14. BÖLÜM

ETİK, ADALET VE YASAL YÜKÜMLÜLÜKLER

ETHICS, JUSTICE AND LEGAL OBLIGATIONS

Doç. Dr. Asiye ŞENGÜL AVŞAR

ETİK, ADALET VE YASAL YÜKÜMLÜLÜKLER

Doç. Dr. Asiye ŞENGÜL AVŞAR
Recep Tayyip Erdoğan Üniversitesi

Öz

Ölçme ve değerlendirme bireylerin akademik başarıları, mesleki yeterlilikleri, gelişim düzeyleri hakkında kararlar verilmesi, kurumların eğitimsel etkililiklerinin ortaya konulması, eğitim politikalarının oluşturulması, sürdürülmesi veya iyileştirilmesi gibi birçok önemli süreçte belirleyici bir rol üstlenmektedir. Bu süreçlerde verilen kararların doğruluğu sadece geçerli ve güvenilir ölçme yapılmasıyla değil aynı zamanda güçlü bir etik anlayışla mümkün olmaktadır. Ölçme araçlarının geliştirilmesinden veya uyarlanmasından uygulanmasına, uygulanmasından puanlanmasına, puanlanmasından sonuçların duyurulmasına kadar tüm aşamalarda; ölçme ve değerlendirme uzmanları, eğitimciler, yöneticiler, sınav uygulayıcıları ve diğer paydaşların mesleki sorumluluklarının farkında olarak hareket etmeleri gerekmektedir. Bu kapsamda ölçme ve değerlendirme süreçlerinde bilimsel ilkelere bağlılık, yansızlık, hesap verebilirlik ve bireysel farklılıklara duyarlılık önemli etik ilkeler olarak öne çıkmaktadır. Etik ilkeler, ölçme ve değerlendirme süreçlerinde adalet, eşitlik, şeffaflık, gizlilik ve hesap verebilirliğin sağlanmasına rehberlik etmektedir. Ölçme ve değerlendirmede takip edilmesi gereken temel etik ilkeler; sınava giren bireylerin psikolojik, duygusal ve fiziksel güvenliğini korumayı, ilgili yasal düzenlemelere uygun hareket etmeyi, ölçme ve değerlendirme alanındaki mesleki yeterlilikleri sürdürmeyi, yalnızca yetkinlik alanı içinde hizmet sunmayı, geçerli uygulamaların anlaşılmasını desteklemeyi ve tüm mesleki sorumlulukları dürüstlük, bütünlük, özen ve adalet ilkeleri doğrultusunda yerine getirmeyi temel almaktadır. Kitabın bu bölümünde, ilk olarak ölçme ve değerlendirme süreçlerinde paydaşları bağlayıcı kılan etik sorumluluklara yer verilmiştir. Daha sonra ölçme ve değerlendirmede adalet kavramının temel bileşenleri olan yansızlık, kapsayıcılık, hesap verebilirlik ve şeffaflık unsurları açıklanmıştır. Son bölümde ise ölçme ve değerlendirme uygulamalarını düzenleyen yasal çerçeve, ilgili mevzuat hükümleri ve paydaşların yerine getirmesi gereken hukuki yükümlülükler incelenmiştir. Böylece ölçme ve değerlendirme süreçlerinin etik ve hukuki boyutlarıyla kapsamlı bütüncül olarak ele alınması amaçlanmıştır.

Anahtar Kelimeler

Etik Sorumluluk, Yansızlık, Kapsayıcılık, Hesap Verebilirlik, Yasal Yükümlülükler

ETHICS, JUSTICE AND LEGAL OBLIGATIONS

Assoc. Prof. Asiye ŞENGÜL AVŞAR
Recep Tayyip Erdoğan University

Abstract

Measurement and evaluation play a crucial role in numerous critical processes, such as making decisions regarding individuals' academic achievements, professional competencies, and developmental levels; demonstrating the educational effectiveness of institutions; and formulating, maintaining, or improving educational policies. The accuracy of decisions made during these processes depends not only on conducting valid and reliable assessments but also on a strong ethical mindset. At every stage from the development or adaptation of assessment tools to their administration, scoring, and the reporting of results measurement and evaluation specialists, educators, administrators, test administrators, and other stakeholders must act with an awareness of their professional responsibilities. In this context, adherence to scientific principles, unbiasedness, accountability, and sensitivity to individual differences emerge as key ethical principles in measurement and evaluation processes. These ethical principles guide the assurance of fairness, equity, transparency, confidentiality, and accountability within these processes. The fundamental ethical principles to be followed in measurement and evaluation are grounded in protecting the psychological, emotional, and physical well-being of examinees; complying with relevant legal regulations; maintaining professional competence in the field; providing services only within one's area of expertise; promoting an understanding of valid practices; and fulfilling all professional responsibilities in accordance with the principles of honesty, integrity, diligence, and fairness. This chapter of the book begins by addressing the ethical responsibilities that are binding upon stakeholders in measurement and evaluation processes. Subsequently, the elements constituting the core components of fairness in measurement and evaluation namely unbiasedness, inclusivity, accountability, and transparency are explained. In the final section, the legal framework governing assessment and evaluation practices, the provisions of relevant legislation, and the legal obligations to be fulfilled by stakeholders are examined. Thus, the aim is to address measurement and evaluation processes in a comprehensive and holistic manner, encompassing both their ethical and legal dimensions.

Keywords

Ethical Responsibility, Unbiasedness, Inclusivity, Accountability, Legal Obligations

1. GİRİŞ

Ölçme ve değerlendirme süreçlerinin tüm aşamalarında uluslararası kabul edilen belli standartların takip edilmesi önemlidir. Ölçme ve değerlendirme süreçleri, yalnızca takip edilmesi gereken adımları içeren teknik bir alan değil, aynı zamanda profesyonel ve etik bir sorumluluk alanıdır. Ölçme ve değerlendirmeye ilişkin standartlar; geçerlik ve güvenirlik kanıtlarının sağlanması, test geliştirme süreçlerinin paydaşları bilgilendirecek şekilde şeffaf bir biçimde planlanması, kapsayıcılığa dikkat edilerek ölçme araçlarının geliştirilmesi ve olası hataların giderilmesi şeklinde özetlenebilir. Ölçme ve değerlendirmede bir diğer konu ise kapsayıcılıktır. Kapsayıcı bir ölçme ve değerlendirme süreci için; eşit öğrenme ve eşit test fırsatları sağlanmalı, dilsel ve kültürel farklılıklara dikkat edilmeli, yanlılık içeren maddeler belirlenmeli ve bu maddelerden uzak durulmalı, özel gereksinimli bireyler için erişilebilir testler ve test düzenlemeleri yapılmalıdır. Ayrıca bilgisayar ve teknoloji tabanlı testlerin formatları ve kullanıcı arayüzleri evrensel tasarım ilkelerine uygun olarak düzenlenmelidir. Kapsayıcılığın yalnızca özel gereksinimli bireylerle ilgili bir konu olmadığı, ölçme araçlarında çeşitlilik sağlanarak bireylerin kendilerini ve yeteneklerini en iyi ifade edebilecekleri süreçler oluşturmayı da içeren çok daha geniş bir anlam taşıdığı unutulmamalıdır. Yasal yükümlülükler ölçme ve değerlendirme için önemli bir konudur. Örneğin, yüksek riskli testler, bireyler, okullar ya da kurumlar üzerinde ciddi sonuçlar doğuran ölçme araçlarıdır. Bu testlerin geliştirilmesinden uygulanmasına, puanlanmasından sonuçların duyurulmasına kadar tüm süreçlerde titizlikle çalışılması gerekmektedir. Her aşamada tüm paydaşları bağlayıcı yasal yükümlülüklerin bulunması, süreçleri güçlendirerek şeffaf ve hesap verebilir hale getirmektedir. Özetle, tüm paydaşları ilgilendiren ve önemli bir sorumluluk alanı olan yasal yükümlülükler, hesap verebilirliği ve şeffaflığı sağlayarak ölçme ve değerlendirme süreçlerini sağlamlaştırmaktadır. Kitabın bu bölümde etik bağlamında ölçme ve değerlendirme faaliyetlerinde bulunan kişilerin mesleki sorumlulukları detaylı bir şekilde açıklanmıştır. Ardından ölçme ve değerlendirmede adaletin bileşenleri olan ölçmede adalet, ölçmede yansızlık, ölçmede kapsayıcılık, ölçmede hesap verebilirlik ve şeffaflık konuları hakkında bilgi verilmiştir. Daha sonra ölçme ve değerlendirmede yasal yükümlülükler açıklanmıştır.

2. ÖLÇME VE DEĞERLENDİRMEDE ETİK

Etik; bir kişi veya grup tarafından kabul edilen ya da belli bir alanda uygun görülen ahlaken doğru davranış ilkeleri/kuralları olarak tanımlanmaktadır (*American Psychological Association [APA]*, n.d.-a). Örneğin Amerikan Eğitim Araştırmaları Derneği (*American Educational Research Association [AERA]*), Amerikan Psikoloji Derneği (*American Psychological Association*), Amerikan Sosyoloji Derneği (*American Sociological Association*), Dünya Tabipler Birliği (*World Medical Association*) gibi önemli kuruluşlar kendi mesleki etik kurallarını (*code of ethics*) yayınlamışlardır. Etik standartlar, eğitim araştırmacılarının etik davranışlarına ilişkin kurallarını belirler (AERA, 2011).

Bu bölümde Eğitimde Ölçme Ulusal Konseyi (*National Council on Measurement in Education [NCME])* ve Etik Kuralları Geliştirme Geçici Komitesi (*Ad Hoc Committee on the Development of a Code of Ethics*) tarafından 1995 yılında hazırlan "Eğitimde Ölçme Alanında Mesleki Sorumluluk Kuralları" (*Code of Professional Responsibilities in Educational Measurement*) hakkında bilgi verilmiştir. Hazırlanan bu kurallar 2023 yılında NCME Standartlar (*NCME Standards*) ve Test Kullanımı Komitesi (*Test Use Committee*) tarafından yeniden düzenlenerek kullanıma sunulmuştur (NCME, 2023).

Hazırlanan ve yeniden düzenlenen bu kurallar NCME (2023) üyelerinin;

- Psikolojik ve duygusal iyi oluşları da dâhil olmak üzere sınava giren tüm bireylerin sağlığını ve güvenliğini korumak,
- Mesleki faaliyetlerin yürütülmesiyle ilgili yerel, ulusal ve uluslararası yasalar hakkında bilgi sahibi olmak ve bunlara uygun davranmak,
- Eğitimde ölçme ve değerlendirmede mesleki yeterliliklerini korumak ve geliştirmek,
- Yalnızca kendi yeterlilik ve deneyim alanlarında ölçme ve değerlendirme hizmetleri sunmak ve mesleki yeterliliklerin tam olarak açıklanmasını sağlamak,
- Eğitimde geçerli ölçme ve değerlendirme uygulamalarının anlaşılmasını desteklemek,
- Eğitim hizmetleri sunan eğitim kurumları ve kuruluşları içinde en yüksek davranış standartlarına uymak ve mesleki açıdan sorumlu davranışı teşvik etmek,
- Tüm mesleki sorumlulukları dürüstlük, bütünlük, gerekli özen ve adaletle yerine getirmek beklentileri üzerine geliştirilmiştir.

Bu beklentilere göre *Eğitimde Ölçme ve Değerlendirme Alanında Mesleki Sorumluluk Kuralları* (NCME, 2023);

- Ölçme ve değerlendirme araçlarını geliştirmek,
- Ölçme ve değerlendirme araçlarını pazarlamak ve satışını yapmak,
- Uygun ölçme ve değerlendirme araçlarını seçmek,
- Ölçme ve değerlendirme süreçlerini yönetmek, ölçme sonuçlarını puanlamak,
- Ölçme sonuçlarını; yorumlamak, kullanmak ve açıklamak,
- Ölçme ve değerlendirme konusunda başkalarına eğitim vermek
- Programları değerlendirmek ve ölçme ve değerlendirme araçlarıyla araştırma yürütmek olmak üzere sekiz ana alanda belirlenmiştir.

Bu kuralların amacı, eğitimde ölçme ve değerlendirme faaliyetlerinde bulunan kişilerin mesleki sorumluluklarını tanımlamaktır. Kitabın bu bölümünde NCME (2023)[1] tarafından ölçme ve değerlendirme faaliyetlerinin sekiz ana alanında belirlenen mesleki sorumluluklar sırasıyla açıklanmıştır.

2.1. Ölçme ve Değerlendirme Araçlarını Geliştirenler ve Değerlendirme Hizmetlerini Verenlerin Sorumlulukları

Ölçme ve değerlendirme ürünleri ve hizmetlerini geliştirenler, en yüksek kalitede ölçme araçlarını geliştirme konusunda mesleki bir sorumluluğa sahiptir. Bu sorumluluklar sırasıyla aşağıda açıklanmıştır (NCME, 2023):

1. Ölçme ve değerlendirme ürün ve hizmetlerinin profesyonel, teknik ve yasal standartları karşılayacak şekilde geliştirilmesi sağlanmalıdır.
2. Ölçme ve değerlendirme ürün ve hizmetleri; cinsiyet, etnik köken, ırk, sosyoekonomik durum, özel gereksinimli olma, din, yaş veya ulusal köken gibi özelliklerden bağımsız olmalı, ölçülen yapılar tüm adaylar (testi alanlar) için benzerlik taşımalıdır. Diğer bir ifadeyle belirtilen bu özelliklerden bağımsız ölçme yapıldığına yönelik kanıtlar sağlanmalıdır.
3. Evrensel tasarım ilkelerini kullanarak ölçülen yapılar üzerindeki etkileri en aza indiren ve erişilebilirliği sağlayan ölçme ve değerlendirme araçları geliştirilmedir. Örneğin özel gereksinimli sınav katılımcıları için uygun düzenlemeler sağlanmalıdır. Başka bir örnek olarak ölçmenin amacı dil yeterliliğini ölçmek değilse değerlendirmede dilde sınırlı yeterliliğe sahip olanlar için gerekli düzenlemeler yapılmalıdır.
4. Test geliştiricilerin tutum, davranış ya da yargılarını etkileyebilecek gerçek veya potansiyel çıkar çatışmaları ilgili taraflara açıklanmalıdır.
5. Ölçme ve değerlendirme ürün ve hizmetlerinde yerel, ulusal ve uluslararası yasalara uygun olarak telif hakkıyla korunan materyaller kullanılmalıdır.
6. Teknik bir kılavuzda; güvenirlik, geçerlik, puanlama ve raporlama süreçleri ile ölçme ve değerlendirmenin diğer ilgili özellikleri hakkında güncel bilgiler de dâhil olmak üzere, ölçme araçlarının geliştirilmesi ve puanlanması için atılan adımlar hakkında detaylı bilgi sağlanmalıdır.
7. Teknik bilgiler, mümkün olduğunca paydaşlara uygun açıklamalarla iletilmelidir.

---

[1] Bu bölümde açıklanan tüm etik kurallar, NCME (2023) tarafından açıklanan kurallardır. NCME (2023) bu etik kuralların uygun atıf verme şartı ve eğitim amacıyla kullanılmak üzere herhangi bir izin alınmaksızın ve ücret ödemeksizin yayınlanıp çoğaltılabileceğini belirtmektedir [bkz: National Council on Measurement in Education. (2023). *Code of professional responsibilities in educational measurement* (2. baskı)].

8. Ölçme ve değerlendirme süreci boyunca; test kaydı, yönetimi ve puanların raporlaması için mekanizmalar geliştirmek de dâhil olmak kişisel olarak tanımlanabilir bilgilerin gizliliği ve güvenliği sağlanmalıdır.
9. Değerlendirme sonuçlarının uygun ve amaçlanan şekilde yorumlanmasını ve kullanılmasını destekleyecek puan raporları ve destek materyalleri geliştirilmelidir.
10. Değerlendirmenin amacı açık ve belirgin bir dille açıklanmalıdır. Değerlendirme sonuçlarının yanlış kullanımına ya da yanlış yorumlanmasına karşı test kullanıcıları uyarılmalıdır.
11. Test hazırlığı ve program destek materyallerinde, bir ölçme aracı ya da sonuçların kullanımı ve yorumlanması hakkında yanlış veya dayanağı olmayan iddialardan kaçınılmalıdır.
12. Değerlendirmelerde ya da değerlendirmeleri destekleyici materyallerde yer alan önemli yanlışlıklar, mümkün olan en kısa sürede düzeltilmelidir.

Bu standartlar özetlendiğinde ölçme ve değerlendirme süreçlerinde profesyonel bir bakış açısıyla geçerlik, güvenirlik kanıtlarının sağlanması, test geliştirme süreçlerinin şeffaf bir şekilde paydaşları da bilgilendirecek şekilde planlanması, kapsayıcılığa dikkat edilerek ölçme araçlarının geliştirilmesi ve eğer ölçme süreçlerinin herhangi bir aşamasında hata yapıldıysa gerekli düzeltmelerin gerçekleştirilmesi gerektiği vurgulanmaktadır. Türkiye'de ÖSYM başkanlığı tarafından yürütülen sınav süreçlerinde; test geliştirme, uygulama ve testin sonuçlarını değerlendirme (puanlama ve yerleştirme) aşamalarına ilişkin yukarıda belirtilen sorumluluklara bağlı kalınarak hareket edilmektedir.

2.2. Ölçme ve Değerlendirme Ürün ve Hizmetlerini Pazarlayan ve Satışını Yapanların Sorumlulukları

Testler ve diğer araçlar, puanlama, sınav hazırlık, danışmanlık ve sınav yorumlama hizmetleri gibi ölçme-değerlendirme ürün ve hizmetlerinin pazarlanması; bunları kullanmayı düşünenler için doğru, eksiksiz ve amaca uygun bilgilere dayanmalıdır. Sorumluluk ekipler arasında paylaşılmalıdır. Örneğin psikometristler değerlendirmelerin teknik özelliklerinin doğru şekilde temsil edilmesini sağlarken satış ve pazarlama çalışanları, müşterilerle etkileşime girmeden önce bu bilgiler hakkında bilgilendirilmelidir. Değerlendirme ürün ve hizmetlerini pazarlayan ve satışını yapan kişilerin mesleki sorumlulukları şunlardır (NCME, 2023):

1. Potansiyel alıcılara ölçme-değerlendirme ürünleri ve hizmetleri ile bunların önerilen kullanımları ve sınırlamaları hakkında doğru bilgiler sunmalıdır.

2. Potansiyel kullanıcılara; ölçme ve değerlendirme ürünlerinin ve hizmetlerinin bilinen yanlış kullanımları konusunda bilgi verilmeli ve yanlış kullanımdan nasıl kaçınılacağına dair önerilerde bulunulmalıdır.
3. Değerlendirme ürün ve hizmetlerini pazarlayan ve satışını yapanlar, sundukları ölçme-değerlendirme ürünleri ve hizmetleri ile bunların eğitimdeki uygun kullanımları hakkında güncel bir anlayışa sahip olmalıdır.
4. Doğru ve uygun bir seçim kararını etkileyebilecek, ölçme ve değerlendirme ürünleri ve hizmetleri hakkındaki bilgiler kasıtlı olarak saklanmamalıdır.
5. Ölçme ve değerlendirme ürünleri ve hizmetleriyle ilgili tüm iddialar, kamuya açık olarak sunulan bilgilerin geçerli yorumlarına dayalı olmalıdır.
6. Ölçme ve değerlendirme ürünleri ve hizmetleri satın alma konusunda nitelikli kullanıcılara eşit fırsat tanınmalıdır.
7. Ölçme ve değerlendirme ürünleri ve hizmetleri için makul ücretler uygulanmalıdır.
8. Ölçme ve değerlendirme ürünleri ve hizmetleriyle ilgili tüm geçerli ücretler, herhangi bir satın alma ya da kullanımdan önce, potansiyel kullanıcılara bildirilmelidir.
9. Ölçme ve değerlendirme ürün ve hizmetleri için gereken ücretleri ödeyemedikleri gerekçesiyle değerlendirme sonuçlarından doğan fırsat ya da yararlardan hiçbir bireyin mahrum bırakılmaması için çaba gösterilmelidir.
10. Ölçme ve değerlendirme ürün ve hizmetlerinin satışına ilişkin ölçütler belirlenmelidir. Örneğin bir değerlendirme ürün ve hizmetinin satışı; ölçme aracını uygun şekilde kullanma ve sonuçlarını uygun şekilde yorumlamada uzman kişilere yapılmalıdır.
11. Ölçme ve değerlendirme ürün ve hizmetlerinin kullanıcıları tarafından onaylandığını ima eden bilgiler, yalnızca kullanıcıların izniyle yayınlanmalıdır.
12. Başka bir kurum ya da kuruluştan resmi bir onay alınmadığı sürece, ölçme ve değerlendirme ürünleri ve hizmetlerinin söz konusu kuruluş tarafından onaylandığına dair iddialarda bulunulmamalıdır.
13. Bireylerin gerçek başarı düzeylerini yanlış yansıtan puanlar almasına neden olabilecek test hazırlık ürünlerinin ve hizmetlerinin pazarlaması yapılmamalıdır.

Ölçme ve değerlendirme ürün ve hizmetlerini pazarlayan ve satışını yapanların uymaları gereken bu kuralların tamamı ÖSYM tarafından dikkate alınmaktadır. ÖSYM testi alan bireylerin kişisel bilgilerini saklama konusunda son derece titiz hareket etmekte, sınavlarda görev alacak kişileri özenle seçmekte, sınavın işleyişinde sorunlara yol açan

ve/veya adayların sınavlarının iptal edilmesine yol açabilecek eylemlerde bulunan görevlilere çeşitli yaptırımlar uygulamakta ve eğitimde fırsat eşitliği ilkesini göz önünde bulundurarak sınav ücretlerini makul düzeyde tutmaya çalışmaktadır. ÖSYM kâr amacı gütmeyen bir kamu kurumudur. Dolayısıyla ÖSYM'de geliştirilen testlerin satış ve pazarlaması yapılmaz. ÖSYM yüksek standartlarda sınavlar hazırlayarak öğrencilerin ve adayların doğru ve adil şekilde değerlendirilmesine yönelik çalışmalarını sürdürmektedir.

2.3. Ölçme ve Değerlendirme Ürünlerini ve Hizmetlerini Seçenlerin Sorumlulukları

Eğitim ortamlarında kullanılmak üzere ölçme ve değerlendirme ürün ve hizmetlerini seçen ya da başkalarının bunu yapmasına yardımcı olan kişiler, ölçme ve değerlendirme araçlarının (ürün ve hizmetlerinin), kullanım amaçlarına uygun olduğundan emin olmak için önemli mesleki sorumluluklara sahiptir. Ölçme ve değerlendirme ürün ve hizmetlerini seçen kişiler aşağıda belirtilen unsurları asgari düzeyde yerine getirmeye çalışmalıdır (NCME, 2023).

1. Ölçme-değerlendirme ürünleri ve hizmetlerini değerlendirmek için gerekli kavramları anlayan paydaşlar sürece dâhil edilmelidir.
2. Amaçlanan kullanımlar için geçerli olacak değerlendirme stratejileri ve ölçme araçları hakkında kapsamlı ve titiz bir inceleme yapılmalıdır.
3. Ölçme ve değerlendirme sonuçlarını karşılaştırırken adil ve tutarlı ölçütler (örneğin dereceli puanlama anahtarları) kullanılmalıdır.
4. Ölçme araçlarını önerirken ya da seçerken teknik kalite ve fayda konusunda kanıtlanmış kamuya açık araçlara dikkate edilmelidir.
5. Satın alınması düşünülen ölçme araçlarının; yazarlarıyla, test yayıncılarıyla veya diğer kişilerle herhangi bir ilişki ya da bağlantı varsa açıklanmalıdır. Bu tür ilişkilerin ölçme araçlarının seçim sürecinin nesnelliğini etkileyebileceği durumlarda katılımdan kaçınılmalıdır.
6. Başkalarının ölçme-değerlendirme araçlarının seçimi sürecine uygunsuz etki ya da baskı girişimleri olduğunda bu durum derhal bildirilmelidir.
7. Karar vericilere ve testin gelecekteki olası kullanıcılarına; ölçme-değerlendirme aracının amaçlanan kullanım için uygunluğu, kullanımın olası sonuçları, sınava/teste girenlerin haklarının korunması, göreceli maliyetler, ölçme-değerlendirme sürecinin yürütülmesi için gereken materyaller ve hizmetler hakkında bilgi verilmelidir. Ayrıca değerlendirme sonuçlarının; bilinen sınırlılıkları, olası yanlış kullanımları ve yanlış yorumları hakkında bilgilendirme yapılmalıdır.

8. İncelenen ölçme-değerlendirme araçları ile ilgili tüm güvenlik politika ve usullerine/prosedürlerine (örneğin değerlendirme içeriğiyle ilgili gizlilik sözleşmeleri) uyulmalıdır.
9. Herhangi bir ölçme ve değerlendirme aracının; amacı dışında kullanılmasına veya yanlış uygulanmasına, puanlanmasına ya da yorumlanmasına karşı durulmalıdır.
10. Ölçme uygulamasından elde edilen puanların yapay biçimde yükselmesine neden olacak uygulamalardan uzak durulmalıdır.

Yukarıda listelenen "ölçme-değerlendirme ürün ve hizmetlerini seçenlerin eğitimde ölçme-değerlendirme okuryazarlığına sahip kişilerden oluşması, ölçme araçlarının seçiminde amaca uygunluğa dikkat edilmesi ve gizlilik kurallarına uyulması" gibi ilkelerin tamamı ÖSYM tarafından göz önünde bulundurulmaktadır. ÖSYM sınavlarda kullandığı testleri kendi öz kaynaklarıyla geliştiren bir devlet kurumudur. Test geliştirme aşamalarında soru hazırlama ihtiyacını gerçek ve tüzel kişiler ile kamu kurum ve kuruluşlarından hizmet alımı suretiyle karşılamaktadır (ÖSYM, 2011). Tüm bu süreçlerde en yüksek güvenlik koşullarını işe koşmakta, nitelikli ölçme yapma hedefine sıkı sıkıya bağlı hareket etmektedir.

2.4. Ölçme ve Değerlendirme Süreçlerini Yönetenlerin Sorumlulukları

Bireyleri ölçme ve değerlendirme süreçlerine hazırlayanlardan sınavların uygulanmasında doğrudan ya da dolaylı olarak görev alan öğretmenler ve yöneticilere, sınav evraklarını taşıyan kuryelerden sınav binalarında görev alan emniyet personeline ölçme süreçlerine dâhil olan tüm kişiler sınavların adil ve uygun bir şekilde uygulanmasını sağlamada önemli bir role sahiptir. Ölçme ve değerlendirme süreçlerini yönetenlerin sorumlulukları sırasıyla şu şekilde açıklanabilir (NCME, 2023):

1. Sınav uygulanmadan önce sınava girecek kişilere; sınavın amacı, kullanımı ve olası sonuçlarını içerek şekilde sınavın nasıl değerlendirileceği ya da puanlanacağı; sonuçların nasıl dosyalanacağı, sonuçlara kimin erişebileceği ve sonuçların nasıl ulaştırılacağı (elektronik posta, çevrimiçi ortam, posta yolu gibi) hakkında bilgi verilmelidir. Ayrıca sınavı alan kişilere; sınav öncesinde, sırasında ve sonrasında sahip oldukları haklara ilişkin bilgiler verilmelidir.
2. Ölçme ve değerlendirme süreçleri nitelikli (gerekli eğitimleri alan, lisans sahibi ya da sertifika sahibi) kişiler tarafından yürütülmelidir.
3. Ölçme-değerlendirme içeriğinin ve süreçlerinin güvenliğini sağlamak, kopya girişimlerini engellemek için ölçme-değerlendirme uygulamaları öncesinde, sırasında ve sonrasında gerekli önlemler alınmalıdır.

4. Ölçme-değerlendirme uygulamaları başlamadan önce uygulama süreçleri (kurallar/yönergeler) anlaşılmalıdır.
5. Bilgisayar tabanlı bir sınavın uygulamasından önce; sınav cihazlarının sistem gereksinimlerini karşıladığı, ağ bant genişliği gereksinimlerinin sağlandığı, gerekli tüm yazılımların yüklü olduğu ve mevcut tüm sistem hazırlık kontrollerinin yapıldığı kontrol edilmelidir.
6. Standartlaştırılmış testler, kuralları belli olan işleyiş ve koşullara göre uygulanmalıdır. Herhangi bir aykırı durum oluştuğunda ilgili kişiler bilgilendirilmelidir.
7. Uygun olan tüm adaylara sınav fırsatları sağlanmalıdır. Örneğin adaylar düşük performans göstereceği beklentisiyle sınavlardan dışlanmamalıdır.
8. Ölçme-değerlendirme uygulaması sırasında sonuçları geçersiz kılabilecek her türlü koşuldan kaçınılmalıdır.
9. Belgelenmiş özel gereksinimi ya da ihtiyaçları olan kişilere yönelik, ölçme-değerlendirme süreçlerinin yönetimi için makul ve izin verilen tüm düzenlemeler sağlanmalı ve kayıt altına alınmalıdır.
10. Ölçme-değerlendirme süreçleri bilgisayar tabanlı ölçme platformları üzerinden gerçekleştiriliyorsa bu uygulamalara aşina olmayan adaylara, ölçme-değerlendirmenin içeriğini ve biçimini tanımaları için test pratiği yapma imkânı sunulmalıdır.
11. Ölçme-değerlendirme uygulamasından önce ve uygulama sırasında belirlenen zamanlarda bireylere uygulama süreçleri ya da talimatları hakkında soru sormaları için uygun fırsatlar sağlanmalıdır.
12. Sınavı alan kişilerin mahremiyeti ve adil süreç hakları (ölçme-değerlendirme sürecinde kuralların adil işlemesi, kişinin hakkını savunabilmesi ve şeffaf, eşit bir uygulamadan geçmesinin güvence altına alması) korunmalıdır.
13. Bireylerin veya grupların gerçek başarı düzeylerini yansıtmayan puanlar almasına izin verecek ya da bunu teşvik edecek eylem ve koşullardan uzak durulmalıdır.

Sıralanan ilkelerin eksiksiz bir biçimde hayata geçirilmesinde ölçme ve değerlendirme süreçlerini yönetenlere sorumluluklar düşmektedir. Sınavı yöneten kurumlar ve kişiler ölçme-değerlendirme süreçlerini yüksek standartlarda yürütmeli, kapsayıcılığı gözeterek özel gereksinimli ya da özel ihtiyacı olan kişilere uygun ortamlar sağlamalı, bilgisayar tabanlı değerlendirmelerde bazı katılımcıların bilgisayara aşina olmamasından kaynaklanabilecek yanlılıkların önüne geçmeye dönük tedbirler almalıdır. Test süreçlerini yönetmede profesyonel bir kurum olan ÖSYM söz konusu

gereklikler doğrultusunda tüm sınav süreçlerini işletmektedir. Bütün sınav görevlileri; sınav öncesinde, sınav sırasında ve sınav sonrasında yapmaları gerekenler hakkında yönergelerle bilgilendirilmektedir. Benzer şekilde adaylara da sınavda uymaları gereken kuralların tamamı tebliği edilmektedir. Tüm bu süreçler kanun ve yönetmeliklerle güvence altına alınmaktadır (ÖSYM, 2025). ÖSYM kapsayıcılık ilkesi doğrultusunda özel gereksinimli adaylara yönelik destek hizmetleri (işaretleyici ve/veya okuyucu yardımı, ek süre ve ekipman desteği vb.) sunmaktadır.

2.5. Ölçme Sonuçlarını Puanlayanların Sorumlulukları

Sınav sonuçlarının sınava girenler, veliler, eğitimciler ve diğer paydaşlara doğru ve zamanında bildirilebilmesi için eğitsel değerlendirmeler uygun ve etkili şekilde puanlanmalıdır. Ölçme sonuçlarını puanlayan ve puanlara dayalı olarak rapor hazırlayan kişilerin mesleki sorumlulukları aşağıda sıralanmıştır (NCME, 2023):

1. Test kullanıcılarına ölçme sonuçlarının nasıl puanlanacağı hakkında (örneğin; raporlama takvimi, puanlama süreci, puanlama yaklaşımının mantıksal gerekçesi, teknik özellikler, kalite kontrol süreçleri, raporlama biçimi ve varsa bu hizmetlere ilişkin ücretler) eksiksiz ve doğru bilgi sağlanmalıdır.
2. Yapılandırılmış yanıt görevlerinde (açık uçlu maddeler gibi) puanlamada bütünlüğünü sağlamak için kalite süreçleri uygulanmalıdır. Bu kapsamda puanlamadan önce puanlayıcı nitelikleri kontrol edilmeli ve puanlayıcılara eğitim verilmeli, puanlama boyunca puanlamanın doğruluğu ve puanlayıcılar arası uyum izlenmeli ve puanlama sonrasında son bir değerlendirme yapılmalıdır.
3. Ölçme ve değerlendirme amaçlarıyla ilgisi olmayan faktörlerin, puanlama üzerindeki etkisi en aza indirilmelidir.
4. Planlanan puanlama ve raporlama hizmetinde ya da programında herhangi bir sapma meydana gelirse test kullanıcıları derhal bilgilendirilmelidir. Test kullanıcılarıyla bir çözüm üzerinde görüşme sağlanmalıdır.
5. Puanlara dayalı olarak yapılan çıkarımları etkileyebilecek hatalar belirlenirse düzeltilmiş puan sonuçları, sınav katılımcısına ya da test kullanıcısına mümkün olan en kısa sürede sağlanmalıdır.
6. Bireyleri tanımlayan bilgilerin gizliliği; yerel, ulusal ve uluslararası yasalarca belirlendiği şekilde korunmalıdır.
7. Değerlendirme sonuçlarının özeti; yalnızca yerel, ulusal veya uluslararası yasalar uyarınca bu bilgiye erişim hakkına sahip kişilere ya da puanlama hizmetlerini talep eden tarafça yetkilendirilmiş kişilere açıklanmalıdır.
8. Gerektiğinde değerlendirme sonuçlarına itiraz ve puanlamanın yeniden yapılabilmesi için adil ve uygun bir itiraz süreci oluşturulmalıdır.

Ölçme sonuçlarının puanlanması, puanlarının değerlendirilmesi ve yayınlanması ile ilgili yukarıda değinilen sorumluluklar ÖSYM'nin sınav süreçlerine de yön vermektedir. ÖSYM objektifliği ve kişisel bilgilerin gizliliğini sağlamaya azami gayret göstermekte, puanlama için hazırladığı veriler ile raporları içermesi gereken teknik detaylara göre şekillendirmektedir. ÖSYM test geliştiren, geliştirdiği testleri uygulayan, test sonuçlarını değerlendiren ve değerlendirme sonuçlarının raporlarını hazırlayan bir kurumdur. Adayların test puanları, Sınav Hizmetleri Daire Başkanlığı tarafından hesaplanmaktadır. Puanlamanın nasıl yapıldığı (düzeltme formülü uygulanması, ortaöğretim başarı puanın ağırlığı gibi) her bir sınav için test yönergesinde açıklanmaktadır. Puanlara itiraz yolu da açıktır. Adayların testlere verdikleri cevaplar bilgi işlem yöntemleriyle incelenerek gerektiğinde (kopya çekme durumu olmasında) bireysel ya da toplu olarak sınavlar iptal edilebilmektedir. Puanlamada ya da geliştirilen maddelerde hata olması halinde de itiraz yolu açıktır. Gerekirse soru iptali gerçekleştirilmekte, bu durum kamuoyu ile açık ve şeffaf olarak paylaşılmaktadır. Ayrıca ÖSYM hazırladığı detaylı sınav raporlarında sınavların istatistikleri (yıllara göre karşılaştırmalı olarak) ile sınav sonuçlarından elde edilen puanlara göre ulaşılan geçerlik ve güvenirlik kanıtlarını tüm paydaşlara sunmaktadır.

2.6. Ölçme ve Değerlendirme Sonuçlarını Yorumlayan, Kullanan ve Açıklayanların Sorumlulukları

Ölçme-değerlendirme sonuçlarının yorumlanmasında, kullanımında ve açıklanmasında geçerli çıkarımlar desteklenmeli; geçersiz çıkarımların engellenmesi sağlanmalıdır. Ölçme sonuçlarını yorumlayan, kullanan ve açıklayan kişilerin sorumlulukları aşağıda maddeler halinde açıklanmıştır (NCME, 2023):

1. Ölçme-değerlendirme sonuçları açıklanırken süreçteki sınırlılıklar göz önüne alınmalıdır. Elde edilen ölçümlerin kullanımının ortaya çıkaracağı sonuçların farkında olunmalıdır. Sonuçlar bilgiye dayalı, nesnel ve adil bir şekilde yorumlanmalı, kullanılmalı ve açıklanmalıdır.
2. Ölçme-değerlendirme sonuçlarını alan kişilere (adaylar, veliler, yasal temsilciler; öğretmenler, eğitimciler, eğitim kurumu yöneticileri, danışmanlar ve rehberlik personeli; işverenler, insan kaynakları yetkilileri [işe alım süreçlerinde]; akademik ya da profesyonel değerlendirme yapan kurumlar; yasal merciler ya da ilgili resmî kurumlar) yürütülen ölçme ve değerlendirme işleminin amaçları, sınırlılıkları, kullanımları ve puanların doğru şekilde yorumlanması için gerekli bilgiler sağlanmalıdır.
3. Puanların raporunu alan kişilere, raporlanan tüm puanların yazılı bir açıklaması yapılmalıdır. Bu açıklama doğru/uygun çıkarımları ve olası yanlış yorumları içeren anlaşılması kolay bir metin olmalıdır.
4. Ölçme ve değerlendirme sonuçları hedef kitleye zamanında açıklanmalıdır.

5. Ölçme-değerlendirme sonuçlarının yorumlanmasında kullanılan normların ya da performans standartlarının yeterliliği ile uygunluğu değerlendirilmeli ve açıklanmalıdır.
6. Ölçme-değerlendirme sürecine dâhil olan taraflar, değerlendirme sonuçlarının kendilerini nasıl etkileyebileceği konusunda bilgilendirilmelidir.
7. Eğitimle ilgili kararlar alınırken kişiler ya da programlar hakkında birden fazla kaynaktan bilgi almaya özen gösterilmeli, konuyla ilgili çeşitli bilgi türlerinden faydalanılmalıdır.
8. Ölçme-değerlendirme sonuçları hakkında; yanlış raporlar, asılsız iddialar, uygun olmayan yorumlar ya da bunların dışında, gerçeğe aykırı ve yanıltıcı diğer ifadeler kullanılmamalı; başkalarının da bu tür davranışlarda bulunması engellenmelidir.
9. Adaylar ve paydaşlar değerlendirme sonuçlarının arşivde saklanıp saklanmayacağı ya da ne kadar süreyle saklanacağı konusunda bildirilmelidir. İtiraz ve yeniden puanlama sürecine, adayların ve ilgili diğer kişilerin değerlendirme sonuçlarına erişme haklarına, ayrıca bu hakları nasıl kullanılabileceklerine ilişkin bilgiler açıklanmalıdır.
10. Ölçme-değerlendirme bilgileriyle ilgili açıkça görülen kötüye kullanımlar, süreçten sorumlu kişilere bildirilmelidir.
11. Ölçme-değerlendirme sürecine dâhil olan bireylerin ve kurumların gizlilik hakları korunmalıdır.

ÖSYM ölçme ve değerlendirme sonuçlarının yorumlanmasında, raporlanmasında ve saklanmasında dikkat edilmesi gereken yukarıdaki ilkelere bağlı kalmakta; düzenlediği sınavların sonuçlarını değerlendirirken takip ettiği yöntemleri açık bir şekilde paylaşmakta; sınav sonuçlarının arşivlenmesi, ölçme ve değerlendirme sonuçlarının zamanında açıklanması, kişisel bilgilerin korunması gibi konularda üzerine düşen sorumluluğu şeffaf bir şekilde yerine getirmektedir.

2.7. Ölçme ve Değerlendirme Konusunda Başkalarına Eğitim Verenlerin Sorumlulukları

Eğitimde ölçme-değerlendirme konusunda başkalarına eğitim verme süreci ister yükseköğrenimin, ister mesleki gelişimin, ister kamu politikası tartışmalarının, isterse de iş eğitiminin bir parçası olsun, bireyleri doğru ölçme uygulamalarını anlamaya ve testler ile test sonuçlarını seçici (eleştirel ve bilinçli) şekilde kullanmaya hazırlamalıdır. Ölçme-değerlendirme konusunda başkalarına eğitim veren ya da bilgilendiren kişiler sorumluluğunun bilincinde profesyonel kişiler olmalı, aşağıdaki ilkelere uygun davranmalıdır (NCME, 2023):

1. Eğitim verilen alanlarda yetkin/uzman olunmalı ve güncel gelişmeler takip edilmelidir. Ayrıca bunlar öğretim süreçlerine yansımalıdır.
2. Ölçme-değerlendirme hakkında eğitim verirken adil ve dengeli bir bakış açısı sağlanmalıdır.
3. Ölçme-değerlendirmenin prensipleri ve kavramları, hedef kitleye uygun olan teknik ayrıntılarla açıklanmalıdır.
4. Başkalarını herhangi bir özel ölçme-değerlendirme yöntemi, ürünü ya da hizmeti hakkında bilgilendirirken kanıtlanmış bilgiler ile kişisel görüşler açık bir biçimde ayrılmalıdır.
5. Eğitim-öğretimin konusu olan, belli bir ölçme-değerlendirme ürünü ya da hizmeti hakkında yapılacak değerlendirmeyi etkileyebileceği düşünülen herhangi bir mali çıkar söz konusu ise bu durum açıkça beyan edilmelidir.
6. İster değerlendirme isterse eğitim-öğretim amaçlı olsun, bir ölçme-değerlendirme uygulaması veya ölçme işleminden elde edilen sonuçların raporlanması, sınava giren kişiye/kişilere yönelik olumsuz etki taşıma ihtimali içeriyorsa bu duruma ilişkin riskler ilgili taraflara açıkça bildirilmelidir.
7. Eğitim-öğretim sürecinde kullanılan tescilli (yetkili kişilerce kullanılabilen ya da telif hakkı olan) ölçme-değerlendirme araçlarının ve materyallerinin güvenliği sağlanmalıdır.
8. Sorumlu ölçme ve değerlendirme uygulamaları modellenmelidir. Eğitim-öğretim, öğrencilerin eğitim ve kariyer hedefleriyle uyumlu hale getirilmelidir. Eğitim alanların, eğitimde ölçme ve değerlendirme konusundaki mesleki sorumluluklarını öğrenmelerine yardımcı olunmalıdır.
9. Politika yapıcılar, ebeveynler ve diğer vatandaşlar tarafından tartışılan ölçme ve değerlendirme konularına ilişkin, adil ve dengeli bakış açıları sağlanmalıdır.

ÖSYM'nin görev ve sorumlulukları arasında doğrudan ölçme-değerlendirme eğitimleri vermek yer almasa da ÖSYM kurumsal sorumluluğunun bir parçası olarak çeşitli kurum ve kuruluşlara ölçme-değerlendirme eğitimleri verebilmektedir (örneğin Karabük Üniversitesi Medya, 2025). Ayrıca düzenlediği hizmet içi eğitimlerle kurum çalışanlarının ölçme-değerlendirme alanındaki bilgilerinin güncel kalması sağlanmaktadır. Bunları yaparken yukarıda listelenen sorumluluklara göre eğitim verecek kişileri belirlemekte ve eğitim içeriğini şekillendirmektedir.

2.8. Programları Değerlendiren ve Değerlendirmelerle Araştırma Yürütenlerin Sorumlulukları

Değerlendirmelere dayalı araştırma yürütmek (program değerlendirme, eğitim müdahale araştırması ya da değerlendirmenin kendisini inceleme amaçlarıyla), değerlendirmelerin ve eğitim programlarının daha iyi anlaşılmasına yardımcı olan

önemli bir faaliyettir. Eğitim programlarının değerlendirilmesine katılan ya da değerlendirmelerle ilgili araştırma yürüten kişiler şu hususlara dikkat etmelidir (NCME, 2023):

1. Değerlendirme ve araştırma etkinlikleri; bilgiye dayalı, nesnel ve adil bir biçimde yürütülmelidir.
2. Gizlilik ve bilgilendirilmiş onam hakkı da dâhil olmak üzere, değerlendirme ya da araştırma çalışmasına katılanların haklarını korumak için geçerli standartlara uyulmalıdır.
3. Eğitim araştırması kapsamında insan katılımcılardan veri toplanacaksa araştırmayı yürütmeden önce Kurumsal Etik Kurul (*Institutional Review Board*) onayı ya da muafiyet belgesi alınmalıdır.
4. Yazarlar, test yayıncıları ya da değerlendirmeye dâhil olan diğer kişilerle ilişkileri varsa bu durumu açıklamalıdır. Bu tür ilişkilerin, araştırmanın ya da değerlendirmenin nesnelliğini etkilemesi ihtimal dâhilinde ise çıkar çatışması söz konusu olabilir. Çıkar çatışması oluşturabilecek bu tür durumlardan kaçınılmalıdır.
5. Değerlendirme veya müdahale araştırmalarında, programın/müdahalenin hedefleriyle uyumlu yapıları ölçen ölçme araçları seçilmeli veya geliştirilmelidir. Örneğin; bu ölçme araçları müdahalenin olası etkilerine karşı duyarlı olmalı, diğer bir ifadeyle yapılan müdahalenin muhtemel etkilerini doğru şekilde ortaya koyabilecek yeterlilikte olmalıdır.
6. Daha önce geliştirilmiş ya da yeni geliştirilen ölçme aracının kullanılması durumunda, çalışmadan elde edilen verilere dayanarak güvenirlik ve geçerlik bilgileri raporlanmalıdır.
7. Araştırma veya değerlendirme sonuçlarının, bir programın/müdahalenin etkinliğine dair geçerli göstergeleri sağlayabilmesi için tüm ölçme ve değerlendirme süreçleri güvence altına alınmalıdır.
8. Araştırma tasarımında ya da değerlendirme sonuçlarında, geçerliği düşürebilecek olası kaynakları en aza indirecek adımlar atılmalıdır. Çalışmanın sonuçlarında yanlılığa neden olabileceği bilinen faktörler (bir ölçme ve değerlendirme uygulamasının birden fazla kez yapılmasının öğrenme ya da uygulama üzerinde etkileri gibi) açıklanmalıdır.
9. Araştırmanın hem amaçlanan hem de amaçlanmayan sonuçları; adil, eksiksiz ve nesnel bir biçimde sunulmalıdır.
10. Başkalarına ait çalışma ve fikirler, eksiksiz ve uygun şekilde belirtilmelidir.
11. Araştırma sonuçları, çalışmanın sınırlılıkları dâhilinde nitelendirilmelidir.

12. Değerlendirme ve araştırma yaparken mümkünse konuyla ilgili farklı kaynaklardan toplanan bilgiler kullanılmalıdır.

Geniş kapsamlı araştırmalara hizmet edebilecek büyük bir veri havuzuna sahip olan ÖSYM yaptığı/ desteklediği araştırmalarda yukarıda açıklanan ilkelere bağlı hareket etmektedir. ÖSYM görev alanı nedeniyle belli sınırlılıklarla da olsa özellikle testlerin psikometrik özelliklerinin iyileştirilmesi üzerine çeşitli araştırmaları desteklemektedir (URL-1).

Ölçme araçlarının geliştirilmesinde, uygulanmasında, puanlanmasında değerlendirmesinde ve değerlendirme raporlarının hazırlanmasında takip edilen adımlar eğitimde ölçme ve değerlendirme süreçlerinin niteliğini doğrudan etkilemektedir. Bireyler hakkında doğru kararlar vermenin ön koşulu geçerli ve güvenilir sonuçlara ulaşmaktır. Ölçümlerin geçerliği ve güvenirliği kadar önemli bir nokta da ölçme-değerlendirme uygulamaları ile araştırma süreçlerinde etik ilkelere ne kadar uygun davranıldığıdır. Etik açıdan uygun ve nitelikli çıktılara ancak yukarıda belirtilen sorumluluklar çerçevesinde yürütülecek ölçme ve değerlendirme süreçleriyle ulaşılabilir.

3. ÖLÇME VE DEĞERLENDİRMEDE ADALETİN BİRLEŞENLERİ

Adalet kelimesi TDK'ya (t.y.-a.) göre; "*hak ve hukuka uygunluk, hakkı gözetme; yasalarla sahip olunan hakların herkes tarafından kullanılmasının sağlanması; herkese kendine uygun düşeni, kendi hakkı olanı verme*" anlamlarına gelmektedir. Eğitimde adalet kavramı, daha çok eğitim eşitliği (*equity*) olarak ifade edilmektedir. Eğitim eşitliği çocuklar ve yetişkinler için eğitime erişim ve fırsatların ne ölçüde adil ve hakkaniyetli olduğuyla ilgilidir. Bu anlamda eğitim eşitliği cinsiyet, sosyo-ekonomik durum, ikamet yeri (coğrafi konum), etnik köken, dil ve diğer özelliklere dayalı eşitsizliklerin azaltılmasına atıfta bulunan bir kavramdır (*United Nations Educational, Scientific and Cultural Organization*-UNESCO, 2008). Açıkça görüldüğü üzere eğitim eşitliğinin, birden çok bileşeni vardır. Ayrıca bu noktada adalet ve eşitlik (*equality*) kavramlarının aynı anlama gelmediğine dikkat edilmelidir. Eğitim eşitliği ile herkes kendine uygun düşeni, kendi hakkını alabilir, diğer bir ifadeyle adalet sağlanır. Oysa tek başına eşitlik kavramı imkân ve koşullarına, özel durumuna vb. faktörlere dikkat edilmeksizin herkese eşit muamelede bulunulmasını belirtir ki bu durum her zaman adaletli olmaz.

Eğitim bir sistem olarak düşünüldüğünde ölçme ve değerlendirme bu sistemin kalite-kontrol mekanizmasıdır (Turgut & Baykul, 2013). Bu başlık altında eğitimin kalite kontrol aşaması olarak tanımlanan ölçme ve değerlendirme süreçlerinde adalet ve adaletin bileşenleri olan yanlılık, kapsayıcılık, hesap verebilirlik ile şeffaflık konularında bilgiler verilmiştir.

3.1. Ölçmede Adalet

Eğitimde ölçme-değerlendirme süreçlerinde adalet, test adaleti (*test fairness*) kavramıyla açıklanabilir. Adalet/adil olma/adillik, bir ölçme aracının geliştirme/uyarlama, uygulama ve raporlaştırma olmak üzere tüm aşamalarında dikkate alınması gereken bir konudur (APA, 2004). APA (t.y.-b) adalet/adil olma (*fairness*) kavramını; farklı ırksal ya da kültürel özelliklere ve diğer çeşitli yaşam deneyimlerine sahip bireylerin test puanlarındaki sistematik varyansı (farklılığı) ortadan kaldırmak amacıyla testi alan kişilere hakkaniyetli ve adil muamele edilmesi olarak tanımlanmaktadır. Temelde teknik değil sosyokültürel bir konu olan adalet; testin tasarımında, uygulamasında ve puanlamasında kalite yönetimini ifade eder. Ölçme açısından adil bir test geliştirebilmek için içyapıya dayalı kanıtlar sağlamaya dönük yeterli çalışma yapılmalı, öğrenme fırsatlarında ve teste erişimde eşitlik sağlanmalı, hazırlanan test hedef içeriği yeterince kapsamalı ve yalnızca incelenen beceriye/yeteneğe odaklanmalıdır. Ayrıca ölçme sonuçlarında test uygulayıcısının geçmiş yaşantılarından kaynaklanan bir varyans söz konusu olmamalı ve test maddeleri ölçülmesi hedeflenen yapının dışında kalan bir özelliği ölçmemelidir (AERA, APA & NCME, 2014).

Testle yapılan ölçmelerin yanlı olmaması ile ilişkili olan ölçmede adalet kavramı, testin tüm gruplar için adil ve eşit fırsat sunmasını da içeren daha geniş bir kavramı ifade eder. Test adaleti; bireyler, gruplar ve ortamlar arasında karşılaştırılabilir içyapıya dayalı kanıtlarla elde edilir (Messick, 1998). AERA, APA ve NCME (2014) test adaletini tanımlarken; test sürecinde muamelede adillik, ölçme yanlılığının olmaması, ölçülen yapı(lar)a erişimde adillik ve bireysel test puanı yorumlarının amaçlanan kullanımlar için geçerli olması olmak üzere dört standart belirlemiştir.

Adalet/adil olma/adillik test sürecinde tüm adaylara eşit muamelede bulunulması anlamına gelir AERA, APA ve NCME (2014). Bu bakımdan ölçmede adalet, tüm adaylara yeterliliklerini gösterebilmeleri için gerekli olanaklar sunmayı ve ölçülen yapıdaki konumlarının doğru belirlenmesi için adayların karşılaştırılabilir fırsatlara sahip olmasını gerektirir. Adil muamele karşılaştırılabilir test koşulları (tüm test katılımcılarının testi eşit ve standart bir ortamda alabilmesi anlamına gelir), teste aşinalık kazandırma ve uygulama materyallerine eşit erişim, her sınav katılımcısının sınava hazırlanma fırsatına sahip olması gibi faktörleri içerir (APA, 2004; Messick, 1998). Adaletin sağlandığı bir ortamda; testler ve uygulama koşulları dikkatli bir şekilde standartlaştırılır, tüm sınav katılımcılarına ilgili alandaki performanslarını ve yetenek düzeylerini gösterebilecekleri fırsatlar sunulur (APA, 2004).

Adalet tüm ölçme-değerlendirme süreçleri için önemli olmakla birlikte yüksek riskli sınavlarda çok daha kritik hale gelmektedir. ÖSYM de yüksek riskli sınavlar yürüten bir kurum olduğundan ölçmede adalet kavramı ÖSYM'nin misyonunda kendine yer bulmuştur. ÖSYM misyonunu "*Uyguladığı sınavlarda geçerli, güvenilir ve adil ölçme, seçme ve yerleştirme yapan bir kurum olmaktır*" şeklinde belirtmektedir (URL-2).

3.2. Ölçmede Yansızlık

Eğitimde ve psikolojide temel amaç bireylere ait özelliklerin geçerli ve güvenilir ölçümlerine ulaşmaktır. Bunun için mümkün olduğunca hatalardan arınık ölçme yapılması gerekir. Bu noktada psikolojik ölçmeleri etkileyen problemleri araştıran ve bu problemleri azaltmaya çalışan bir alan olarak tanımlanan test kuramları (Crocker & Algina, 1986) önem taşır. Çeşitli test kuramlarıyla, çeşitli istatistiksel süreçler yürütülerek ölçme araçlarının farklı örneklemlere ya da farklı kültürlere uygulanmasında geçerli ve güvenilir sonuçlara ulaşılması hedeflenir.

Farklı uygulamalarda geçerli ve güvenilir sonuçlara ulaşmanın yolu ölçme araçlarının ölçmeyi hedeflediği özelliği, grubun çeşitli özelliklerinden bağımsız ölçmesinden geçer. Ölçme aracı, uygulandığı grup içinde yer alan alt grupların/örneklemlerin özelliklerine bağlı olarak bir gruba avantaj ya da dezavantaj sağlamamalıdır (Crocker & Algina, 1986). Diğer bir ifadeyle yanlı ölçme yapmamalıdır. APA (t.y.-c) yanlılığı (*bias*);

- Taraflılık (*partiality*) bir şeye karşı veya onun lehine eğilimli olma ya da yatkın olma,
- Tepki yanlılığı (*response bias*) ya da test yanlılığı (*test bias*) gibi herhangi bir eğilim veya tercih,
- Örnekleme, veri toplama ya da veri analizi sırasında ortaya çıkan sistematik hata (yanlı kestirimler, yanlı örnekleme),
- Ölçülen ya da hesaplanan bir niceliğin gerçek (doğru) değeri arasında bir sapma olması (örneğin ölçüm ya da hesaplama, ilgili unsurun temsili niteliğinde değildir) olacak şekilde tanımlamaktadır.

Ölçme aracındaki maddeler ölçülen özellik bakımından aynı yetenek düzeyinde bulunan fakat farklı alt grupların/örneklemlerin üyesi olan bir gruba avantaj ya da dezavantaj sağlamamalıdır (Zumbo, 1999). Örneğin ulusal çapta uygulanan bir genel yetenek sınavında metro istasyonlarının krokileri üzerinden uzamsal akıl yürütmeye yönelik bir madde yer alsın. Uygulama sonucunda aynı yetenek düzeyindeki öğrencilerden yaşadıkları şehirde metro olanların daha yüksek başarı gösterdiği belirlenmiş olsun. Bu örnekte, metrosu olan bir şehirde yaşayan öğrencilerin aynı yetenek düzeyinde yer alan fakat yaşadıkları şehirde metro bulunmayan öğrencilere kıyasla avantajlı olduğu görülmektedir. Burada ölçülmek istenen yapının dışında başka bir faktörün (şehirde metro olması), ölçme sonuçlarına karışarak yanlılık oluşturduğu düşünülebilir. Bir ölçme yapıldığında ölçülmek istenen yapının dışında kalan özellikler (cinsiyet, ırk, sosyoekonomik düzey gibi) aynı yetenek düzeyindeki bireylerden bir gruba avantaj ya da dezavantaj sağlamamalıdır. Aksi takdirde sistematik hata devreye girer ve yanlılık olarak nitelendirilen bu durum ölçme sonuçlarının geçerliğini zedeler; daha açık bir anlatımla ölçme sonuçlarına dayalı çıkarımların, verilecek kararların ve ulaşılacak sonuçların hatalı olmasına yol açar. Bu anlamda ölçme sonuçlarından hareketle yapılan çıkarımların doğruluğu, ölçümlerin yansız olmasına bağlıdır.

Kültürlerarası karşılaştırmaların yapıldığı test uyarlama çalışmaları ya da farklı kültürlerden çok sayıda bireyin aynı testleri aldığı ve ülkelerin akademik başarılarının göstergesi olarak değerlendirilen testlerde (örneğin PISA, TIMSS, PIRLS gibi geniş ölçekli başarı testleri), kültürler arası çalışmalar yapılması hedeflendiğinden yanlılık daha da dikkat çekici bir konu haline gelir. Yapılan ölçmenin kültürden, ölçme aracının dilinden bağımsız olması geçerli sonuçlara ulaşmada kritik bir rol üstlenir. Farklı kültürler için ölçülen yapının aynı kalması ölçme değişmezliği (*measurement invariance*) olarak tanımlanır (Şengül Avşar & Emons, 2021).

Ölçme değişmezliğini ve yanlılığı belirlemenin çeşitli istatistiksel yolları vardır. Bu yollar içinde Değişen Madde Fonksiyonu (*Differential Item Functioning*-DMF) ve Çoklu-Grup Doğrulayıcı Faktör Analizi (*Multi-Group Confirmatory Factor Analysis*-ÇGDFA) sıklıkla dikkate alınan yaklaşımlardır (Stark ve ark., 2006; Vandenberg & Lance, 2000; Zumbo ve ark., 2003). DMF bir maddenin kültürler arasında ya da farklı bir grubun üyesi olmanın bir avantaj sağlayıp sağlamadığını belirler (Ercikan ve ark., 2004). Çeşitli istatistiksel yöntemlerle maddelerin DMF içerip içermediği belirlenir. Bunun için testi alan gruplar, referans ve odak grupları olarak iki ayrılırlar ve uygun istatistiksel yöntemlerle maddelerin DMF içerip içermediği incelenir (Arıkan ve ark., 2016).

Bir maddenin DMF içermesi o maddenin yanlı ölçme yaptığının tek başına kanıtı olmaz. Yanlılık çalışmalarında ilk aşama maddelerin DMF içerip içermediğinin istatistiksel olarak test edilmesidir. İkinci aşamada DMF içerdiği belirlenen maddeler uzman görüşlerine sunulur ve bu maddelerin gerçekten yanlı olup olmadığı uzmanlar tarafından yapılan detaylı incelemeler sonrasında netleşir (Dorans & Holland, 1992). Bu noktada uzmanların niteliği de önem taşır. İnceleme yapacak uzmanların test geliştirme süreçlerine hâkim ve testin ölçtüğü konu ya da yapı hakkında bilgi sahibi olmaları gerekir (Ayan, 2024). Bir maddenin yanlı olarak etiketlenmesi (*flagged*) için madde içeriğinde testle ölçülmesi hedeflenmeyen bir özelliğin ölçülmesi gerekmektedir.

Bir maddede DMF iki farklı şekilde belirebilir. İlki DMF'nin test maddesiyle ölçülmek istenen temel yetenek bakımından gruplar arasında gerçek farklılıklar olması nedeniyle var olduğu madde etkisi (*item impact*) durumudur. Burada yanlılık söz konusu olmayıp DMF'ye kaynaklık eden ölçülen özellik bakımdan var olan gerçek farklılıktan bahsedilmektedir. Diğeri ise madde yanlılığı durumudur. Bu durum, DMF'nin test maddesinin ölçülmek istenen temel yetenek bakımından (ve dolayısıyla testin amacıyla) ilgili olmayan bir özelliği nedeniyle var olduğu durumlar olarak ifade edilmektedir (Zumbo, 2007).

Ölçme değişmezliğini belirlemenin bir diğer yolu olan ÇGDFA, ölçülen yapının farklı örneklemler arasında benzerliğini test etmeye dönük istatistiksel bir yöntemdir (Brown ve ark., 2017). Bu yöntemde iç içe geçmiş dört hiyerarşik model test edilir. Bu modeller

sırasıyla; biçimsel değişmezlik (*configural invariance*), metrik değişmezlik (*metric invariance*), skaler değişmezlik (*scalar invariance*) ve katı değişmezlik (*strict invariance*) olarak isimlendirilir (Vandenberg & Lance, 2000.) Kurulan bu modellerden elde edilen uyum iyiliği istatistikleri arasındaki farklar incelenerek ölçülen yapının gruplar arasında değişip değişmediği test edilir. Ancak katı değişmezliği sağlamanın zor olduğu, eğitim ve psikoloji araştırmalarında katı değişmezliği araştırmaya gerek olmadığı da belirtilmektedir (Bowen & Masa, 2015). Katı değişmezliğin sağlanmadığı durumda kısmi değişmezliğin (*partial invariance*) incelenmesi önerilmektedir (Bryne, 2016; Tse ve ark., 2024).

Özetle bir testi geliştirirken, uyarlarken, uygularken ve puanlarken hata kaynaklarını ve ölçmede hata türlerini dikkatli bir şekilde ele almak gerekir. Ölçme sonuçlarının yanlılık içermesi, geçerliği düşürerek yapılacak çıkarımları olumsuz etkiler. Yanlı maddeler, ölçme yapılan gruplardan birine avantaj sağlayarak test adaletini de zedeler. Bu nedenle geçerlik çalışmaları kapsamında madde bazında DMF incelemelerine ve test bazında ÇGDFA çalışmalarına mutlaka yer verilmelidir. Böylece sonuçların yanlı olmadığına yönelik kanıtlar sunulup ölçümlerin geçerliği desteklenir.

3.3. Ölçmede Kapsayıcılık

UNESCO, *Education* 2030 girişimiyle herkes için kapsayıcı, eşitlikçi ve kaliteli eğitim (*Education 2030: Inclusive, equitable, and quality education for all*-EFA) programını yürütmektedir. EFA programı, sürdürülebilir kalkınma hedefleri çerçevesinde eğitim alanında dünyayı yönlendirmektedir (URL-3). UNESCO (2015) Eğitim 2030 Incheon Deklarasyonu ve Eylem Çerçevesi programında; kapsayıcı ve eşitlikçi kaliteli eğitimin sağlanması ve herkes için yaşam boyu öğrenme fırsatlarının teşvik edilmesi için Sürdürülebilir Kalkınma Hedefi 4'ün hedeflerini düzenlenmiştir. Tüm bireylerin kapsayıcı, eşit ve kaliteli eğitim almasını ve yaşam boyu öğrenme fırsatlarına erişmesini sağlamayı amaçlamayan bu hedefler; çocukların temel eğitimi tamamlamalarını, gençlerin ve yetişkinlerin teknik, mesleki ve yükseköğretime eşit erişimlerini ve istihdam için gerekli becerileri kazanmalarını, özellikle özel gereksinimli, yerli (*indigenous*) ve kırılgan grupların eğitim haklarının güvence altına alınmasını, güvenli ve kapsayıcı öğrenme ortamlarının oluşturulmasını amaçlamaktadır. Kapsayıcı eğitimin sağlanması, eğitimin tüm süreçlerinin kapsayıcılığa uygun olarak tasarlanmasıyla mümkün olur. Bunun için;

- Politika ve planlamaların eğitimde kapsayıcılığı ve eşitliği sağlayacak şekilde düzenlenmesi,
- Öğretmen eğitiminde öğretmen adaylarının kapsayıcı eğitimin bir parçası olarak nitelikli biçimde yetiştirilmesi ve görevde olan öğretmenler için bunun hizmet içi eğitimlerle desteklenmesi,

- Eğitim programlarının ve eğitim materyallerin tüm öğrenciler için uygun, erişilebilir ve farklılıklara duyarlı bir şekilde hazırlanması,
- Ölçme ve değerlendirme uygulamalarının tüm bu aşamalara uygun olarak tasarlanması gerekir.

Ölçme ve değerlendirme süreçlerinin kapsayıcı olması için test geliştirme/uyarlama, uygulama ve raporlama süreçleri kapsayıcılığa uygun şekilde tasarlanmalıdır. Bu kapsamda testlerin farklı formatlarda uygulanması, sonuçların adil şekilde hesaplanması, kapsayıcı eğitimin hedeflerine ulaşılıp ulaşılmadığının belirlenmesine yönelik verilerin toplanması ve analizi ile eğitim süreçlerinin izlenmesi gibi işlemler yürütülebilir.

Kapsayıcı değerlendirmeler, başlangıçta özel gereksinimli öğrenciler gibi dezavantajlı öğrenci gruplarının gereksinimlerini karşılamaya odaklanmıştır. Bununla beraber kapsayıcı değerlendirme tasarımı giderek genişlemiş, daha iyi bir eğitim uygulaması olarak kabul edilmiş ve tüm öğrencilere fayda sağlamayı hedeflemiştir (Advance HE, 2019). Kapsayıcı değerlendirmeler öğrencilerin değerlendirilme biçimleri aracılığıyla dışlanma, göz ardı edilme ve/veya dezavantajlı duruma düşme olasılığını en aza indirmeye çalışmaktadır (Oxford University Centre for Teaching and Learning, 2025). Thurlow ve arkadaşları (2016) yayınladığı raporda özel gereksinimli öğrencilerin yanı sıra İngilizce öğrenenler ve özel gereksinimli İngilizce öğrenenler için tüm değerlendirme türlerinde dikkate alınabilecek bir takım kapsayıcılık ilkelerine yer vermiştir. *University of Bristol, Centre for Teaching and Learning* (2025) dört temel kapsayıcı eğitim ilkesini şöyle açıklamaktadır:

1. **Çeşitli değerlendirme yöntemleri, seçenekleri ve konuları sunmak:** Bir program genelinde yapılacak kapsayıcı değerlendirme, öğrencilerin güçlü yönlerini ortaya koymalarına olanak tanımalı ve dezavantajlarını en aza indirecek çeşitli yöntemler, seçenekler ve konular sunmalıdır. Diğer bir ifadeyle tek tipte bir ölçme süreci takip edilmemelidir. Bir program boyunca tüm öğrencilerin öğrendiklerini gösterebilmelerine imkân tanıyacak çeşitli değerlendirme yöntemleri tasarlanmalıdır. Örneğin, öğrencilerin etkili iletişim becerilerini göstermeleri gerekiyorsa öğrencilere çeşitli seçenekler (poster, kayıtlı sunum, blog gibi) sunulmalı, iletişim kurmada kendilerini en güvende ve rahat hissettikleri ortamı seçmelerine izin verilmelidir. Kapsayıcı değerlendirme yaklaşımlarında öğrencilerin geçmişleri de değerli bir kaynak olarak görülür. Öğrencilere kendi deneyimlerinden yararlanarak anlayışlarını/kavrayışlarını ortaya koyma fırsatları verilir. Konu seçme olanağı sağlamak (örneğin; proje, kompozisyon ya da sınav sorusu için) öğrencilere odak noktalarını yönetebilme fırsatı sunar. Ayrıca öğrencilerin ilgi duydukları alanlarda çalışmalarına imkân verir.

2. **Açık, adil ve şeffaf değerlendirme ölçütlerini sağlamak:** Kapsayıcı değerlendirme yaklaşımlarında; açık, adil ve şeffaf değerlendirme ölçütlerinin kullanılması gerekir. Bu ölçütler sade bir dille yazılmalı, öğrencilerin öğretim elemanları ve akranlarıyla tartışmalarına olanak tanımalıdır. Örnek olarak öğrencilerin (a) ölçütleri birbirleriyle tartışması ve bunları değerlendirme yönergesiyle ilişkilendirmesi, (b) kendi çalışmalarını ölçütlere göre gözden geçirmesi, (c) örnek çalışmaları ölçütler çerçevesinde inceleyip tartışması, (d) birbirlerinin biçimlendirici (*formative*) değerlendirmelerini ölçütlere göre gözden geçirip geri bildirim vermesi ve (e) aldıkları geri bildirimleri ölçütlerle eşleştirmesi gösterilebilir. Bu çalışmalar hem yaparak yaşayarak öğrenme süreci hem de ölçütlerin nasıl kullanılacağının gösterilmesi açısından önemlidir.
3. **Öğrencileri değerlendirmeye hazırlamak:** Öğrencileri değerlendirmeye hazırlamak, kapsayıcı tasarımın ayrılmaz bir parçasıdır. Bu hazırlık; uygulama görevlerinin hem iyi hem de yetersiz örneklerin kullanımını ve yukarıda açıklandığı üzere bu örneklerin değerlendirme ölçütleriyle karşılaştırıldığı çalışmaları içerebilir. Ayrıca öğrencilerin güçlü yönlerini vurgulayan ve gelişim yollarını destekleyen düzenli biçimlendirici geri bildirim fırsatlarını da içine almalıdır. Kapsayıcı geri bildirim; metin, ses veya ekran kaydı gibi çeşitli formatlarda sunulabilir. Temel nokta geri bildirimlerin zamanında, yapıcı ve açık bir şekilde iletilmesi, öğrencilerin özgüvenini geliştirmeye ve başarıya ulaşmalarını desteklemeye hizmet etmesidir.
4. **Öğrencinin iş yükünü göz önünde bulundurma ve değerlendirme takviminde oluşabilecek çakışmalardan kaçınma**: Değerlendirme tasarımları, öğrencilerin sıkılmadan derinlemesine öğrenmelerini sağlamalıdır. Bunun için tüm program genelindeki değerlendirme yoğunluğu hesaba katılmalıdır. Program boyunca yapılan değerlendirmeler, öğrencilerin her bir değerlendirmeye hazırlanabilmeleri için yeterli zamanı sunmalıdır. Ayrıca değerlendirme görevlerinin teslim tarihleri arasında çakışmaların önlenmesi önemlidir.

Kapsayıcı değerlendirme tasarımı için Tablo 1'de verilen kontrol listesi kullanılabilir.

Tablo 1

Kontrol Listesi (University of Guelph, Office of Teaching and Learning'ten (2025) uyarlanmıştır)

Öğrenme Çıktıları ve Öğretimle Uyum:	
☐	Değerlendirmelerin içeriği ve biçimi, derste öğretilen temel konuları ve becerileri doğru şekilde yansıtmıştır.
☐	Değerlendirmelerin içeriği ve biçimi, doğrudan dersin öğrenme çıktılarıyla ilişkilendirilmiştir.
Genel Kurs/Ders Değerlendirme Tasarımı:	
☐	Ders içinde farklı değerlendirme türleri kullanılmıştır.
☐	Yüksek riskli değerlendirmeler, birden fazla bileşene ayrılmış veya yüksek riskli sınavların denemesi niteliğinde olan düşük riskli değerlendirmeler sunularak aşamalı hâle getirilmiştir.
☐	Yüksek riskli değerlendirmelerde gereken içerik ve becerilerle ilgili açık uygulama imkânı sağlanmıştır (Bunun için notlandırmaya dayanmayan veya düşük riskli değerlendirmeler kullanılmalı, gelişimi destekleyen geri bildirimler zamanında verilmelidir).
☐	Değerlendirmelerle ilişkili iş yükü dikkate alınmıştır.
Değerlendirme Süreçleri:	
☐	Öğrencilere, öğrenmelerini nasıl gösterecekleri konusunda seçenekler sunulmuştur (Örneğin herkesin son bir makale teslim etmesi yerine; öğrencilere sunum, podcast, infografik ya da poster teslim edebilme alternatifleri sunulmalıdır).
☐	Değerlendirme beklentileri değerlendirme öncesi açık bir şekilde paylaşılmıştır.
☐	Dereceli puanlama anahtarları (rubrikler) ve örnek ödevler kullanarak mükemmel çalışmanın nasıl göründüğü açıklanmıştır.
☐	Not vermek için net ölçütler belirlenmiştir.
☐	Mümkün olduğunda öğrenci ödevleri anonim (gizli) olarak değerlendirilmiştir.
Ölçme ve Değerlendirme İçeriği:	
☐	Ölçme aracında yer alan maddeler/sorular; farklı kültürlerden ya da geçmişlerden gelen öğrenciler ile ailesinde daha önce üniversite mezunu olmayan öğrenciler için belirsiz ya da bilinmeyen kaynaklardan uzak olarak hazırlanmıştır.
☐	Ölçme ve değerlendirme içerikleri, senaryolar ve sorular aracılığıyla farklı bakış açılarının temsil edilmesini sağlamıştır.

AERA ve arkadaşları (2014) yayınladığı Eğitim ve Psikolojik Testler için Standartlarda (*Standards for Educational and Psychological Testing*) eğitimde ölçmede kapsayıcılık için çeşitli standartlar tanımlamıştır.

Standart 3.9. "Test geliştiricileri/kullanıcıları, uygun ve mümkün olduğunda, adayların hedeflenen yapılar üzerindeki yeterliklerini ortaya koymalarını engelleyebilecek, ölçülen yapıyla ilişkisiz faktörleri ortadan kaldırmaya yönelik sınav uyarlamaları geliştirmekten ve sağlamaktan sorumludur" şeklinde ifade edilmektedir. Bu standart ile öğrencilerin ölçülen yapıyla ilgili olmayan engeller nedeniyle dezavantaj yaşamalarının önüne geçilmesini sağlayacak sınav uyarlamalarının (*test accommodations*) yapılması beklenmektedir. Bunun için sınav ortamı, sunum biçimi, etkileşim/katılım, yanıt verme şekli gibi alanlarda değişiklikler yapılabilir ve gerektiğinde okuyucu veya yazıcı desteği

sağlanabilir. Uygun bir uyarlama bireysel özelliklere/ihtiyaçlara yanıt verir ama bunu yaparken testin ölçtüğü yapıyı ve puanların anlamını değiştirmez. Kapsayıcı bir uygulama için bu gereklidir. Test geliştiricileri, uyarlamanın ölçülen yapıyı değiştirmediğine yönelik kanıtlarını sunmalıdır. Uyarlamalar; bilişsel, dilsel, duyusal ya da fiziksel ihtiyaçlara yönelik olabilir ve çoğu zaman yasal zorunluluk taşır. Örneğin; dil yeterliliği olmayan öğrenciler için dilsel uyarlamalar, görme yetersizliği olan öğrenciler için metin büyütme, işaretleyici ve/veya okuyucu desteği sağlanması gibi düzenlemeler yapılabilir.

Standart 3.10. "Sınav uyarlamalarına izin verildiğinde test geliştiricileri/kullanıcıları, uyarlamaların nasıl kullanılacağına ilişkin standart düzenlemelerini belgelemekten ve uyarlamanın uygun şekilde gerçekleştirilip gerçekleştirilmediğin, izlemekten sorumludur" şeklinde ifade edilmektedir. Bu standart, uyarlamaların bir prosedüre bağlanmasını içermektedir. Test uyarlamaları yalnızca özel durumunu ve ihtiyacını belgeleyen bireylere sağlanmalı, sınav öncesinde bu ihtiyaç uzmanlarca onaylanmalı ve sürece dair bilgiler bireylere şeffaf biçimde sunulmalıdır. Uyarlamaların türü, uygulanma biçimi ve kuralları standart protokollerle belirlenmeli; test yöneticileri bu süreçte gerekli bilgi ve eğitime sahip olmalıdır. Uygulamalar standart şekilde yürütülmeli, tüm sapmalar kaydedilmeli ve özellikle büyük ölçekli sınavlarda uyarlamaların doğru ve adil biçimde uygulanıp uygulanmadığı takip edilmelidir. Böylece sınav süreci, bireysel farklılıkları gözetirken eşitlik ve kapsayıcılığı güvence altına alır. Örneğin ÖSYM görme yetersizliği olan adaylara engel durumlarını belgelemeleri ve talep etmeleri halinde okuyucu ve işaretleyici desteği sağlamakta ve okuyucu yardımı talep eden adaylarını kitapçığa bakmalarını ve soruları kendilerinin okumalarını yasaklamaktadır. Görme engelli adayların soru kitapçığını kendilerinin okuduğunun tespit edilmesi durumunda sınavlarının geçersiz sayılacağı belirtilmektedir.

Standart 3.11. "Bir testte, ölçülmek istenen yapıya erişimdeki engelleri kaldırmak için değişiklikler yapıldıysa test geliştiricileri/kullanıcıları yapılan değişiklik sonrası testin geçerliğini kanıtlayan verileri toplamak ve belgelemekle sorumludur. Geçerlik çalışmaları mümkünse yeterli büyülükte bir örneklem üzerinde yapılmalıdır" şeklinde ifade edilmektedir. Standart testlerde sınav süresinin uzatılması, dilin sadeleştirilmesi gibi değişiklikler yapıldığında pilot uygulamalar ya da saha çalışmalarıyla, gerçekleştirilen değişikliklerin geçerlik üzerindeki etkisi hedeflenen alt grupları (örneğin özel gereksinimli, dilsel/kültürel farklılıkları olan adaylar) temsil eden bireyler üzerinde test edilmelidir. Geçerlik çalışmaları hem bu değişikliğin ilgili grup için etkinliğine hem de değişiklik yapılan test ile orijinal test arasındaki puanlarının karşılaştırılabilirliğine odaklanmalıdır. Örneklem büyüklüğünün yeterli olmadığı durumlarda geçerlik kanıtlarına ulaşmak amacıyla daha küçük gruplarla nitel çalışmalar yürütülebilir ya da uzman görüşlerine başvurulabilir. Özetle kapsayıcılık adına yapılan uyarlamaların, testin psikometrik özellikleri üzerindeki etkisi araştırılmalı ve ampirik kanıtlarla destek sağlanmalıdır.

Standart 3.13. "Bir test, testin amacına en uygun olan dilde uygulanmalıdır" şeklinde ifade edilmektedir. Buna göre standart testlerde çift dilli (*bilingual*) ya da çok dilli öğrenciler göz önünde bulundurulmalı daha açık bir anlatımla test kullanıcıları adayların dilsel ve kültürel özellikleri ile dil yeterliliklerini dikkate almalıdır. Testin uygulanacağı en uygun dili belirlerken testin amacı ve bağlamı titizlikle değerlendirilmelidir. Testin amacı belirli bir dilde yeterliliği ölçmek değilse adaylar en iyi bildikleri dilde sınava alınmalıdır. Ancak bazen adayların en iyi bildiği dil, test edilen kazanımlarla ilgili eğitim aldıkları dil olmayabilir. Bu tür durumlarda test, adayın en iyi bildiği dilde değil de öğretim diliyle uygulanabilir. Adayın en iyi bildiği dilin seçilmesi, test puanlarının geçerliğinin sağlandığını garanti etmez; dil ile öğrenilen yapı arasındaki uyumsuzluklar puan yorumlarını etkileyebilir.

Kapsayıcılıkla ilgili bir diğer önemli konu da teknoloji destekli ölçme ve değerlendirmelerle ilgilidir. Uluslararası Test Komisyonu ve Test Yayıncıları Derneği (International Test Commission and Association of Test Publishers [ITC-TAP]) (2025) yayınladığı Teknoloji Tabanlı Değerlendirme Rehberi kılavuz kitabında teknoloji destekli ölçme ve değerlendirme süreçleri hakkında detaylı bilgiler vermektedir. Bu bilgilerden biri Evrensel Test Tasarımı (Universal Test Design-[ETT]) ile ilgilidir. Evrensel Tasarım, duyusal ve fiziksel engelleri olan bireyler de dâhil olmak üzere tüm bireylerin binalara ve fiziksel araçlara erişimini ve bunları kullanabilmesini sağlamak adına mimarlık alanında ortaya çıkmıştır (Story ve ark., 1998, akt.: ITC-TAP, 2025). Testlerin dijital uygulamalarıyla birlikte testlere ve testlerin uygulandığı arayüzlere erişilebilirlik üzerinde de çalışmalar başlatılmıştır. Erişilebilirlik sadece özel gereksinimli bireyler için test uyarlamaları olarak algılanmamalı, tüm test kullanıcılarını etkileyen ve ETT ile desteklenebilen bir konu olarak dikkate alınmalıdır (ITC-TAP, 2025). Erişilebilirlik; görsel, işitsel, fiziksel, konuşma, bilişsel, dil, öğrenme ve nörolojik problemler de dâhil olmak üzere oldukça çeşitli bir alanı kapsar (*Web Content Accessibility Guidelines,* 2025). Bu anlamda ETT özel gereksinimi olanlar, çok dilli öğrenenler, farklı yaş, cinsiyet ya da kültürel gruplardan gelenler gibi tüm bireylere fayda sağlar. ETT ölçülmek istenen yapıyla ilgisi olmayan engelleri, diğer bir ifadeyle ölçme hedefi dışındaki faktörlerin yol açtığı engelleri ortadan kaldırmayı hedefler. Elbette teknoloji desteği ile tasarlanan ölçme araçlarında da geleneksel yöntemlerde olduğu gibi en üst düzeyde dikkatli olmak gerekir. Teknoloji desteğiyle tasarlanmış bir testte yer alan maddeler özel gereksinimli bireylere bilişsel yük getirebilir (ITC-TAP, 2025). Bu noktada kullanıcı merkezli bir tasarım oluşturulmalı, diğer bir ifadeyle tasarım bireylere uygun hale getirilmelidir. Dijital tasarımlarda uygunluk, bireylerle fiziksel bir obje arasındaki karşılıklı ilişki olarak ifade edilir (Norman, 2013). Bilişsel yükü azaltan, testi alanlara kolaylık sağlayan, bireylere uygun olan ve bunları gerçekleştirirken psikometrik özelliklerden taviz vermeyen bir test, ETT'ye uygun kapsayıcı bir test olarak ifade edilebilir.

Teknoloji tabanlı değerlendirmelerde kapsayıcılık önemli bir avantaj olarak öne çıksa da bu değerlendirmelerde kullanılan ses ve video içerikleri, farklı özel gereksinimli bireylere uygun olmayabilir. İşitme kaybı olan bireyler için ses içerikli maddelerin mutlaka yazılı versiyonlarının olması ya da görme kaybı yaşayanlar için sesli açıklamalara yer verilmesi gerekir. Yapılabilecek bu tür düzenlemeleri planlarken özel gereksinimli bireyler için *Web*'i daha erişilebilir hale getirmeye yardımcı olacak stratejiler, standartlar ve destekleyici kaynaklar sunan *the World Wide Consortium-W3C*'nin *Web* İçeriği Erişilebilirlik Yönergeleri dikkate alınabilir (ITC-TAP, 2025).

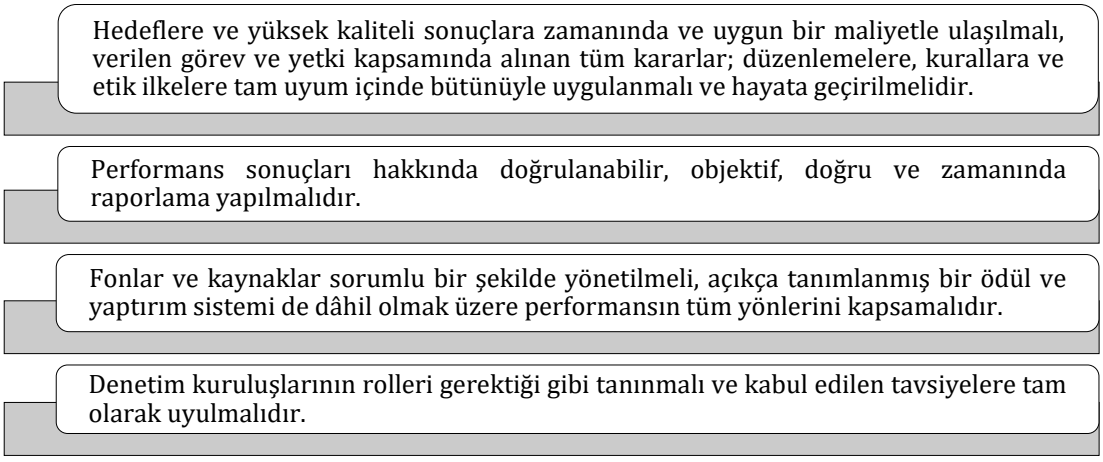
Ölçme ve değerlendirmede kapsayıcılık için yapılabilecekler özetlenecek olursa; eşit öğrenme ve eşit test fırsatları sağlanmalı, dilsel ve kültürel farklılıklara dikkat edilmeli, yanlılık içeren maddeler belirlenmeli ve bu maddelerden uzak durulmalı, özel gereksinimli bireyler için erişilebilir testler ve test düzenlemeleri yapılmalı, bilgisayar ve teknoloji tabanlı testlerin formatları ve kullanıcı arayüzleri evrensel tasarım ilkelerine uygun olarak düzenlenmelidir. Kapsayıcılığın sadece özel gereksinimli bireylerle ilgili bir konu olmadığı, ölçme ve değerlendirme süreçlerinde ölçme araçlarında çeşitlilik sağlanarak kişilerin kendilerini, yeteneklerini en iyi ifade edebilecekleri süreçler oluşturmak gibi çok daha geniş bir anlam taşıdığının farkında olunmalıdır.

3.4. Ölçmede Hesap Verebilirlik ve Şeffaflık

Hesap verebilirlik (*accountability*) bir kuruluşun/organizasyonun ve personelinin, alınan tüm kararlardan ve gerçekleştirilen tüm eylemlerden sorumlu olma ve taahhütlerini hiçbir kayıt ya da istisna olmaksızın yerine getirme yükümlülüğüdür (United Nations Development Programme [UNDP], 2025). APA (t.y.-d) hesap verebilirliği; bireylerin davranışları, kararları ya da yargıları hakkında başkalarına (örneğin yönetici, denetmen, danışman, resmi denetim organı, akran grubu) yanıt verebilme/hesap verebilme derecesi olarak tanımlamaktadır. UNDP (2025) hesap verebilirliğin içermesi gereken bileşenleri Şekil 1'deki gibi açıklamaktadır.

Şekil 1

UNDP'ye (2025) Göre Hesap Verebilirliğin Bileşenleri



UNESCO (2017) eğitimde hesap verebilirliği; öğretmenlerin, okulların, üniversitelerin, hükümetlerin ve diğer aktörlerin eğitimin sonuçları konusunda sorumluluk taşıması ve bunu şeffaflıkla raporlaması olarak ifade etmekte; ayrıca hesap verebilirliğin açık biçimde tanımlanmış sorumluluklar, sorumlulukların nasıl yerine getirildiğine dair hesap verme yükümlülüğü ve hesap verme yükümlülüğünün yasal, politik, sosyal ya da ahlaki gerekçesi olmak üzere üç temel unsuru olduğunu belirtmektedir. Eğitimde; düzenlemelere (bazen aynı anda uygulanan) uyum, mesleki normlara bağlılık ve sonuç odaklılık olmak üzere üç tür hesap verebilirlik türü olduğunu bildirmektedir (Anderson, 2005). Tüm bu tanımlar ve açıklamalar birlikte değerlendirildiğinde hesap verebilirliğin, sorumluluğa ve sonuçların şeffaf bir şekilde açıklanmasına odaklanan bir kavram olduğu ifade edilebilir.

Eğitimde testler ve test sonuçları hesap verebilirliğin bir ölçüsü olarak kullanılabilir. AERA ve arkadaşları (2014) eğitimde hesap verebilirlik amacıyla kullanılan testlere ilişkin belgelerin tüm potansiyel kullanıcıların/ paydaşların (okul personeli, veliler, öğrenciler) erişimine açık olacak biçimde, erişilebilir bir formatta ve anlaşılır bir dilde yayımlanması gerektiğini belirtmektedir. Hesap verebilirlik, eğitimde ölçme ve değerlendirme sürecinde paydaşların sonuçların doğruluğu, güvenirliği ve adilliği hakkında bilgi sahibi olmalarını sağlar. AERA (2014) hesap verebilirlikle ilgili bazı standartlar tanımlamıştır.

Standart 13.3. "Hesap verebilirlik indeksleri, program değerlendirmelerinde ve politika çalışmalarında etkililik göstergeleri ya da katma-değer modelleri gibi diğer istatistiksel modeller olarak kullanıldığında söz konusu indekslerin, göstergelerin veya modellerin hangi yöntemle oluşturulduğu açıklanmalı, gerekçelendirilmeli ve teknik özellikleri rapor edilmelidir." şeklinde ifade edilmektedir. Katma değerli değerlendirme yaklaşımı,

öğrenci ve okullara özgü farklılıkları kontrol altına alarak her bir öğrencinin sınav sonuçlarını inceleyerek öğrencinin ne kadar gelişim gösterdiğini belirleyen ve bu gelişimi okullarla ya da öğretmenlerle ilişkilendirmede objektif bir değerlendirme sağlayan sistem olarak ifade edilebilir (Şen ve ark., 2020). Test puanlarının birleştirilmesiyle oluşturulan indeksler, dayanak noktaları olan test puanlarında olduğu gibi geçerlik, güvenirlik ve adalet bakımından incelenip şeffaf biçimde raporlanmalıdır. Bu tür indekslerin oluşturulmasına ilişkin yöntem ve kurallar, teknik özelliklerinin dokümantasyonu ile birlikte kullanıcılara sunulmalıdır. Puanları birleştirmeye yönelik çeşitli yaklaşımların güçlü ve zayıf yönleri değerlendirilmeli; indekslerin, göstergelerin ya da modellerin bilgileri bağımsız araştırmacıların süreci tekrar edebilmesine imkân verecek şekilde açıkça paylaşılmalıdır.

Standart 13.4. "Program değerlendirmede, politika çalışmasında ya da hesap verebilirlik sisteminde kullanılan bir teste ilişkin geçerlik, güvenilirlik ve adalet kanıtları toplanmalı, bu kanıtlar ulaşılabilir olmalıdır" şeklinde ifade edilmektedir. Bir testin program değerlendirmede, politika çalışmalarında ya da hesap verebilirlik sistemlerinde kullanıma ve hedef örnekleme uygunluğu kanıtlarla ortaya konulmalıdır. Test sonuçlarının yayımlanmasından ya da raporlanmasından sorumlu kişiler, verilerin yanlış yorumlanmasını ve kötüye kullanımını önleyecek ek bilgileri açıklamalıdır. Özellikle bir değerlendirme ya da hesap verebilirlik sistemi; bir program, kurum ya da hizmet sağlayıcısının etkililiğine ilişkin yorumları desteklemek için tasarlanmışsa bu yorumların amaçlanan kullanımlar açısından geçerliği araştırılmalı ve belgelenmelidir. Bir test birden fazla amaç için kullanılıyorsa (örneğin; öğrencilerin bireysel başarısının izlenmesi, bireyler ya da öğrenci grupları için öğretim planlamasına yardımcı olacak bilgi sağlanması, bölgelerin, okulların veya öğretmenlerin değerlendirilmesi), bu kullanımların her biri için yorumların geçerliğine dayalı kanıtlar toplanmalı ve kullanıcılara sunulmalıdır. Personelin performansını değerlendirmek için kullanılan bir test söz konusu ise bu testin farklı personel grupları (örneğin öğretmenler, özel eğitim öğretmenleri, müdürler) için uygunluğu incelenmelidir.

Standart 13.5. "Testlerin değerlendirme ya da hesap verebilirlik amaçlarıyla geliştirilmesinden ve kullanılmasından sorumlu olan kişiler, test sonuçlarının uygulanacağı tüm gruplar için doğru yorumları ve uygun kullanım biçimlerini teşvik edecek adımlar atmalıdır" şeklinde ifade edilmektedir. Ölçme işleminden sorumlu kişiler, program ya da politika değerlendirmesinde yer alan farklı alt gruplardaki test katılımcıları için erişimi teşvik edecek ve yorumların geçerliğini en üst düzeye çıkaracak (örneğin uygun düzenlemeler sağlayarak) bir test süreci tasarlamalıdır. Diğer bir ifadeyle, test kullanıcıları, hesap verebilirlik amaçları doğrultusunda ilgili tüm alt grupları test programına dâhil etmek için her türlü çabayı göstermelidir. Ayrıca varsa dışlama kuralları, test düzenlemeleri (*modifications*) veya test/uygulama koşullarındaki diğer değişikliklerle ilgili belgeleri sağlamalı ve alt gruplar için test puanlarına dayalı olarak yapılan yorumlarının geçerliğine ilişkin kanıtlar sunmalıdır. Test puanlarının özetleri alt gruplara göre (örneğin ırk/etnik grup) ayrı ayrı raporlandığında test

kullanıcıları bu gruplar için puanların güvenilirliğini/kesinliğini ve puanlardan hareketle yapılan yorumların geçerliğini değerlendirmek üzere analizler yapmalı, sonuçların özetlerini yayınlarken bu bilgileri paylaşmalıdır. Hesap verebilirlik ya da uygulanan programın etkililiğini ölçmek için kullanılan karmaşık indekslerin analizleri, belirli alt gruplara ya da bu alt gruplara hizmet veren programlara/kurumlara karşı yanlılık olasılığını ele almalıdır. Örneğin indeksteki puanların, sınava giren kişilerin ırk/etnik köken gibi özellikleriyle ilgili bir yanlılık içerdiği belirlenirse bu puanlar yanlılık giderilmeden kullanılmamalıdır.

Standart 13.6. "Test performansındaki grup farklılıklarına ilişkin raporlar, farklılıkların anlamlı bir şekilde yorumlanabilmesi için mümkünse ilgili bağlamsal bilgilerle birlikte sunulmalıdır. Uygun bağlamsal bilgi mevcut değilse kullanıcılar yanlış yorumlamaya karşı uyarılmalıdır" şeklinde ifade edilmektedir. Grupların (örneğin cinsiyet, ırk/etnik köken, özel gereksinimli olma, dil yeterliliği, sosyoekonomik durum ya da coğrafi bölge) ortalama test puanları arasında gözlemlenen farklar; öğrenme fırsatı, eğitim deneyimi, çaba, eğitmen kalitesi, ebeveyn desteğinin düzeyi ve türü gibi faktörlerdeki değişkenliklerden kaynaklanabilmektedir. Eğitimde zaman içinde grup performansında ortaya çıkan farklılıklar, test edilen kişilerin yer aldığı örneklemdeki değişikliklerden (örneklem büyüklüğündeki değişiklikler dâhil) ya da bu bireylerin deneyimlerindeki değişikliklerden etkilenebilmektedir. Kullanıcılara, bu grup farklılıklarını yorumlarken ve bu farklılıklara dönük politikalar/uygulamalar tasarlarken uygun bağlamsal bilgileri dikkate almaları tavsiye edilmelidir. Ayrıca değerlendirmeler ulusal sınırlar içinde test puanlarının karşılaştırılmasını içeriyorsa puanların karşılaştırılabilirliğine ilişkin kanıtlar sunulmalıdır.

Standart 13.7. "Bir test değerlendirme ya da hesap verebilirlik uygulamalarında kullanılmak üzere seçildiğinde testin kullanım amacı ile beklenen sonuçları açıkça belirtilmeli ve uygunsuz kullanımlara karşı uyarılarda bulunulmalıdır." şeklinde ifade edilmektedir. Bazı durumlarda testin amacı, belli bir müfredat programını değerlendirmek gibi daha sınırlı olabilir ya da test değerlendirme hakkında bilgilendirme yapmak dışında belli sonuçları desteklemek için tasarlanmamış olabilir. Özellikle test temelli hesap verebilirlik sistemlerinde, karar vericilere işlevsel bilgiler sağladığı ve eğitimciler ile öğrencilerin daha iyi performans göstermelerini destekleyecek teşvikler oluşturarak eğitimin kalitesini artırdığı gerekçeleriyle testlerin kullanımı desteklenmektedir. Eğer böyle bir bakış açısı benimseniyor ya da testlerin kullanımı zorunlu hale getiriliyorsa mutlaka bu iddiaları destekleyen geçerlik kanıtları sunulmalıdır. Belirli bir geçerlik iddiasına ilişkin kanıtların toplanması ve raporlanması, program tasarımına dâhil edilmesi gereken işlerdir. Öğrencilerin başarısını artırmak gibi test kullanımının faydalarına ilişkin belirli bir iddia, ampirik verinin yanı sıra teorik ya da mantıksal argümanlarla da desteklenebilir. Belirtilen iddiayla çelişkili olan bilimsel literatürdeki bulgulara da dikkat edilmelidir.

Standart 13.8. "Testlerin politika, değerlendirme ve hesap verebilirlik bağlamlarında kullanılmasını zorunlu kılanlar ve testleri bu bağlamlara uygun olarak kullananlar, testlerin etkilerini izlemeli ve olumsuz sonuçları belirleyip en aza indirmelidir." şeklinde ifade edilmektedir. Testlerin politika, değerlendirme ve hesap verebilirlik amaçlarıyla kullanımı bazı durumlarda beklenmedik sonuçlara yol açabilir. Özellikle testin taşıdığı riskler yüksek olduğunda testleri zorunlu kılanlar ile test sonuçlarını kullananlar olası beklenmedik sonuçların belirlenmesine yönelik adımlar atmalıdır. Beklenmeyen olumsuz sonuçlar arasında; test maddelerinin önceden öğretilmesi, test uygulama prosedürlerinin değiştirilmesi, belirli katılımcıların teste girmesinin engellenmesi ya da test dışında tutulması yer alabilmektedir. Bu uygulamalar, testle ölçülmek istenen yapının gerçek performansı yansıtmamasına yol açabilmekte ve sahte yüksek puanlara ulaşılmasına neden olabilmektedir. Ayrıca bu uygulamaların kanunen yasaklanması söz konusu olabilir. Test prosedürleri bu tür sonuçların meydana gelme olasılığını en aza indirecek şekilde tasarlanmalı, kullanıcılara doğru olmayan test hazırlık uygulamalarından kaçınmaları için rehberlik ve teşvik sağlanmalıdır.

Standart 13.9. "Değerlendirme ya da hesap verebilirlik uygulamalarında, ek bilgi kullanımı genel yorumun geçerliğine katkıda bulunuyorsa test sonuçları diğer kaynaklardan gelen bilgilerle birlikte kullanılmalıdır" şeklinde ifade edilmektedir. Test dışındaki performans göstergeleri çoğu durumda gerekli ve neredeyse her zaman fayda sağlayıcıdır. İncelenen programın ya da politikanın kapsamlı bir resmini sunmak ve test sonuçlarının yorumlanmasına yardımcı olmak için genellikle hizmet alanların/yararlanıcıların özellikleri ile seçim ölçütleri, sunulan hizmetler, ortam ve kaynaklar gibi değişkenleri betimlemek ya da analiz etmek gerekir. Hesap verebilirlik bağlamında, bir öğretmen veya sağlık hizmeti sağlayıcısı gibi bireyler ya da kuruluşlar üzerinde büyük etkiye sahip olacak bir karar alırken test puanlarına yanında ilgili diğer bilgiler de göz önünde bulundurulmalıdır. Değerlendirmelere ya da hesap verebilirlik sistemlerine dâhil edilebilecek diğer bilgilere; eğitimcilerin ya da sağlık hizmeti sağlayıcılarının uygulamalarına ait ölçümler (örneğin sınıf gözlemleri, kontrol listeleri) ile öğrenci başarısının test dışı ölçümleri (ders alma, üniversiteye devam) örnek olarak verilebilir. Katma değerli modelleme durumunda, bazı araştırmacılar öğrencilerin demografik özelliklerini (örneğin ırk/etnik köken ve sosyoekonomik durum) kontrol değişkeni olarak ele almak gerektiğini savunurken diğer bazı araştırmacılar bu tür değişkenlerin ölçümlerin performansını iyileştirmediğini ve istenmeyen sonuçlara (bazı öğrenciler için diğerlerinden daha düşük standartlar belirlenmesi gibi) yol açabileceğini öne sürmektedir. Bu tür modellere hangi değişkenleri dahil etmek gerektiğine ilişkin dair kararlar, söz konusu değişkenleri dahil etmenin ya da hariç tutmanın ne tür etkilerinin olabileceğine ilişkin ampirik kanıtlara dayanmalıdır. Politika düzenlemelerinde test sonuçlarının yorumlanmasıyla ilgili ek bir bilgi türü, sınava girenlerin/test uygulanan kişilerin motivasyon derecesidir. Özellikle bireysel puanların test katılımcılarına bildirilmediği ya da puanların test katılımcıları için sonuçlarla ilişkilendirilmediği

durumlarda katılımcıların sınav deneyimini ciddiye alıp almadıklarını belirlemek önemli hale gelmektedir. Şüpheli motivasyona sahip bireylerin puanlarının dahil edilip edilmeyeceğine ilişkin karar kriterleri açıkça belgelenmelidir.

Tüm bu standartlar özetlendiğinde eğitimde hesap verebilirliğin; süreçlerin objektif bir şekilde yürütülmesi, sonuçların ve raporların herkes için anlaşılır ve şeffaf bir şekilde oluşturulması, geçerlik, güvenirlik ve adalet kanıtlarının sunulması ve bu süreçlerin farklı zamanlarda bağımsız araştırmacılar tarafından tekrar edilebilirliğini sağlayacak bilgilerin paylaşımı ile gerçekleştirilebileceği anlaşılmaktadır.

Hesap verebilirliğin bir parçası olarak şeffaflık (*transparency*) ölçme ve değerlendirme süreçlerinin tüm paydaşlar için açık, anlaşılır, erişilebilir olmasını gerektirir. Şeffaf(lık), açıkça görünen ve belirgin olan anlamına gelmektedir (TDK, n.d.-b). Ölçme ve değerlendirme süreçlerinde; testlerin geliştirilmesi/uyarlanması, uygulanması, puanlanması ve puanların yorumlaması ve tüm bunların raporlaştırılması aşamalarında şeffaf olunması gerekmektedir. Şeffaflık, hesap verebilirlik için önemli bir unsurdur. Öğrencilerin kullanılacak ölçme aracı ile hangi performans düzeyine hangi puanın karşılık geleceği hakkında bilgilendirilmesi, öğretmenin puanlama yaptığı aracı (dereceli puanlama anahtarı, cevap kâğıdı gibi) öğrencilerle paylaşması ve ölçümlerle ilgili olarak zamanında ve anlamlı geri bildirimlerde bulunması eğitimde ölçme-değerlendirme uygulamalarında şeffaflığın sağlanmasına katkı getirecek örneklerdir. Şeffaflığın sağlandığı sistemlerde hesap verebilirlik süreçleri hızlı ve verimli bir şekilde gerçekleşmektedir.

4. ÖLÇME VE DEĞERLENDİRMEDE YASAL YÜKÜMLÜLÜKLER

Yasal yükümlülük; bireylerin veya kurumların kanunlar ve ilgili mevzuat çerçevesinde yerine getirmek zorunda oldukları görevler ve sorumluluklar anlamına gelir (Dworkin, 2007). Ölçme ve değerlendirme süreçlerinde yasal yükümlülük; ölçmeyi yapan öğretmenler, eğitim kurumları ve kuruluşları ile sınava giren adayların uymaları gereken yükümlülükleri ifade eder. Hukuki bağlayıcılığı olan bu yükümlülükler hesap verebilirlik ve şeffaflık sağlar.

Millî Eğitim Bakanlığı (MEB) ilk kez 2023 yılında Ölçme ve Değerlendirme Yönetmeliği'ni (URL-4) yayınlamıştır. Bu yönetmeliğin amacı Madde 1'de "*Millî Eğitim Bakanlığı tarafından yapılan merkezî sistem sınavları, ulusal/uluslararası izleme araştırmaları, okul öncesi eğitim kurumları ve ilkokullarda akademik ve sosyal gelişimin takibi, ortaokul ve ortaöğretim kurumlarındaki ortak yazılı sınav faaliyetlerinin usul ve esasları ile ölçme değerlendirme merkezi müdürlüklerinin görev, yetki ve sorumluluklarını düzenlemektir.*" şeklinde açıklanmaktadır. İlgili yönetmelik "*Başlangıç Hükümleri; Ölçme ve Değerlendirme İlke ve Esasları; Sınav Uygulama Esasları ile Ulusal ve Uluslararası İzleme Araştırmaları; Merkezî Sistem Sınavlarına İlişkin Usul ve Esaslar; Ölçme Değerlendirme Merkezi Müdürlüklerinin (ÖDM) Kuruluşu, Açılması, Kapatılması ile Görev, Yetki ve*

Sorumlulukları ve *Çeşitli ve Son Hükümler*" olmak üzere altı bölümden oluşmaktadır. Yönetmeliğin içeriği özetlenecek olursa ulusal ve sınıf içi ölçmeler başta olmak üzere uluslararası ölçme ve değerlendirme süreçleri hakkında bilgiler verildiği, Türkçe dil becerisinin dört boyutta ölçüleceği, öğrencilerin akademik ve sosyal gelişimlerinin takibinin yapılacağı, ölçme ve değerlendirme süreçlerinin güçlendirilmesi, ÖDM'lerin etkililiklerinin artırılması, ülke çapındaki izleme çalışmalarıyla (ABİDE, okul tabanlı izleme) öğrenciler hakkında ulusal düzeyde geri bildirim sağlanması konuları üzerinde durulduğu görülmektedir (Şengül Avşar, 2023).

MEB Ortaöğretim Kurumları Yönetmeliği'nde MEB'e bağlı resmî ve özel örgün ortaöğretim kurumlarında eğitim, öğretim, yönetim ve işleyişe ilişkin usul ve esasları düzenlemiştir (URL-5). Bu yönetmeliğin dördüncü kısmı "Öğrenci Başarısının Değerlendirilmesi" olup "*Ölçme ve Değerlendirme*; *Sınavlar*; *Sınıf Geçme*; *Telafi Programı, Yoğunlaştırılmış Eğitim, Tamamlayıcı Eğitim ve Önceki Öğrenmelerin Tanınması*; *Okul Birincilerinin Tespiti"* şeklinde beş bölümden oluşmaktadır. Bu bölümler incelendiğinde kazanımlara ve ölçülen özelliğe uygun ölçme, üst düzey zihinsel becerilerin ölçülmesi, psikometrik özellikleri belirlenen nitelikli ölçme araçlarının kullanımı, sınavların, performans çalışmalarının, proje ve uygulamaların nasıl puanlanacağı, sınavların uygulanış biçimi (sınav sayısı, sınav süresi), mazeret sınavları, sınav sonuçlarının nasıl duyurulacağı, öğrencilerin sınav sonuçlarına itirazı, sınava katılmayan öğrencilerin değerlendirilmesi, sınıf geçme ölçütü, mezuniyet puanı gibi konularda yapılacaklar hakkında bilgiler verildiği görülmektedir (Şengül Avşar, 2023).

ÖSYM'ye ilişkin; "*6114 sayılı Ölçme, Seçme ve Yerleştirme Merkezi Hizmetleri Hakkında Kanun, Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı Personeli Görevde Yükselme ve Unvan Değişikliği Yönetmeliği; Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı Disiplin Amirleri Yönetmeliği; ÖSYM Uzman ve Uzman Yardımcılığı Sınav, Atama, Yetiştirilme, Görev ve Çalışma Usul ve Esasları Hakkında Yönetmelik; Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığında Çalıştırılacak Yerli ve Yabancı Sözleşmeli Personel Hakkında Yönetmelik; Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığının Sınav Hizmetlerine Yönelik Mal ve Hizmet Alımları Yönetmeliği; Adayların ve Sınav Görevlilerinin Sınav Binalarına Giriş Koşullarına İlişkin Yönetmelik; Soruların Hazırlanması, Soru Havuzunun Oluşturulması ve Soruların Güvenliğinin Sağlanmasına İlişkin Yönetmelik; Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı Sınav Koordinatörlükleri ve Başvuru Merkezlerinin Çalışma Usul ve Esasları Hakkında Yönetmelik*" olmak üzere çeşitli kanun ve yönetmeliklere ulaşılabilmektedir (ÖSYM, 2025). Bu kanun ve yönetmelikler genel olarak kurum içi düzenlemeler, aday ve sınav görevlileri ile sınav başvuru merkezleri üzerinde durmaktadır. İlgili kanun ve yönetmelikler detaylı incelendiğinde hem kurumun hem de paydaşların yerine getirmek zorunda olduğu görev ve sorumluluklara odaklandığı görülür. Örneğin *Adayların ve Sınav Görevlilerinin Sınav Binalarına Giriş Koşullarına İlişkin Yönetmelik'*te adaylar ile sınav görevlilerinin; sınav binasına ve

salonlarına ne zaman alınacağı, yanlarında neleri getirip getiremeyecekleri, kimliklerini beyan ederken kullanabilecekleri kimlik kartları, adaylar için sınav giriş ve görevliler için görevlendirme belgesi taşıma zorunluluğu, sınav görevlilerinin adayın kimliğinden şüphelendiği durumlarda neler yapması gerektiği gibi hususlar detayları ile açıklanmıştır.

Yasal yükümlükler, bireyler üzerinde önemli etkileri olan yüksek riskli testler için ayrı bir önem taşımaktadır. Bu testler sonuçlarının bireyler, okullar ya da kurumlar üzerinde önemli ve ciddi sonuçlar doğurduğu ölçme araçlarıdır. Yüksek riskli testler; bir kişinin gelecekteki akademik kariyerini ve iş fırsatlarını belirleyen, öğrencilerin bir okulu bitirmek, bir eğitim programına girmek, bir üniversiteye devam etmek, burs alabilmek ya da bir başvuru için yeterlilik elde etmek için geçmeleri gereken testleri ifade etmektedir (Cizek, 2001; Jones ve ark., 2003). Bu testler sadece bireyleri değil öğretmenleri, aileleri, sosyal çevreyi de etkilemektedir (Johnson, 2017). Yüksek riskli testler; *öğrenciler* için lise mezuniyeti ya da sınıf yükseltme, *eğitimciler* için iş süreçleri/terfiler, maaşlar/promosyonlar vb. kararlar alma, *okullar ve bölgeler* için bütçe belirleme ve okulları yeniden yapılandırma gibi çalışmalara yön verebilmektedir (Blazer, 2011). ÖSYM tarafından düzenlenen üniversiteye giriş sınavları, kamu personeli seçme sınavları bu testlere örnek olarak gösterilebilir. Böylesine önemli testlerin geliştirilmesinden uygulanmasına, uygulanmasından puanlanmasına, puanlanmasından sonuçların duyurulmasına kadar tüm süreçlerde titiz çalışmak gerekmektedir. Her aşamada tüm paydaşları bağlayıcı yasal yükümlülüklerin olması, süreçleri güçlendirerek şeffaf ve hesap verebilir hale getirmektedir. Özetle tüm paydaşları ilgilendiren ve önemli bir sorumluluk alanı olan yasal yükümlülükler, hesap verebilirliği ve şeffaflığı sağlayarak ölçme ve değerlendirme süreçlerini sağlamlaştırmaktadır.

5. SONUÇ

Ölçme ve değerlendirme süreçleri çoklu paydaşlarıyla birlikte ciddi bir sorumluluk gerektirmektedir. Testleri geliştiren ve değerlendirme hizmeti verenler; öncelikle teknik anlamda psikometrik özellikleri üzerinde çalışılmış nitelikli ölçme araçlarını geliştirmeli/uyarlamalı, kapsayıcılığa dikkat etmeli, gerekli gizliliği koruyarak süreçleri şeffaf bir şekilde yürütmelidir. Eğitimde kullanılacak ölçme ve değerlendirme ürün ve hizmetlerini seçen veya bu seçime rehberlik edenlerin de sahip olması gereken sorumluluklar vardır. Bu kişiler seçtikleri ölçme araçlarının ölçme amacına uygunluğundan emin olmalıdırlar. Bu da bu kişilerin alan uzmanı ve ölçme ve değerlendirme okuryazarlığına sahip kişiler olmasını gerektirir. Ölçme ve değerlendirme süreçlerini yönetenler de sınavların adil ve uygun bir şekilde uygulanmasını sağlamalıdırlar. Bunu yaparken yüksek standartlara sahip olunmalı, kapsayıcılığa dikkat edilmeli ve yanlılığa neden olabilecek unsurlar ortadan kaldırılmalıdır. Ölçme sonuçlarının uygun puanlama yöntemleriyle tüm paydaşlara zamanında ve doğru bir şekilde bildirilmesi esastır. Gerekli durumlarda sonuçlara itiraz hakkı verilmesi, bu

itirazlara bağlı olarak soru iptali veya cevap anahtarı değişikliği gibi süreçler objektif bir şekilde yürütülmelidir. Ölçme ve değerlendirme süreç içinde değişen, gelişen ya da tekrarlı eğitimlerle bilgilerin güncel tutulması gereken bir alandır. Bunun için ölçme ve değerlendirme alanında eğitim veren kişilerin alanında uzman profesyonel kişiler olması, farklı alan uzmanlarıyla uyumlu bir şekilde çalışması gerekir. Ayrıca ölçme ve değerlendirme sonuçları sistematik bir şekilde arşivlenip üzerinde gerekli araştırmalar yapılarak eğitim programlarının ve değerlendirmelerinin niteliği artırılabilir.

Adalet ölçme ve değerlendirmede üzerinde durulması gereken önemli bir kavramdır. En temel haliyle ölçmede adaletin sağlanabilmesi için testlerin tüm gruplara eşit fırsat sunması gerekir. Ölçme ve değerlendirmede temel standart geçerli ve güvenilir sonuçlara ulaşmaktır. Farklı uygulamalarda geçerli ve güvenilir sonuçlara ulaşmanın yolu ölçme araçlarının ölçmeyi hedeflediği özelliği, grubun özelliklerinden bağımsız ölçmesinden geçer. Bu da ölçmede yansızlığı gerektirir. Adil ölçme yapmanın bir diğer koşulu da yanlı ölçme yapmamaktır.

Ölçme ve değerlendirme süreçlerinde dikkat edilmesi gereken bir diğer konu kapsayıcılıktır. Kapsayıcılığın sağlanabilmesi için ölçme ve değerlendirme sürecinin planlama aşamasından başlanarak uygulama ve puanların duyurulmasına kadar tüm aşamalarda kapsayıcılık ilkeleri gözetilmelidir. Eğitim ortamlarında çeşitli değerlendirme yöntemleri, seçenekleri ve konuları sunma, açık, adil ve şeffaf değerlendirme ölçütlerini sağlama gibi çabalar kapsayıcılığı güçlendirir. Ayrıca kapsayıcılığın sadece özel gereksinimli bireylerle ilgili bir konu olmadığı, ölçme araçlarında çeşitlilik sağlanarak kişilerin kendilerini, yeteneklerini en iyi ifade edebilecekleri süreçler oluşturmak gibi çok daha geniş bir anlam taşıdığının farkında olunmalıdır.

Ölçme ve değerlendirmede etik sorumluluklar, adalet ve kapsayıcılık beraberinde bir diğer önemli kavram olan hesap verebilirliği gündeme getirir. Eğitimde testler ve test sonuçları hesap verebilirliğin bir ölçüsü olarak kullanılabilirken bu testlere ilişkin belgelerin de tüm paydaşlar için anlaşılabilir bir formatta erişime açık olması gerekir. Hesap verebilirlikle birlikte paydaşlar; sonuçların doğruluğu, güvenirliği ve adilliği hakkında bilgi sahibi olur. Son olarak eğitimciler, eğitim kurumları ve kuruluşları ile sınava giren adayların uymaları gereken yükümlülükleri ifade eden yasal yükümlülükler, hukuki bağlayıcılığıyla hesap verebilirlik ve şeffaflığı sağlar.

KAYNAKÇA | REFERENCES

Advance HE. (2019, 24 Ocak). *Inclusive assessment: Where next?* https://www.advance-he.ac.uk/news-and-views/inclusive-assessment-where-next

American Educational Research Association. (2011). *Code of ethics*. American Educational Research Association.

American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

American Psychological Association. (2004). *Code of fair testing practices in education*. Joint Committee on Testing Practices. https://www.apa.org/science/programs/testing/fair-testing.pdf

American Psychological Association. (t.y.-a). Ethics. *APA dictionary of psychology* içinde. https://dictionary.apa.org/ethics adresinden 7 Ekim 2025 tarihinde alınmıştır.

American Psychological Association. (t.y.-b). Fairness. *APA dictionary of psychology* içinde. https://dictionary.apa.org/fairness adresinden 7 Ekim 2025 tarihinde alınmıştır.

American Psychological Association. (t.y.-c). Bias. *APA dictionary of psychology* içinde. https://dictionary.apa.org/bias adresinden 7 Ekim 2025 tarihinde alınmıştır.

American Psychological Association. (t.y.-d). Accountability. *APA dictionary of psychology i*çinde. https://dictionary.apa.org/accountability adresinden 7 Ekim 2025 tarihinde alınmıştır.

Anderson, J. A. (2005). *Accountability in education*. International Institute for Educational Planning.

Arıkan, Ç. A., Uğurlu, S. & Atar, B. (2016). Mimic, sibtest, lojistik regresyon ve Mantel-Haenszel yöntemleriyle gerçekleştirilen DMF ve yanlılık çalışması. *Hacettepe University Journal of Education, 31*(1), 34–52. https://doi.org/10.16986/HUJE.2015014226

Ayan, C. (2024). Değişen madde fonksiyonu analizleri. A. Şengül Avşar (Ed.), *Jamovi ile kolay istatistik* (3. baskı, s. 277–290) içinde. Nobel Akademik Yayıncılık.

Blazer, C. (2011). *Unintended consequences of high-stakes testing* (Information Capsule No. 1008). Research Services, Miami-Dade County Public Schools.

Bowen, N. K. & Masa, R. D. (2015). Conducting measurement invariance tests with ordinal data: A guide for social work researchers. *Journal of the Society for Social Work and Research, 6*(2), 229–249. https://doi.org/10.1086/681607

Brown, G. T. L., Harris, L. R., O'Quin, C. & Lane, K. E. (2017). Using multi-group confirmatory factor analysis to evaluate cross-cultural research: Identifying and understanding non-invariance. *International Journal of Research & Method in Education, 40*(1), 66–90. https://doi.org/10.1080/1743727X.2015.1070823

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3. baskı). Routledge.

Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice, 20*(4), 19–27. https://doi.org/10.1111/j.1745-3992.2001.tb00072.x

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.

Dorans, N. J. & Holland, P. W. (1992). DIF detection and description: Mantel–Haenszel and standardization. *ETS Research Report Series, 1992*(1), i–40. https://doi.org/10.1002/j.2333-8504.1992.tb01440.x

Dworkin, R. (2007). The model of rules. L. B. Nielsen (Ed.), *Theoretical and empirical studies of rights* (s. 14–46) içinde. Routledge.

Ercikan, K., Gierl, M. J., McCreith, T., Puhan, G. & Koh, K. (2004). Comparability of bilingual versions of assessments. *Applied Measurement in Education, 17*(3), 301–321. https://doi.org/10.1207/s15324818ame1703_4

International Test Commission & Association of Test Publishers. (2025). *Technology-based assessment: Guidance for test developers and users*. https://www.intestcom.org/upload/media-library/tba-guidelines-ver-11-july-2025-1754044782Z3vV9.pdf

Johnson, P. B. (2017). High-stakes testing and NCLB. *What is authentic educational reform?* (s. 55–69) içinde. Routledge.

Jones, G. M., Jones, B. D. & Hargrove, T. Y. (2003). *The unintended consequences of high-stakes testing*. Bloomsbury.

Karabük Üniversitesi Medya. (2025, 21 Haziran). *ÖSYM'den KBÜ'lü akademisyenlere soru hazırlama ve test geliştirme eğitimi.* https://kbumedya.karabuk.edu.tr/osymden-kbulu-akademisyenlere-soru-hazirlama-ve-test-gelistirme-egitimi

Messick, S. (1998). Consequences of test interpretation and use. *Educational Measurement: Issues and Practice, 17*(2), 35–39. https://doi.org/10.1002/j.2333-8504.1998.tb01797.x

National Council on Measurement in Education. (2023). *Code of professional responsibilities in educational measurement* (2. baskı). https://ncme.org/wp-content/uploads/2025/10/NCME_STUC_2023_CodeofProfResp_Final_Updated.pdf

Norman, D. A. (2013). *The design of everyday things*. MIT Press.

Oxford University Centre for Teaching and Learning. (2025). *IncludED: Designing inclusive assessment*. https://www.ctl.ox.ac.uk/included-designing-inclusive-assessment

Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı. (2011). *Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığının sınav hizmetlerine yönelik mal ve hizmet alımları yönetmeliği*. https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=15562&MevzuatTur=7&MevzuatTertip=5

Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı. (2025). *Kanun ve yönetmelikler*. https://www.osym.gov.tr/TR,12849/kanun-ve-yonetmelikler.html

Stark, S., Chernyshenko, O. S. & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology, 91*(6), 1292–1306. https://doi.org/10.1037/0021-9010.91.6.1292

Şen, S., Yıldırım, İ. & Karacabey, M. F. (2020). The implementation of value-added assessment. *Education and Science, 45*(202), 79–91. https://doi.org/10.15390/EB.2020.8023

Şengül Avşar, A. & Emons, W. H. M. (2021). A cross-cultural comparison of non-cognitive outputs. *Studies in Educational Evaluation, 70*, 101053. https://doi.org/10.1016/j.stueduc.2021.101053

Şengül Avşar, A. (2023). Ortaöğretim ve yükseköğretimde ölçme ve değerlendirme. H. Tabak & V. Özdemir (Ed.), *Sınıf içi öğrenmeleri ölçme ve değerlendirme* (s. 379–417) içinde. Pegem Akademi.

Türk Dil Kurumu. (n.d.-a). Adalet. *Güncel Türkçe Sözlük* içinde. https://sozluk.gov.tr/?q=adalet adresinden 7 Ekim 2025 tarihinde alınmıştır.

Türk Dil Kurumu. (n.d.-b). Şeffaflık. İçinde *Güncel Türkçe Sözlük* içinde. https://sozluk.gov.tr/?q=seffaflik adresinden 7 Ekim 2025 tarihinde alınmıştır.

Thurlow, M. L., Lazarus, S. S., Christensen, L. L. & Shyyan, V. (2016). *Principles and characteristics of inclusive assessment systems* (NCEO Report 400). https://nceo.umn.edu/docs/OnlinePubs/Report400/NCEOReport400.pdf

Tse, W. W. Y., Lai, M. H. C. & Zhang, Y. (2024). Does strict invariance matter? *Behavior Research Methods, 56*(4), 3117–3139. https://doi.org/10.3758/s13428-023-02247-6

Turgut, M. F. & Baykul, Y. (2013). *Eğitimde ölçme ve değerlendirme* (5. baskı). Pegem Akademi.

United Nations Development Programme. (2025). *Accountability*. https://www.undp.org/sites/g/files/zskgke326/files/2025-04/accountability-system-policy-popp-final-1.pdf

UNESCO. (2008). *Education for all by 2015: Will we make it?* https://unesdoc.unesco.org/ark:/48223/pf0000154743

UNESCO. (2015). *Education 2030: Incheon declaration and framework for action*. https://unesdoc.unesco.org/ark:/48223/pf0000245656

UNESCO. (2017). *Accountability in education: Meeting our commitments*. https://unesdoc.unesco.org/ark:/48223/pf0000259338

University of Bristol. (2025). *Principles for inclusive assessment*. https://www.bristol.ac.uk/bilt/sharing-practice/guides/inclusive-teaching-and-assessment/principles-for-inclusive-assessment/

University of Guelph. (t.y.). *Assessing student learning in in-person courses*. https://otl.uoguelph.ca/teaching-assessment-resources/assessing-student-learning-person-courses

URL-1. T.C. Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı. https://www.osym.gov.tr/TR,6548/arastirma-hakkinda.html

URL-2. T.C. Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı. https://www.osym.gov.tr/TR,8789/hakkinda.html

URL-3. UNESCO. *Global Education Monitoring Report.* https://www.unesco.org/gem-report/en/monitoring-sdg4

URL-4. Millî Eğitim Bakanliği Ölçme ve Değerlendirme Yönetmeliği. *Resmî Gazete,* 32304. T.C. Cumhurbaşkanlığı Mevzuat Bilgi Sistemi. https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=40317&MevzuatTur=7&MevzuatTertip=5

URL-5. Millî Eğitim Bakanlığı Ortaöğretim Kurumları Yönetmeliği. *Resmi Gazete*, 28758. T.C. Cumhurbaşkanlığı Mevzuat Bilgi Sistemi. https://mevzuat.gov.tr/mevzuat?MevzuatNo=18812&MevzuatTur=7&MevzuatTertip=5

Vandenberg, R. J. & Lance, C. E. (2000). A review of MI literature. *Organizational Research Methods, 3*, 4–69. https://doi.org/10.1177/109442810031002

World Wide Web Consortium. (2025). *Web content accessibility guidelines (WCAG) 2.3*. https://www.w3.org/WAI/standards-guidelines/wcag/

Zumbo, B. D. (1999). *A handbook on the theory and methods of DIF*. Directorate of Human Resources Research and Evaluation.

Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where ıt ıs going. *Language Assessment Quarterly, 4*(2), 223–233. https://doi.org/10.1080/15434300701375832

Zumbo, B. D., Sireci, S. G. & Hambleton, R. K. (2003, Nisan). Re-visiting exploratory methods for construct comparability: Is there something to be gained from the ways of old?. Paper presented at the National Council on Measurement in Education, Chicago, IL.

YAZAR HAKKINDA | ABOUT AUTHOR

DOÇ. DR. ASİYE ŞENGÜL AVŞAR

Recep Tayyip Erdoğan Üniversitesi | asiye.sengul[at]erdogan.edu.tr

ORCID: 0000-0001-5522-2514

Doç. Dr. Asiye Şengül Avşar, Recep Tayyip Erdoğan Üniversitesi Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalında öğretim üyesidir. Lisansüstü eğitimini Ankara Üniversitesi Eğitimde Ölçme ve Değerlendirme alanında tamamlayan Şengül Avşar, doktora sonrası araştırmacı olarak 2017-2018 yılları arasında Hollanda Tilburg Üniversitesi, Metodoloji ve İstatistik bölümünde, 2025-2026 yılları arasında Amerika Birleşik Devletleri'nde Georgia Institute of Technology Üniversitesi'nde psikoloji bölümünde bulunmuştur. Uzmanlık alanları arasında; veri bilimi, sosyal bilimlerde istatistik uygulamaları, öğrenci başarısının ölçülmesi, ölçme aracı geliştirme ve uyarlama, madde tepki kuramı modelleri ve uygulamaları, bireysel test puanlarının geçerliğinin belirlenmesi konuları yer almaktadır.

ASSOC. PROF. DR. ASİYE ŞENGÜL AVŞAR

Recep Tayyip Erdoğan University | asiye.sengul[at]erdogan.edu.tr

ORCID: 0000-0001-5522-2514

Assoc. Prof. Dr. Asiye Şengül Avşar is a faculty member in the Department of Measurement and Evaluation in Education at the Faculty of Education, Recep Tayyip Erdoğan University. Having completed her graduate studies in Measurement and Evaluation in Education at Ankara University, Şengül Avşar conducted postdoctoral research at Tilburg University (Netherlands) in the Department of Methodology and Statistics (2017–2018) and at the Georgia Institute of Technology (USA) in the Department of Psychology (2025–2026). Her areas of expertise include data science, statistical applications in the social sciences, the measurement of student achievement, the development and adaptation of measurement instruments, item response theory models and applications, and the determination of the validity of individual test scores.