ciddi bir şekilde yetersiz olduğu görülmüştür. Sonuç olarak hem sınıf içi hem de geniş ölçekli ölçme ve değerlendirme faaliyetlerinde geçerliğin çok düşük olduğu ve bundan dolayı ölçme ve değerlendirme faaliyetlerinin amaca hizmet etmediği görülmektedir. Ayrıca geçerliği olmayan sınavların (hem sınıf içi hem de geniş ölçekli olsun) sonuçları kullanılarak yapılacak olan değerlendirmeler sonucunda öğrenciler hakkında doğru kararların verilmesinin mümkün olmadığı görülmektedir.

Anahtar Kelimeler: Sınıf İçi, Geniş Ölçekli, Ölçme ve Değerlendirme, Geçerlik.

A General Outlook on Classroom and Large-Scale Measurement and Evaluation in Turkish Education System

Abstract

The understanding of measurement and evaluation which is one of the teacher qualifications in the new curriculum has come into prominence. Therefore, classroom measurement and evaluation applications have become as significant as large-scale measurement and evaluation applications. The purpose of the study is to investigate how classroom and large-scale measurement and evaluation practices are applied in Turkish Education System. As a result of the examination of national literature, it was observed that classroom and large-scale measurement and evaluation applications were inadequate. Consequently, it was seen that the validity of classroom and large-scale measurement applications was very low and that the applications of measurement and evaluation did not serve their purposes. Besides, the right decisions on students would not be true because of the judgments made from the results of the invalid exams (classroom or large-scale).

Keywords: Classroom, Large-Scale, Measurement and Evaluation, Validity.

Makale Bilgileri:

Gönderim / Recieved:
11.01.2016

Kabul / Accepted:
29.03.2016

Sorumlu Yazar:

Ağrı İbrahim Çeçen
Üniversitesi,
Eğitim Fakültesi,
Ağrı, Türkiye

mehmetwsata@gmail.com

1. Giriş

Eğitim ve öğretim etkinlikleri öğrencilerin konu alanına yönelik bilgi ve zihinsel becerilerini kapsadığı gibi, tutumlar, ilgiler, öğrenme stilleri gibi duyuşsal özellikleri de kapsamaktadır. İyi bir ölçme ve değerlendirme işlemi tüm bu boyutları kapsayan, öğrencinin gelişimini farklı boyutlarla ele alan bir işlemdir. Sınıfındaki öğrencilerin öğrenim gelişimleri hakkında bilgi edinmek isteyen bir öğretmen, üç adımlı bir süreci izlemek zorundadır. Bu adımlar sırasıyla ölçme, durum belirleme (assessment) ve değerlendirme (evaluation) basamakları olarak tanımlanmaktadır (Berberoğlu, 2006).

Ölçme ve değerlendirmenin eğitim bilimlerindeki önemi son yıllarda daha çok vurgulanmaya başlanmıştır. Bir eğitim programının işlevliğinin belirlenmesinde, eğitim yöntem ve stratejilerinin etkililik seviyelerinin tespit edilmesinde, öğrenci başarısının, öğrencilerin öğrenmedeki güçlü ve zayıf yönlerinin tanımlanmasında ölçme ve değerlendirmeden faydalanmaktadır. Bundan dolayı eğitim sisteminin girdi, süreç ve ürün boyutlarının hepsinde ölçme ve değerlendirmeye ihtiyaç duyulduğu söylenebilir (Güler ve Gelbal, 2010). Ölçme ve değerlendirme, öğrencilerin ne öğrendikleri hakkında bilgi ve kanıt toplama süreci olarak tanımlanmaktadır (Chen 2003; Akt: Şahin ve Karaman, 2013). Ölçme, bir özelliğin/yapının gözlenerek gözlem sonuçlarının sayı veya sembollerle temsil edilmesi olarak tanımlanırken, değerlendirme daha geniş kapsamlı bir süreç olup; ölçüm, ölçüt ve karar verme aşamalarından oluşur (Baykul, 2010).

Öğretmenlerin öğrenci başarısı ve performansına ilişkin ölçme ve değerlendirme süreçlerini güvenilir ve geçerli bir biçimde yapabilmeleri için ölçme ve değerlendirme alanında yeterli donanıma sahip olmaları gereklidir (Özbaşı, 2009). Türkiye'de 2005 yılında gerçekleştirilen ve daha sonraki yıllarda güncellemeleri devam eden öğretim programlarındaki değişimin önemli bir kısmı, ölçme ve değerlendirme alanında gerçekleştirilmiştir. Değişim ve beraberinde getirdiği uygulamalar, öğretmenlerin ölçme ve değerlendirme alanına ilişkin bilgi ve becerilerini geliştirmeye yöneltmiştir. 2008 yılında yayınlanan Öğretmenlik Mesleği Genel Yeterlikleri de öğretmenlerin ölçme ve değerlendirmenin hangi boyutlarında yeterliklere sahip olmaları gerektiğini ortaya koymuş, yapılmakta olan hizmet içi eğitim faaliyetleri arasında ölçme ve değerlendirmeyle ilgili olanların sayısı ve süresi artmaya başlamıştır (MEB, 2008).

Turgut ve Baykul (2014) öğretmenlerin ölçme ve değerlendirme yeterlik alanına ilişkin sahip olmaları gereken yeterliklerin; öğretimin amacını saptama, amaçlara uygun soru hazırlama, ölçme aracı geliştirme, ölçme aracını uygulama, analiz etme ve değerlendirme olduğunu vurgulamaktadır. Karaca (2004) ise, öğretmenlerin; amaç-amaç davranışlara uygun, geçerli ve güvenilir ölçme değerlendirme araçlarını hazırlama, öğretimin etkinliğini veya öğrencilerin hedef davranışlara ulaşma düzeyini değerlendirme, değerlendirme sonuçlarını öğrencilerin başarısını arttırmada kullanma ve sonuçları diğer eğitimcilere, ailelere, öğrencilere ve diğer paydaşlara iletme yeterliklerine sahip olmalarının gerektiğini vurgulamaktadır.

2. Sınıf İçi Ölçme ve Değerlendirme

Ölçme ve değerlendirme öğretimin önemli parçalarından biridir. Okul eğitiminde ölçme ve değerlendirmeden yararlanmanın en önemli amacı, öğrencinin sınıf içi performansını tespit etmektir (Kutlu, Doğan ve Karakaya, 2014, s. 30). Sınıf içi ölçme ve değerlendirme faaliyetlerinin iki temel amacı, öğrenmeyi geliştirmek ve öğretim etkinliklerini nitelik bakımından arttırmaktır. Öğrenmeyi geliştirmek için öğrencilerin güçlü ve geliştirmeye açık yanlarının ortaya çıkarılması ile başlanmalıdır. Buradan elde edilen bilgilerle öğretim, materyal ve etkinliklerinin sürekli olarak geliştirilmesine olanak sağlayan bir süreç olarak devam edecektir. Ancak ülkemizdeki geniş ölçekli test uygulamaları sınıf içi ölçme ve değerlendirme sonuçlarının bu şekilde yorumlanmasına imkân vermemektedir. Okullarımızda genellikle yalnızca öğrencilere not vermek amacı ile ölçme ve değerlendirme faaliyetleri (sınavlar) yapılmakta, sonuçların öğrenciyi geliştirme amacıyla kullanılmasına nadiren rastlanmaktadır (Berberoğlu, 2010).

Okullarda hazırlanan sınavların temel ögesi program hedeflerine ulaşılma seviyelerini belirlemek olduğundan dolayı bu sınavlarda kapsam geçerliği hayati bir önem taşımaktadır (Berberoğlu, 2012). Bir sınavın geçerliği, amacına ne ölçüde hizmet ettiğine dair toplanan kanıtlarla değerlendirilmektedir (Messick, 1989). Geçerliği olmayan sınavlarla yapılan değerlendirmeler neticesinde öğrenciler ile ilgili doğru kararların verilmesi mümkün görünmemektedir. Okullarda yapılan ölçme ve değerlendirme faaliyetlerinin temel maksadı öğrenmeleri geliştirmektir. Öğrenmeleri geliştirmek için ölçme sonuçlarının öğrenciye ve öğretmenlere geri bildirim sağlaması ile olası olabilir. Okullardaki ölçme ve değerlendirme etkinlikleri kazanımlara ulaşılma konusunda bilgi toplamayı hedeflemelidir. Öğretmenler öğretim yöntemlerini, kullandıkları materyalleri sınav sonuçlarına göre değerlendirebilir.

Eğitim sisteminin işlevselliğini yerine getirmesinde bir başka değişle geçerli öğrenme yaşantılarının meydana gelmesinde öğrenme-öğretme süreci önemlidir. Bu yaşantıların düzenlenmesinde öğrenmelerin nasıl gerçekleştiği ve bireye nasıl öğretilceğinin bilinmesi gerekmektedir (Ayva, 2010). Öğretmenin rolleri, sorumlulukları ve gücünü nasıl kullandığı öğretim sürecindeki başarı ve başarısızlık ile doğrudan ilgili olmaktadır. Sıcak, hevesli, heyecanlı, işe yönelik ve düzenli olma, çeşitli materyalleri kullanma, açıklamalar yapma, derse bir önceki dersi özetleyerek başlama, arka arkaya soru sorma yerine tek ve açık soru sorma öğrenci başarısı ile olumlu ilişkiler gösteren öğretmen davranışı olarak bulunmuştur. Öğretmenin başarısı öğrencinin başarı, tutum, ilgi ve becerisine yansımaktadır (Sünbül, 1996).

İlköğretimden üniversiteye kadar öğretimin her kademesinde eğitim-öğretim etkinliklerinin büyük çoğunluğu sınıf ortamında olmaktadır. Etkili bir öğrenme-öğretme ortamı oluşturmak isteyen bir öğretmen, elindeki tüm olanakları ve materyalleri kullanarak, sınıfı en iyi nasıl örgütleyeceği konusunda çok yönlü olarak düşünmelidir. Öğrenme-öğretme süreci boyunca, konuların özelliklerine ve öğrencilerin beklentilerine en uygun olan etkinlikler tespit edilerek sınıf içi öğretim süreçleri gerçekleştirilmelidir (Gökçe, 2002). Yapılan çalışmaların önemli bir bölümü öğretmenlerin sınıf içi ölçme ve değerlendirme alanında bilgi ve becerilerinin istenen ve olması gereken seviyenin altında olduğunu göstermiştir (Bıçak ve Çakan, 2004; Güven, 2001).

Öğretmenlerin eksik ya da yetersiz bilgi ve beceriye sahip oldukları ölçme ve değerlendirme tekniklerini kullanmaları sakıncalı sonuçlara neden olmaktadır. Türkiye’de, öğretmenlerin ölçme ve değerlendirme yeterliklerinin araştırıldığı çeşitli çalışmalarda; öğretmenlerin yeni programda yer alan ölçme ve değerlendirme faaliyetlerini yapma konusunda sorunlar yaşadığı (Güven 2001; Bıçak ve Çakan, 2004), ölçme ve değerlendirme konusunda yeterince bilgilerinin olmadığı ve bu yüzden ölçme ve değerlendirmede eğitim ihtiyacı içinde oldukları (Gelbal ve Kelecioğlu, 2007), ölçme ve değerlendirmeye dönük belli konu alanlarında kendilerini yetersiz gördükleri (Yapıcı ve Demirdelen 2007; Gözütok, Akgün ve Karacaoğlu 2005) ve gelişime ihtiyaç duydukları, daha çok kendilerini yeterli hissettikleri ölçme yöntemlerini kullandıkları (Kahramanoğlu ve Ay, 2013) gibi sonuçlar elde edilmiştir. Bunun yanında Acat ve Demir (2010), tarafından yapılan çalışmada, öğretmenler alternatif ölçme ve değerlendirme araçları konusunda kendilerini genel olarak yeterli hissetmediklerini belirtilmektedir. Buna rağmen yapılan araştırmalara bakıldığında öğretmenlerin bir bölümü yetersiz oldukları durumlarda bile çeşitli ölçme ve değerlendirme yöntemlerini uygulamaktadırlar (Nolen, Haladyna ve Haas, 1992; Stiggins ve Conklin, 1992). Bunun neticesinde yararlı bir yöntem yanlış durum ve şartlarda kullanıldığı veya doğru zamanda yanlış kullanıldığı için yarar yerine zarara neden olabilmektedir. Bu zarar en başta sınava giren öğrenciler olmak üzere, öğretim programına, öğretim ortamına yönelik yanlış veya eksik kararların alınması biçiminde de gerçekleşebilmektedir (Çakan, 2004).

Sınıf içi ölçme ve değerlendirme faaliyetlerinde geleneksel ve yapılandırmacı yaklaşım olmak üzere iki yaklaşım kullanılmaktadır. Geleneksel ölçme ve değerlendirme yaklaşımı, yazılı ve sözlü yoklamalar, çoktan seçmeli, kısa cevaplı, eşleştirmeli ve doğru-yanlış gibi klasik test araçları kullanılan sürece çok

sonuca odaklı bir yaklaşımdır. Geleneksel ölçme ve değerlendirme yaklaşımının mevcut eğitim sisteminde üst düzey becerileri ölçmede yetersiz kaldığı görülmektedir (Kutlu, 2006). Geleneksel ölçme ve değerlendirme yaklaşımı öğrenciyi yargılayan, öğretmen merkezli, ürüne odaklı, geribildirim geç sağlayan, norm dayanaklı değerlendirme yapan, ezber ve hatırlamaya dayalı olan bir yaklaşımdır. Yapılandırmacı ölçme ve değerlendirme yaklaşımı ise performans belirleme, durum belirleme ve portfolyo gibi tamamlayıcı test araçlarını kullanan sonuçtan çok sürece odaklı bir yaklaşımdır. Yapılandırmacı yaklaşım ürünle beraber sürece odaklı, öğrenci merkezli, öğrencinin gelişimine katkı sağlayan, zamanında geribildirim sağlayan, ölçüt dayanaklı bir yaklaşımdır.

Geleneksel ölçme ve değerlendirme yaklaşımında öğrenci başarısının değerlendirilmesi, genellikle öğretim sürecinden ayrı ve daha çok ürüne ağırlık verecek bir biçimde ele alınmakta; bu amaçla daha çok seçmeli ve kısa cevaplı testlerle, yazılı ve sözlü yoklamalara önem verilmektedir. Yapılandırmacı öğrenme yaklaşımında ölçme ve değerlendirme ise, öğretim sürecinin bir parçasıdır ve sadece öğrenmenin başında ve sonunda değil, aynı zamanda öğrenme süreci boyunca her önemli noktada yer alır. Bu yaklaşımda sürecin ağırlıklı olması sebebiyle geleneksel yaklaşıma göre daha çok ve çeşitli ölçme araçlarının ya da yöntemlerinin kullanılmasını gerektirir. Geleneksel olarak uygulanan kağıt-kalem testleri ile birlikte, öğrencinin sınıf içi ve sınıf dışındaki davranışlarını izleyerek, süreç içindeki performansını gözleyerek, ilgisini ve tutumunu ölçerek ve öğrenciyi de değerlendirme sürecine katarak ölçme ve değerlendirmeyi geniş bir açıdan ele almaktadır. Böylelikle öğrenci performansını her yönüyle değerlendirebilmek mümkün olabilmektedir. Öğretmenlerin alışık olması sebebiyle yapılandırmacı yaklaşım yöntemlerinden daha çok geleneksel nitelikte olanları tercih edebilmekte; ancak diğerleri ile ilgili uygun araçlar geliştirilip, uygun zamanda kullanılması sağlanarak bunların da yaygınlaştırılması mümkün olabilir (Gelbal ve Kelecioğlu, 2007).

3. Geniş Ölçekli Ölçme ve Değerlendirme

Geniş ölçekli ölçme ve değerlendirme, büyük bir öğrenci popülasyonunu kapsayan bölgesel ulusal ya da uluslararası ölçekte gerçekleştirilen standart faaliyetler olarak tanımlanmaktadır (Simon, Ercikan ve Rousseau, 2013). Geniş ölçekli ölçme ve değerlendirme uygulamaları eğitim politikalarının kararlaştırılması sürecinde önemli bir yere sahiptir. Kirsch, Lennon, von Davier, Gonzalez ve Yamamoto (2013)'ya göre geniş ölçekli ölçme ve değerlendirme, belirli bir etki alanındaki bilgi, beceri araştırmaları, ya da davranışlardır. Bu faaliyetlerin amacı ise araştırılan bir evreni ya da evrenleri tanımlamaktır.

Ancak problem yalnızca çoktan seçmeli soruların uygulanmasında değildir. Seviye Belirleme Sınavı ve Yükseköğretime Geçiş Sınavı gibi geniş ölçekli sınavlardan testleri alan öğrenci gruplarının yetenek dağılımı ile uyuşmayan, özellikle fen ve matematik alanlarında son derece zor soruların kullanılması (Ayvacı, Keleş ve Çepni, 1999; Çepni, Ayvacı ve Keleş, 2001; Mutlu, Uşak ve Aydoğdu, 2003; Güler, Özek ve Yaprak, 2004; Karaman, Salar, Dilber ve Turgut, 2014) bu etkiyi daha da arttırmaktadır. Ayrıca buna ek olarak bu sınavlarda kullanılan sorular teknik olarak da kapsam geçerliği açısından kuşku vericidir (Berberoğlu, 2012). Tek bir soruyla neyin ölçülmeye çalışıldığı net olarak tanımlanmadıkça öğrencilerin hangi düşünme süreçlerinde daha iyi, hangilerinde desteğe ihtiyaçlarının olduğu gibi soruları da yanıtız kalmaktadır.

Test uygulamaları eğitim sisteminde bir amaç değil araçtır. Yapılan eğitim öğretim etkinliklerinin ölçüde başarılı olduğunu belirlemek ve buna dayanarak öğretim program ve yöntemlerinde gerekli düzenlemeleri yapmak test sonuçları ile mümkün olabilmektedir. Ancak ülkemizdeki geniş ölçekli ölçme ve değerlendirme faaliyetleri bu amaca hizmet etme kapasitesine sahip değildir. Özellikle öğrencilerin gelişimini izlemek amacı ile uygulanan Seviye Belirleme Sınavı'nda test planının yetersiz olması ve soruların ölçme ve değerlendirme ilkelerine uygun olmadan hazırlanması amaca hizmet etmeyen bir

uygulamanın devam etmesine neden olmuştur. Amaca hizmet etmeyen testler ölçme ve değerlendirme alanında geçerliği olmayan araçlar olarak değerlendirilmektedir (Berberoğlu ve İş Güzel, 2013).

Türkiye’de yapılan geniş ölçekli sınavlara bakıldığında (YGS, TEOG vs..) ortak madde seti kullanılmadığından dolayı karşılaştırma yapılamamaktadır. Bu tür karşılaştırmaların yapılması hatalı bir yorum olmaktan öteye geçememektedir. Ülkemizde bu sınavların yapılmasının hemen ardından televizyon programlarına katılan bazı eğitimciler, sınav sonuçlarını bir yıldan diğerine göre yorumlamaktadırlar. Ancak her yıl aynı güçlükte soruların hazırlanıp hazırlanmadığı bilinmediğinden, testleri alan öğrencilerin de benzer olup olmadığına dair herhangi bir kanıt olmadığından bu yorumlar tamamen geçersizdir. Sınava giren öğrenci kitlesinin benzer olduğu farz edilse bile hazırlanan soruların bir yıl öncesine oranla daha zor hazırlanması ortalamayı düşürecektir. Bundan dolayı testi alan öğrenci grubunun daha kötü olduğuna dair veya son yıl liseden mezun olan öğrencilerin daha kötü bir eğitim sürecinden geçtiğine dair yorum yapmak yanlış olmaktadır. Benzer biçimde daha düşük ortalamaya dayanarak öğrencilerin öğretim programlarındaki hedeflere ulaşamadıklarını söylemek de yanlıştır. Çünkü sınavlarda kullanılan soruların öğretim programındaki hedefleri ne derecede doğru ölçtüğü tespit edilememektedir.

Oldukça pahalıya mal olan ve büyük bir öğrenci grubuna uygulanan Yükseköğretime Geçiş Sınavı, öğrencilerin hangi düşünme süreçlerinde yetersiz hangilerinde daha iyi olduğuna dair bilgi vermemektedir. Yalnızca ortalama olarak çözülen soru sayıları gündeme gelmekte, düşük ortalamalar başarısızlık olarak yorumlanmaktadır (Berberoğlu ve İş Güzel, 2013). Türkiye’de yapılan geniş ölçekli ölçme ve değerlendirme faaliyetlerinde başka bir problem ise bu sınavların yalnızca seçme amacıyla kullanılmasından kaynaklanmaktadır. Çünkü gelişmiş ülkelerde yapılan geniş ölçekli ölçme ve değerlendirme uygulamalarına bakıldığında sıralamaya dayalı sınavların artık kullanılmadığı görülmektedir. Diğer gelişmiş ülkelerde ölçme ve değerlendirme alanının yıllar içinde gelişmiş olması da mevcut durum üzerinde etkili olmuştur (Berberoğlu ve İş Güzel, 2013). Yükseköğretime Geçiş Sınavı’nda üniversite programlarına yerleştirilen öğrencilerin hangi düşünme süreçlerinde olduğu ve yerleştirilmeyenlere oranla daha iyi olduğuna dair bir bilgi elde edilememektedir. Benzer biçimde kazanamayan öğrencilerin hangi becerileri geliştirip sınava tekrar girmeleri durumunda kazanma ihtimallerinin daha yüksek olacağı da bilinmemektedir. Öğrenciler sınava hazırlanmak amacı ile sürekli olarak soru çözmektedir. Ancak amaçsız yapılan bu egzersizlerin başarıyı arttırması bu şartlar altında mümkün değildir (Berberoğlu ve İş Güzel, 2013). Bu problemin çözümü için Demirtaşlı (2009) sınavda alınan puanların birbirini izleyen yıllar boyunca denkliklerinin sağlanmasının gerekliliğini, öğrenci öğrenmelerindeki gelişimi izleyebilecek niteliğe ulaştırılmasının sağlanması gerektiğini belirtmektedir. Bunun yapılabilmesi için uluslararası geniş ölçekli test uygulamalarında yaygın olarak kullanılan madde tepki kuramına dayalı ölçeklemelerden yararlanılması gerekmektedir. Çünkü ülkemiz uluslararası geniş ölçekli ölçme ve değerlendirme faaliyetlerine 1999 yılından beri düzenli olarak katılmakta olduğundan dolayı bu faaliyetlerdeki veri işlenerek öğrenci gelişimi ile ilgili sonuçlar elde edilebilir.

Ülkemizde yapılan geniş ölçekli sınavlarda başka bir problem de Berberoğlu ve İş Güzel’in (2013) belirttiği gibi bu sınavların uygulama yapılan kitlenin yetenek düzeyi ile eşleşmemesidir. Ülkemizde “sınav ne kadar zor olursa o kadar iyi ayırıcılığa sahip olur” düşüncesi her nedense herkes tarafından kabul görmektedir. Ülkemizde yapılan geniş ölçekli sınavların çok zor olduğu da alan yazında belirtilmiştir Çok zor ve çok kolay sınavlara dayanarak verilecek kararlarda her zaman çok daha fazla hata yapılır. Çünkü madde ayırıcılık indeksinin yüksek olması için zor, orta ve kolay güçlükte soruların bulunduğu bir test formunun olması gerekmektedir. En az hata sınav güçlüğü’nün testi alan kitlenin yetenek düzeyiyle eşleştiği durumda elde edilir. Sınava giren grubun yetenek düzeyine uygun olmayan bazı madde setlerine hiç yanıt verilmemesi (ki bu özellikle fen ve matematik alanlarında dikkati çekmektedir) öğrenci kitlesinden bilgi alınmasını engellemektedir. Boş bırakılma oranlarının yüksek olduğu bazı alt testlerde yanıtlanan soru sayısı daha da düşmektedir. Bu geçerliği zedeleyen çok önemli bir faktördür (Berberoğlu ve İş Güzel, 2013).

Geniş ölçekli ölçme ve değerlendirme faaliyetlerinde diğer bir konu da bu sınavların güvenilirliklerinin çok yüksek olmasının ön plana çıkmasıdır. Bu tür sınavların güvenilirliklerinin yüksek çıkması gayet doğaldır, çünkü çok geniş bir varyans mevcuttur (Berberoğlu ve İş Güzel, 2013). Ancak yüksek güvenilirlik yanıltıcıdır ve hiç bir şekilde testlerin geçerliğinin de yüksek olacağını garanti etmez. Özellikle çok düşük ve çok yüksek puanlı öğrencilerin puanları daha büyük oranda hata içermektedir. Bu düşük puanlılarda belki bir sorun teşkil etmez, ancak yüksek puanlı öğrenciler okullara ve üniversite programlarına yerleştirilmektedir. Yüksek puanlı öğrencilerin seçilmesi ve yerleştirilmesinde özellikle kesme puanları belirlenirken ciddi hataların yapıyor olması kaçınılmaz olmaktadır. Bu hata oranları hiç bir şekilde dikkate alınmamaktadır. Bunun bir nedeni söz konusu testlerde klasik test kuramının kullanılıyor olmasıdır. Bu kuram hatayı tüm puan aralıklarında aynı büyüklükte kabul etmektedir. Ancak bu çok büyük bir varsayımdır (Berberoğlu ve İş Güzel, 2013).

Ulusal geniş ölçekli sınavlarda bir diğer ölçme ve değerlendirme problemi de hazırlanan maddelerin yalnızca çoktan seçmeli formatında olmasından kaynaklanmaktadır. Çünkü ölçme işlemi farklı soru formatlarının kullanımı ile gerçekleşmektedir. Çoktan seçmeli maddelerin üst düzey becerileri ölçmemesi ve sürekli tekrarlardan dolayı nitelik olarak düşmesi bu sınavların geçerliklerini –özellikle kapsam geçerliğini- ciddi bir şekilde düşürmektedir. Bu problemin çözümü için Demirtaşlı (2009), ulusal ölçme ve değerlendirme uygulamalarında soru formatlarında çeşitlilik sağlanmasını ve bunların puanlamasıyla ilgili çözüm örnekleri incelenerek uygulamaya geçilmesinin gerekliliğini belirtmiştir. Ayrıca Demirtaşlı (2009), ulusal ölçme ve değerlendirme faaliyetlerinden alınan sonuçların sisteme açık ve kapsamlı geri bildirimler verecek biçimde yapılandırılmasının ve bu yolla puanların anlamlandırılmasının mümkün olabileceğini, böylelikle öğrenci başarısı “doğru yanıtlanan soru sayısı” olarak değil, yeterlikler çerçevesinde değerlendirilebileceğini belirtmiştir.

Berberoğlu vd. (2012) tarafından yapılan araştırmada ülkemizde yapılan geniş ölçekli sınavların kapsam geçerliği örnek sorular seçilerek incelenmiştir. Sınavların kapsam geçerliği incelenmeden önce testlerin amacının ne olduğunu, test planlarının ne ölçüde amaca yönelik hazırlandığını, hazırlanan planların test kapsamında ne ölçüde doğru yansıtıldığı gibi ölçütler dikkate alınmıştır. Araştırmacılar Seviye Belirleme Sınavı ile Yüksek Öğretime Geçiş Sınavlarını incelemişlerdir. Araştırmanın sonucuna göre geniş ölçekli sınavların geçerliklerinin olmadığı bir başka deyişle amaca hizmet etmediği belirtilmiştir. Bunun sonucunda da geniş ölçekli ölçme ve değerlendirme faaliyetlerine giren öğrencilerin aldıkları toplam sınav puanlarının belli bir anlama sahip olmadığı belirtilmiştir.

4. Sonuç

Sonuç olarak hem sınıf içi hem de geniş ölçekli ölçme ve değerlendirmede geçerlik amaca hizmet etme derecesi olarak değerlendirildiğinde bu sınavların amaca hizmet etmediği açıktır. Geçerliği olmayan sınavların (hem sınıf içi hem de geniş ölçekli olsun) sonuçları kullanılarak yapılacak olan değerlendirmeler sonucunda öğrenciler hakkında doğru kararların verilmesinin mümkün olmadığı görülmektedir. Okullarda yapılan ölçme ve değerlendirme çalışmalarının temel amacı öğrenmeleri geliştirmektir. Öğrenmeleri geliştirmek, ölçme sonuçlarının öğrenciye ve öğretmenlere geri bildirim sağlaması ile mümkün görünmektedir. Uluslararası yapılan sınavlarda ülkemiz sınav sonuçlarının geri bildiriminde diğer ülkelere göre çok düşüktür (EARGED, 2003). Dolayısıyla öğrenmeleri geliştirmeyi engellemektedir.

Dünyadaki diğer ülkelere bakıldığında yapılan geniş ölçekli ölçme ve değerlendirme faaliyetleri (örneğin PISA, TIMSS ve PIRLS gibi) madde tepki kuramını temel almakta iken, ülkemizdeki geniş ölçekli sınavlar hala klasik test kuramına dayanmaktadır. Klasik test kuramında her madde için hata eşit kabul edilmekte iken, madde tepki kuramında her maddenin hata miktarı farklı olabilmekte ve bu hata oranları için tedbirler alınmaktadır. Böylece karar verme süreci daha az hata içermektedir. Ancak ülkemizde karar sürecinde oluşan hata hiçbir şekilde dikkate alınmamaktadır.

Diğer bir sonuç olarak ülkemizdeki öğretmenlerin çok büyük bir kısmının ölçme ve değerlendirmeyi bilmemesinden dolayı öğrencilerin üst düzey bilişsel seviyelere ulaşamamasıdır. Öğretmenlerin madde hazırlama ve analiz etme gibi temel yeterlik alanlarının eksik olması sonucu kaliteli sorular hazırlanamamakta ve bundan dolayı belirlenen amaçlara ulaşılmamaktadır. Öğretmenlerin öğrenciye sadece not vermesi ve farklı ölçme ve değerlendirme faaliyetlerini sınıf ortamında kullanmaması sonucu öğrenciler hazıra alışmakta ve üst bilişsel seviyelere ulaşamamaktadır. Ayrıca ülkemizde geniş ölçekli sınavlarda toplumda sadece yapılan soru sayısı dikkate alındığı için sınıf içi ölçme ve değerlendirme faaliyetleri de bu doğrultuda devam etmekte ve öğrenci gelişimi izlenememektedir.

Sınıf içi ya da geniş ölçekli ölçme ve değerlendirmede öncelikli olarak test planının (belirtke tablosu gibi) hazırlanması gerekir. Test planı hazırlanırken bazı kavram/kavramlar çoktan seçmeli maddelerle değerlendirilebilir durumda iken bazıları değerlendirilmemektedir. Bu yüzden hem sınıf içi hem de geniş ölçekli sınavlarda yalnızca çoktan seçmeli maddelerin kullanılması sınavların geçerliklerini ciddi anlamda düşürmektedir.

Sonuç olarak amaç, gerek sınıf içi ölçme ve değerlendirme gibi gelişimi izlemek, gerekse geniş ölçekli ölçme ve değerlendirme gibi seçme yapmak olsun her iki durumda da kullanılan testlerin iyileştirilmesi ve farklı formatlarda dikkate alınması gerekmektedir. Test hazırlama profesyonel ve uzmanlık gerektiren bir süreçtir. Bu süreç, içinde program uzmanlarının, ölçme ve değerlendirme uzmanlarının, istatistiksel analiz yapan uzmanların yer aldığı karmaşık bir yapıdan meydana gelmektedir. Kalite kontrolü süreç içinde kaçınılmaz olarak dikkate alınması gereken bir işlemdir. Yapılan her işin en az iki kez kontrol edilmesi, uzmanların ortak çalışarak daha geçerli sınavlar hazırlaması yapılacak hataları azaltacak, sistemin gelişmesine olumlu etkiler sağlayacaktır.

Kaynakça

- Acat, M. B. ve Demir, E. (2010). İlköğretim programlarındaki alternatif değerlendirme yöntemlerinin uygulanmasında karşılaşılan sorunlara ilişkin sınıf öğretmenlerinin görüşleri. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi*, 29, 337-356.
- Ayva, Ö. (2010, Ekim). Sosyal bilgiler dersi öğrenme öğretme süreci ile ilgili öğrenci görüşleri. *In International Conference on New Trends in Education and Their Implications* (pp. 11-13).
- Ayvacı, H.Ş., Keleş, E. ve Çepni, S. (1999, Ekim). ÖSS'de sorulan fizik sorularının liselerde uygulanan öğretim etkinlikleri açısından irdelenmesi, *Türk Fizik Derneği 18. Fizik Kongresinde Sunulmuş Bildiri*, Çukurova Üniversitesi, Adana.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: Pegem Akademi Yayıncılık.
- Berberoğlu, G. (2006). *Sınıf içi ölçme değerlendirme teknikleri*. Ankara: Morpa Kültür Yayınları.
- Berberoğlu, G. (2010). Ölçme ve değerlendirmenin uygulamaya yansıyan boyutları. *Cito Eğitim: Kuram ve Uygulama*, 10, 10-14.
- Berberoğlu, G. (2012). Kapsam geçerliği. *Cito Eğitim: Kuram ve Uygulama*, 15, 10-16.
- Berberoğlu, G. ve İş Güzel Ç. (2013). Eğitim sistemimizdeki ölçme ve değerlendirme uygulamaları nasıl olmalıdır?. *Cito Eğitim: Kuram ve Uygulama*, 21, 9-16.
- Berberoğlu, G., Arıkan, S., Çalışkan, M., Ekinci, P., İpekçioğlu S. ve Sevgi, S. (2012). Geniş ölçekli sınavlarda kapsam geçerliği nasıl sağlanır?. *Cito Eğitim: Kuram ve Uygulama*, 17, 10-22.
- Bıçak, B. ve Çakan, M. (2004, Aralık). Lise öğretmenlerinin sınıf içi ölçme ve değerlendirme uygulamalarına dönük görüşleri. *Orta Öğretimde Yeniden Yapılanma Sempozyumu'nda sunulmuş bildiri*, Milli Eğitim Bakanlığı, Ankara.
- Çakan, M. (2004). Öğretmenlerin ölçme-değerlendirme uygulamaları ve yeterlik düzeyleri: İlk ve ortaöğretim. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 37(2), 99-114.
- Çepni, S., Ayvacı, H.Ş. ve Keleş, E. (2001). Okullarda ve lise giriş sınavlarında sorulan fen bilgisi sorularının Bloom taksonomisine göre karşılaştırılması. *Yeni Bin Yılın Başında Türkiye'de Fen Bilimleri Eğitimi Sempozyumu 1*, 144-150.
- EARGED (2003). *TIMSS 1999 ulusal raporu*. Milli Eğitim Bakanlığı, Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı.
- Gelbal, S. ve Kelecioğlu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 33, 135-145.

- Gökçe, E. (2002). İlköğretim öğrencilerinin görüşlerine göre öğretmenlerin etkililiği. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 35(1-2), 111-119.
- Gözütok, F. D., Akgün, Ö. E. ve Karacaoğlu, Ö. C. (2005). İlköğretim programlarının öğretmen yeterlilikleri açısından değerlendirilmesi. *Eğitimde Yansımalar*, 8, 17-40.
- Güler, N. ve Gelbal, S. (2010). Açık uçlu matematik sorularının güvenilirliğinin klasik test kuramı ve genellenebilirlik kuramına göre incelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri*, 10(2), 991-1019.
- Güler, G., Özek, N. ve Yaprak, G. (2004). 1999-2001 ÖSS fizik sınav sorularının bilişsel gelişim seviyelerinin incelenmesi, dersane ve liselerde sorulan soruların bilişsel gelişim seviyeleriyle karşılaştırılması, *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 8(2), 63-66.
- Güven, S. (2001, Haziran). Sınıf öğretmenlerinin ölçme ve değerlendirmede kullandıkları yöntem ve tekniklerin belirlenmesi. *10. Ulusal Eğitim Bilimleri Kongresi'nde sunulmuş bildiri*, Abant İzzet Baysal Üniversitesi, Bolu.
- Karaca, E. (2004). Öğretmen adaylarının ölçme ve değerlendirme yeterliklerine ilişkin likert tipi bir yeterlik algısı ölçeğinin geliştirilmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10, 15-29.
- Karaman, İ., Salar, R., Dilber, R. ve Turgut, Ü. (2014). YGS ve LYS sınavlarındaki fizik sorularının öğretim programı açısından ve Bloom taksonomisi bilişsel alan düzeyi açısından analizi. *Akademik Sosyal Araştırmalar Dergisi*, 2(6), 309-315.
- Kirsch, I., Lennon, M., von Davier, M., Gonzalez, E. & Yamamoto, K. (2013). On the growing importance of international large-scale assessments. In M. von Davier E. Gonzalez I. Kirsch & K. Yamamoto (Eds.), *The role of international large-scale assessments: perspectives from technology, economy, and educational research* (pp. 1-12). New York: Springer Dordrecht Heidelberg.
- Kutlu, Ö. (2006). Üst düzey zihinsel süreçleri belirleme yolları: yeni durum belirleme yaklaşımları, *Çağdaş Eğitim*, 335.
- Kutlu, Ö., Doğan, C.D. ve Karakaya, İ. (2014). *Ölçme ve değerlendirme performansa ve portfolyoya dayalı durum belirleme*. Ankara: Pegem Akademi.
- MEB (2008). *Öğretmen yeterlikleri, öğretmenlik mesleği genel ve özel alan yeterlikleri*. Ankara: Devlet Kitapları Müdürlüğü.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13-103). New York: Macmillan.
- Mutlu, M., Uşak, M. ve Aydoğdu, M., (2003). Fen bilgisi sınav sorularının Bloom taksonomisi'ne göre değerlendirilmesi. *G.Ü. Kırşehir Eğitim Fakültesi Dergisi*, 4(2), 87-95.
- Nolen, S. B., Haladyna, T. M. & Haas, N. S. (1992). Uses and abuses of achievement test scores. *Educational Measurement: Issues and Practice*, 11(2), 9-15.
- Özbaşı, D. (2009). *Sınıf öğretmenleri için öğrenci başarısını ölçme ve değerlendirme ile ilgili yeterlik göstergelerinin ve bunlara ilişkin algılarının incelenmesi*. Yüksek Lisans Tezi, Anlra Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Simon, M., Ercikan, K. & Rousseau, M. (2013). Introduction. In M. Simon K. Ercikan & M. Rousseau (Eds.), *Improving large-scale assessment in education theory, issues, and practice* (pp. 1-9). New York: Routledge.
- Stiggins, R. J. & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany: State University of New York Press.
- Sünbül, A. M. (1996). Öğretmen niteliği ve öğretimdeki rolleri. *Eğitim Yönetimi*, 2(1) 597-607.
- Şahin, Ç ve Karaman, P. (2013). Sınıf öğretmen adaylarının ölçme ve değerlendirmeye ilişkin inançları. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 28(2), 394-407.
- Turgut, M. F. ve Baykul, Y. (2014). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- Yapıcı, M. ve Demirdelen, C. (2007). İlköğretim 4. sınıf programına ilişkin öğretmen görüşleri. *İlköğretim Online*, 6(2), 204- 212.