



HİTİT
ÜNİVERSİTESİ
UZAKTAN EĞİTİM U.A.M.

YÜKSEKÖĞRETİMDE YETERLİLİK TEMELLİ ÖLÇME-DEĞERLENDİRME KILAVUZU

Modül 1: **Ölçme ve Değerlendirmede** **Temel Kavramlar**



hitit.edu.tr/tr/uzem



@hitituzem

KASIM - 2025



REKTÖRÜN MESAJI



İçinde yaşadığımız dijital çağın dinamikleri eğitimi yeniden düşünmeyi gerektiren birtakım zorluklar sunmaktadır. Özellikle yükseköğretim kurumlarını yeniden yapılandırmak, akademik gerekliliklerin yanı sıra geleceğin mesleklerinin isterlerini gözeterek üniversitelerdeki öğrenme-öğretme süreçlerini yeniden tasarlamak ve buna paralel olarak da ölçme-değerlendirme sürecini sadece bilgi odaklı değil aynı zamanda çeşitli beceri ve yetkinlikleri kapsayacak şekilde performans ve süreç temelli olarak yeniden düşünmeyi zorunlu kılmaktadır. Kısaca eğitimi yeniden düşünmek ve bunun için de ölçme-değerlendirmeyi yeniden düşünmek bir zorunluluk halini almıştır.

Günümüzde öğrenmenin yalnızca kısa süreli, düzey belirleyici, sonuç odaklı sınavlarla ölçülmesi ve buna bağlı olarak öğrenme kanıtları oluşturulması yeterli değildir. Öğrencinin bilgiye ulaşma, bilgiyi yapılandırma, problem çözme ve öğrenme sürecinde aktif rol alması ve kendi öğrenme sürecini yönetmesi öne çıkmaktadır. Bu dönüşüm, yükseköğretimde ölçme-değerlendirme anlayışının klasik yöntemlerin ötesine geçerek katılımcı, kapsayıcı, öğrenci katılımını teşvik eden, gerçek yaşam problemlerine odaklanan, otantik, süreç odaklı, öğrenmeyi biçimlendirici ve yeterlilik temelli ölçme değerlendirme uygulamalarını zorunlu kılmaktadır.

Hitit Üniversitesi olarak, 2026-2030 Stratejik Planında vurgulanan öğrenci odaklılık, dijital dönüşüm ve sürekli iyileştirme ilkeleri doğrultusunda öğretim-öğrenme süreçlerinin niteliğini artırma hedefine katkı sağlamak amacıyla kurum genelinde süreç odaklı, öğrenmeyi biçimlendirici ve yeterlilik temelli ölçme-değerlendirme uygulamalarını teşvik ediyoruz. Bu doğrultuda hazırlanan “Yükseköğretimde Yeterlilik Temelli Ölçme-Değerlendirme Kılavuzu”, öğretim elemanlarına derslerinde kullanabilecekleri çağdaş, kanıta dayalı ve uygulamaya dönük yöntemleri sunmak amacıyla tasarlanmıştır.

Bu kılavuz, yükseköğretim düzeyinde öğrenme çıktılarını daha etkili biçimde ölçmek, öğrencinin gelişimini sürece yayılan değerlendirmelerle izlemek ve alternatif ölçme-değerlendirme araçlarının sunduğu olanakları ölçme-değerlendirme kültürüne entegre etmek için pratik örnekler içermektedir. Böylece her bir ders, öğrencinin bilgi, beceri ve yetkinliklerini bütüncül biçimde ortaya koyabileceği bir öğrenme ortamına dönüşecektir.

Bu çalışmanın hazırlanmasında emeği geçen Uzaktan Eğitim Uygulama ve Araştırma Merkezi ekibine, ölçme-değerlendirme alanında katkı sunan akademisyenlerimize ve kılavuzun geliştirilmesine destek veren tüm paydaşlara teşekkür ederim. Bu kılavuzun, öğretim elemanlarımızın ders tasarımlarında ve değerlendirme uygulamalarında yenilikçi yaklaşımları benimsemelerine katkı sağlamasını diliyorum.

Prof. Dr. Ali Osman ÖZTÜRK
Hitit Üniversitesi Rektörü

SUNUŞ

Günümüz yükseköğretim anlayışı, öğrencilerin yalnızca teorik bilgiyi ezberlemesini değil, bu bilgiyi kullanarak analiz, sentez, problem çözme, eleştirel düşünme ve yaratıcılık gibi üst düzey beceriler kazanmasını hedeflemektedir. Yapay zeka çağında, yükseköğretim kurumlarından beklenen, "soru çözen değil, sorun çözen" mezunlar yetiştirmektir. Bunu gerçekleştirebilmek için en kritik adımlardan biri de, öğretme-öğrenme süreçlerini ve özellikle de ölçme- değerlendirme uygulamalarını yeniden yapılandırmaktır.

Bu doğrultuda, Hitit Üniversitesinin 25.09.2024 tarih ve 2024/132 sayılı Senato kararı ile derslerde ölçme ve değerlendirme yöntemlerinin çeşitlendirilmesi tavsiye edilmiş ve bu kapsamda tüm eğitim birimlerinde öğrenci merkezli ve süreç odaklı bir ölçme-değerlendirme yaklaşımı benimsenmiştir. Bu kararı destekleyen kritik bir değişiklik olarak, üniversitemiz genelinde geleneksel "ara sınav haftası" uygulaması kaldırılmış ve ölçme-değerlendirmenin bir hafta süreli dar bir sürece sıkıştırıldığı uygulamanın yerine tüm sürece yayıldığı yeni bir anlayışa geçilmiştir. Sınav haftasının kaldırılması, öğretim elemanları ve öğrenciler için sonuç odaklı, geleneksel düzey belirleyici sınava dayalı bir ölçme-değerlendirme yaklaşımının ötesine geçerek; süreç odaklı, yetkinlik-temelli ve öğrenme sürecini biçimlendirici ölçme-değerlendirme yöntemlerinin (*proje, portfolyo, performans görevi, saha çalışması, olgu sunumu, rubrik, derecelendirilmiş ölçek, gözlem formu, öz/akran değerlendirme gibi*) kullanımını teşvik edici bir pencere açmıştır.

Bu bağlamda hazırlanan Yükseköğretimde Yeterlilik Temelli Ölçme Değerlendirme Kılavuzu ile (1) öğretim elemanlarına yetkinlik temelli ölçme değerlendirme yöntemleri konusunda kavramsal çerçeve sunmak, (2) teorik bilgilerin yanı sıra yetkinlik temelli yöntemlerin çeşitli derslere nasıl entegre edileceği konusunda yol göstermek, (3) öğrenme kanıtlarının nasıl oluşturulacağına dönük pratik öneriler sunmak ve (4) birim bazında hali hazırda uygulanan örnekleri paylaşarak görünür kılmak amaçlanmıştır.

Üniversitemiz eğitim-öğretim kalitesinin iyileştirilmesine katkı sağlaması dileğiyle...

Hazırlayanlar

Doç. Dr. Mehmet Kemal AYDIN

Doç. Dr. Metin KUŞ

Öğr. Gör. Servet Ebrar BAYRAM

İlayda ERASLAN

Aysima TEMÜR

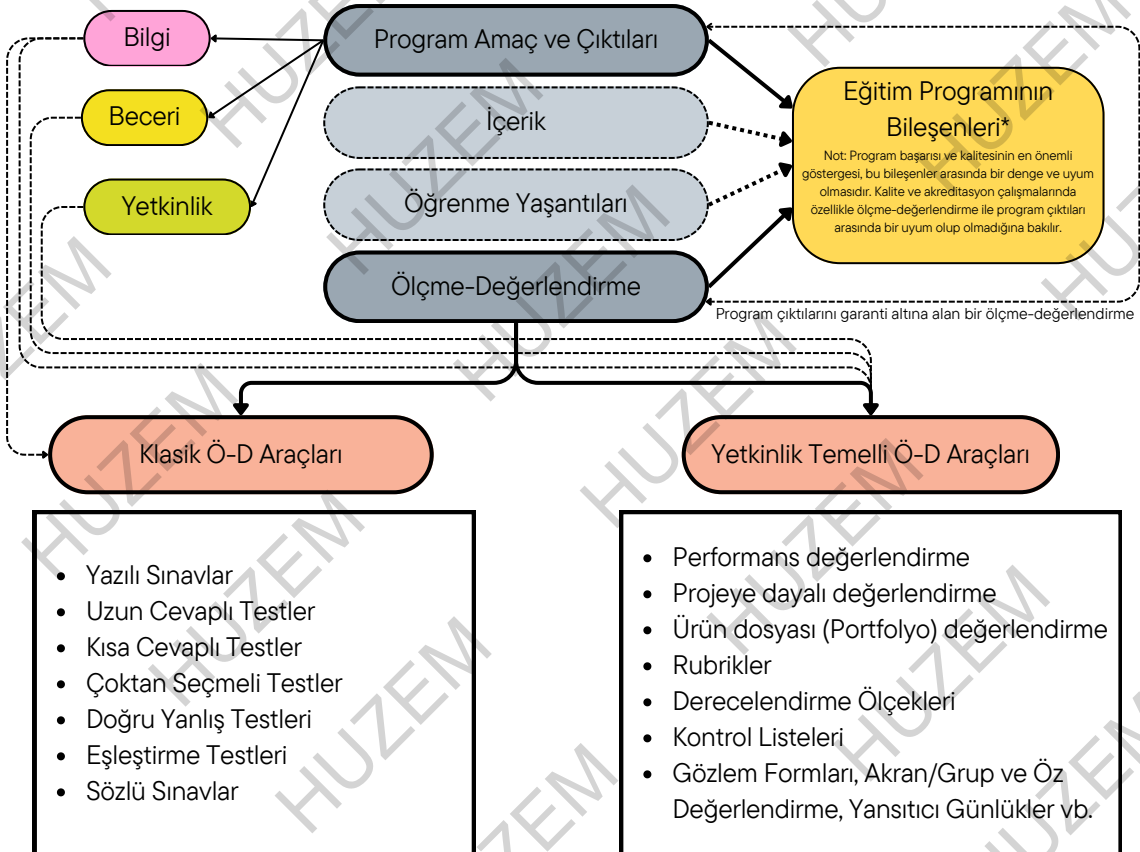
KISALTMALAR

TYYÇ	Türkiye Yükseköğretim Yeterlilikler Çerçevesi
Ö-D	Ölçme-Değerlendirme
YTD	Yeterlik Temelli Değerlendirme
ÖÇ	Öğrenme Çıktıları
KR 20-21	Kuder - Richardson 20-21 Yöntemleri

YÖNETİCİ ÖZETİ

Uzun yıllar boyunca yükseköğretimde ölçme ve değerlendirme anlayışı, büyük ölçüde klasik sınavlara dayalı, sonuç odaklı ve bilgi düzeyini ölçmeye yönelik bir yapıda süregelmiştir. Bu geleneksel model, öğrencilerin dönem sonlarında ulaştıkları bilişsel çıktıları ölçmede etkili olsa da, bilginin uygulanma biçimini ve üst düzey becerilerin gelişimini yeterince görünür kılamamıştır. Oysa çağdaş yükseköğretim anlayışı, öğrenciden yalnızca “ne bildiğini” değil, aynı zamanda “bilgiyi nasıl kullandığını” ve “ne yapabildiğini” de göstermesini talep etmektedir.

Bu doğrultuda, Türkiye Yükseköğretim Yeterlilikler Çerçevesi (TYYÇ), yükseköğretim kurumlarından, bilgi, beceri ve yetkinliklerden oluşan program çıktıları/yeterliliklerini bütüncül biçimde ölçen değerlendirme sistemleri geliştirmesini beklemektedir. Bu beklentiyle uyumlu olarak, Hitit Üniversitesi Senatosu’nun 25.09.2024 tarihli ve 2024/132 sayılı kararı ile ölçme yöntemlerinin çeşitlendirilmesi teşvik edilmiş; ayrıca geleneksel “sınav haftası” uygulamasının kaldırılarak değerlendirmelerin tüm döneme yayılması yönünde önemli bir adım atılmıştır. Böylece, değerlendirme yalnızca öğretimin sonunda yapılan ve bilişsel süreçlere odaklanan bir etkinlik olmaktan çıkıp, öğrenmeyi destekleyen sürekli bir süreç hâline gelmiştir. Şekil 1’de yeterlilik temelli ölçme uygulamaları ve program yeterlilikleri arasındaki ilişki gösterilmiştir.



Şekil 1. Eğitim Programlarının Bileşenleri ve Ölçme-Değerlendirme İlişkisi

Ölçme-değerlendirme anlayışındaki bu paradigma değişimi, öğretim elemanları için yeni pedagojik fırsatlar ve sorumluluklar doğurmuştur. Yeterlilik temelli değerlendirme, öğrencilerin bilgiye sahip olma düzeylerini belirlemeyi amaçlamanın ötesinde, alanıyla ilgili bilgileri gerçek yaşam bağlamlarında nasıl kullandıklarını, problem çözme, eleştirel düşünme, sorumluluk alma ve bağımsız çalışma gibi yetkinliklere ne düzeyde sahip olduklarını belirlemeyi hedefler. Bu çerçevede, performans görevlerinin, projelerin ve portfolyoların nasıl tasarlanacağı; rubrikler, dercelendirilmiş ölçekler ve kontrol listeleri gibi puanlama araçlarının nasıl kullanılacağı; öğrencilere nasıl etkili geri bildirim verileceği gibi pratik bilgiler içeren bir kılavuz hazırlanması gerekliliği ortaya çıkmıştır.

Bu bağlamda hazırlanan bu kılavuz ile özellikle mesleğe yeni başlamış öğretim elemanlarının ölçme-değerlendirme yetkinliklerini geliştirmek ve mevcut öğretim elemanlarının ölçme uygulamalarını yeniden yapılandırmalarında yol göstermek, çağdaş değerlendirme yaklaşımlarına ilişkin kavramsal bir çerçeve sunmak ve farklı fakültelerden derlenen uygulama örnekleri ile desteklenene uygulama temelli bir yol haritası oluşturmak amaçlanmıştır. Tablo 1'de bazı birimlerde öne çıkan beceriler ve yetkinlikler ile bunların ölçülmesinde ve değerlendirilmesinde etkili olabilecek bazı yöntemlere yer verilmiştir.

Tablo 1.1.Fakülte/Birirlere Göre Yetkinlik Temelli Ölçme-Değerlendirme Örnekleri

Fakülte/ Birim	Ölçülecek Beceri ve Yetkinlikler	Uygun ÖD Yöntemi	Örnek Senaryolar
Sağlık Bilimleri Fakültesi	Klinik karar verme, analitik düşünme, psikomotor beceriler, güvenlik prosedürlerini uygulama.	Vaka analizi/Olgu Sunumu, Performans Değerlendirme, Portfolyo Değerlendirme	Vaka Analizi/Olgu Sunumu: Hemşirelik öğrencileri “akut solunum yetmezliği olan bir hastanın bakım planını” hazırlayıp çözüm yollarını tartışır. Performans Görevi: Öğrenciden laboratuvarında bir beceriyi uygulaması istenir ve performansı “Hijyen”, “Doğru prosedürü uygulama” gibi kriterleri içeren bir rubrikle değerlendirilir. E-Portfolyo: Beslenme ve diyetetik öğrencileri “bir aylık diyet danışmanlığı sürecini” raporlar ve hasta iletişimi üzerine yansıtıcı günlük yazarlar.
Tıp Fakültesi	Klinik karar verme, analitik düşünme, hasta yönetimi, duyuşsal ve psikomotor beceriler, etik prosedürleri uygulama, etkili iletişim	Nesnel Yapılandırılmış Klinik Sınav (OSCE), Otantik Değerlendirme, Vaka analizi/Olgu Sunumu, Portfolyo Değerlendirme, Yansıtıcı Günlük	Nesnel Yapılandırılmış Klinik Sınav: Öğrencilerin, “hasta öyküsü alma ve fiziksel muayene” istasyonlarında performans göstermesi beklenir, kontrol listeleri ya da rubriklerle performans değerlendirilir. Vaka Analizi/Olgu Sunumu: Öğrenciler belirli bir hasta üzerinden “tanı koyma ve tedavi planı oluşturma” süreçlerini yazarak analiz eder, Öz değerlendirme formu ile kendi performanslarını değerlendirirler. Portfolyo Değerlendirme: Staj dönemindeki öğrenciler staj süresince öğrenme kanıtlarını içeren ürün dosyası oluştururlar ve “hasta-hekim iletişimi” üzerine haftalık yansıtma raporları hazırlarlar.

Tablo 1.1.(devamı)

Fakülte/ Birim	Ölçülecek Beceri ve Yetkinlikler	Uygun ÖD Yöntemi	Örnek Senaryolar
Mühendislik ve Doğa Bilimleri Fakültesi	Problem çözme, analitik düşünme, araştırma tasarlama ve uygulama, etkili iletişim.	Performans Değerlendirme, Projeyle Dayalı Değerlendirme, Portfolyo Değerlendirme, Kavram Haritaları	• Performans Görevi: Elektrik-Elektronik mühendisliği öğrencileri “akıllı sensör devresi” geliştirmeye yönelik bir performans görevi verilebilir. • Proje Değerlendirme: Bilgisayar mühendisliği öğrencileri, “mobil sağlık uygulaması” geliştirme projesi, sonuçların sunumu ve analitik rubrikle öğrencilerin performansının değerlendirilmesi. • Kavram Haritası: Kimya mühendisliği öğrencilerinden “reaksiyon kinetiği ve katalizör ilişkisini” gösteren bir kavram haritası hazırlamaları istenebilir.
Spor Bilimleri Fakültesi	Psikomotor beceriler, antrenman programı tasarlama, performans analizi, öğretim becerileri, etkili iletişim.	Performans Değerlendirme, Otantik Değerlendirme, Portfolyo Değerlendirme	• Performans Görevi: Antrenörlük öğrencilerinden, “10-12 yaş grubu için koordinasyon odaklı antrenman tasarımları ve uygulamaları” istenir ve bir analitik rubrik ile performans değerlendirilir. • Portfolyo: “Beden Eğitimi Öğretmenliği öğrencilerinde, öğretmenlik uygulaması boyunca gözlemleri, ders planları ve yaptıkları dersleri içeren bir portfolyo oluşturmaları istenebilir. Haftalık gözlemleri için yansıtıcı günlükler oluşturmaları istenir. • Video ile Performans Analizi: Öğrencilerden tenis dersinde, “servis atışı tekniklerini” videoya kaydedip öz ve akran değerlendirme yapmaları istenir.
İlahiyat Fakültesi	Dinî metinleri anlama, yorumlama, duyuşsal ve psikomotor beceriler, sosyal beceriler ve etkili iletişim.	Performans Değerlendirme, Vaka Analizi/Olgu Sunumu, Yansıtıcı Günlükler	• Performans Görevi: Öğrenciler “hutbe veya vaaz metni hazırlayıp sunar,” iletişim ve içerik uygunluğu rubrikle değerlendirilir. • Vaka Analizi/Olgu Sunumu: Öğrenciler, senaryo gereği bir vatandaşın faiz konusunda yaşadığı “ahlakî bir ikilem karşısında İslam düşüncesi perspektifinden çözüm” geliştirirler. • Yansıtıcı Günlükler: Öğrencilerden “Kuran-ı Kerim öğretiminde karşılaştıkları pedagojik zorluklar” üzerine haftalık yansıtma yazıları yazmaları istenir.
İktisadi ve İdari Bilimler Fakültesi	Problem çözme, karar verme, eleştirel ve analitik düşünme, stratejik düşünme, etkili iletişim	Performans Değerlendirme, Otantik Değerlendirme, Projeyle Dayalı Değerlendirme, Vaka Analizi/Olgu Sunumu,	• Performans Görevi: Öğrencilerden, bir örnek işletmenin muhasebe defterleri (veya dijital muhasebe kayıtları) üzerinden cari yıl içerisinde oluşan stopaj vergisi yükümlülüklerini tespit etmeleri, ilgili kanun, oran ve kesinti türlerini belirlemeleri, ardından doğru beyan ve ödeme planı hazırlamaları beklenmektedir. • Vaka Analizi/Olgu Sunumu: Öğrencilerden, yerel bir kamu idaresi sınırlarında yaşandığı varsayılan bir su kesintisi ve halk tepkisi vakasını analiz etmeleri beklenmektedir. Vaka, kamu yönetiminin temel fonksiyonları (planlama, koordinasyon, iletişim, hesap verebilirlik) çerçevesinde incelenmeli; öğrencilerden kriz yönetimi sürecinde alınan kararların doğruluğunu, kamu etiği açısından uygunluğunu ve alternatif politika önerilerini değerlendirmeleri istenir.

Tablo 1.1. (devamı)

Fakülte/ Birim	Ölçülecek Beceri ve Yetkinlikler	Uygun ÖD Yöntemi	Örnek Senaryolar
İnsan ve Toplum Bilimleri Fakültesi	Kaynak tarama, dokümantasyon, eleştirel ve analitik düşünme, araştırma tasarlama ve uygulama, sergileme ve sunum becerileri, yazılı ve sözlü iletişim.	Performans Değerlendirme, Projeyle Dayalı Değerlendirme, Portfolyo Değerlendirme, Kavram Haritaları Vaka Analizi/Olgü Sunumu	• Performans Görevi: Arkeoloji öğrencilerinden üniversite bahçesi veya laboratuvar ortamında kurulan yapay kazı alanında, buluntuları sınıflandırma, belgeleme ve yorumlama sürecini uygulamaları istenir. Öğrencilerin performansı Kazı yöntemi ve ekip çalışması, belgelerin düzeni ve doğruluğu, bulguların yorumlanması, zaman yönetimi ve iş güvenliği gibi çeşitli becerileri içeren bir analitik rubrik ile değerlendirilir. • Portfolyo Değerlendirme: Türk Dili öğrencilerinden seçki dosyası şeklinde “<i>Edebi Türlerde Yaratıcı Yazma Portfolyosu</i>” başlıklı, dönem boyunca yazdıkları kısa öykü, şiir, deneme ve çözümleme metinlerini içeren bir dijital portfolyo hazırlamaları istenir. Öğrenci performansı yansıtma raporu, öz değerlendirme formu ve çeşitli becerileri içeren (Dil bilgisi kurallarını uygulama, yazılı anlatım yetkinliği, yaratıcı yazma becerisi, türlere uygun metin seçimi vb.) bir analitik rubrikle değerlendirilir. • Vaka Analizi: Tarih bölümü öğrencilerinden, belirli bir olayı (ör. II. Meşrutiyet’in ilanı) farklı kaynak türleri (gazete, hatırat, arşiv belgesi) üzerinden analiz ederek elde ettikleri bilgileri sentezleyerek yorumlamaları istenir. Öğrencilerin performansı, kaynak kullanımı, eleştirel ve analitik düşünme, akademik yazım ve sentez becerilerini içeren bir analitik rubrikle değerlendirilir.
Güzel Sanatlar, Tasarım ve Mimarlık Fakültesi	Estetik düşünme, Yaratıcılık, Eleştirel düşünme, teknik beceriler, duyuşsal ve psikomotor beceriler, tasarım becerileri.	Performans Değerlendirme, Otantik Değerlendirme, Portfolyo Değerlendirme	• Performans Görevi: Resim öğrencilerinden, canlı model veya obje karşısında gözleme dayalı bir resim oluşturmaları istenir. Oluşturdukları resimle ilgili akran değerlendirmesi ya da öz değerlendirme formu ile değerlendirme yapmaları beklenir. • Proje: Müzik öğrencilerinden “Yerel Ezgilerin Çağdaş Düzenlemesi” başlıklı bir proje hazırlamaları ve bu kapsamda öğrencilerin, bir Türk halk ezgisini seçip çağdaş müzik formlarında yeniden düzenlemeleri istenir. Yaratıcılık, teknik yeterlilik (armoni, ritim, tempo), geleneksel-çağdaş sentezleme gibi becerileri analitik rubrik ile değerlendirilir. • Portfolyo: Resim bölümü öğrencilerinden seçki dosyası şeklinde “Sanatsal Gelişim Dosyası” hazırlayarak dönem boyunca ürettikleri çalışmalar, eskizler, proje raporları ve yansıtıcı notlarını içeren dijital veya basılı portfolyo hazırlamaları istenir.

İÇİNDEKİLER

REKTÖRÜN MESAJI.....	III
SUNUŞ.....	V
KISALTMALAR.....	VI
YÖNETİCİ ÖZETİ.....	VII
İÇİNDEKİLER.....	XI
Temel Kavramlar.....	1
Ölçme.....	2
Değerlendirme.....	3
Değerlendirme Yaklaşımları.....	7
Geleneksel Değerlendirme Yaklaşımları.....	8
Ölçüte Göre Değerlendirme Türleri.....	8
Mutlak Değerlendirme.....	8
Bağıl Değerlendirme.....	9
Amaca Göre Değerlendirme Türleri.....	10
Tanıma ve Yerleştirmeye Yönelik Değerlendirme.....	11
Bıçimlendirme ve Yetiştirmeye Yönelik Değerlendirme.....	11
Düzy Belirleme/Değer Bıçmeye Yönelik Değerlendirme.....	12
Geçerlilik.....	13
Kapsam Geçerliliği.....	13
Görünü Geçerliliği.....	15
Ölçüt Geçerliliği.....	15
Yapı Geçerliliği.....	16
Güvenirlik.....	17
Kararlılık.....	17
Tutarlılık.....	17
Duyarlılık.....	17
Objektiflik.....	18
Puanlama Güvenirliği ve Artırma Yolları.....	19
Güvenirliği Belirleme Yöntemleri.....	20
İki Uygulamalı Yöntemler.....	20
Test-Tekrar-Test Yöntemi.....	20
Paralel Testler Yöntemi.....	21
Tek Uygulamalı Yöntemler.....	23
Testi Yarılama Yöntemi.....	22
Kuder-Richardson 20-21 Yöntemleri.....	22
Cronbach - Alfa Yöntemi.....	22
Kullanışlılık.....	23
KAYNAKÇA.....	24



ASSESSMENT-OF-LEARNING- ASSESSMENT FOR LEARNING



Öğrencinin öğrenme sürecinin sonunda neyi ne kadar öğrendiğine odaklanan klasik "öğrenmenin değerlendirilmesi" anlayışının yerini, günümüzde öğrencinin öğrenme sürecini iyileştirmeye odaklanan "öğrenme için değerlendirme" anlayışı almıştır.

Temel Kavramlar

Ölçme

Ölçme, en genel anlamıyla bir nesne ya da bireyin belirli bir özelliğinin gözlenerek sayı ya da sembollerle ifade edilmesi sürecidir. Ölçme işleminin temel amacı, incelenen niteliği olabildiğince nesnel, kesin ve güvenilir bir şekilde ortaya koymaktır. Bu nedenle ölçme sürecinde, ölçülecek özelliğe uygun bir ölçme aracının seçilmesi ve doğru şekilde kullanılması önemlidir.

Günlük yaşamda da sıkça karşılaşılan ölçme işlemleri; bir nesnenin uzunluğunu, ağırlığını, sıcaklığını, hızını veya büyüklüğünü belirlemeyi içerir. Bu tür ölçümlerde genellikle standart ölçme araçları (*cetvel, terazi, termometre, kronometre vb.*) kullanılır ve sonuçlar, tanımlı birimlerle ifade edilir. Örneğin, bir odanın büyüklüğü metrekare (m^2), havanın sıcaklığı santigrat derece ($^{\circ}C$) ya da bir masanın yüksekliği santimetre (cm) cinsinden belirtilir. Burada dikkat edilmesi gereken en önemli nokta, kullanılan birimlerin ölçülen özelliklerle uyumlu olmasıdır.

Ölçme yalnızca doğrudan gözlenebilen fiziksel özelliklerle sınırlı olmayıp; bireylerin zekâ, tutum, kişilik veya benzeri psikolojik özelliklerinin belirlenmesinde de kullanılmaktadır. Ancak bu tür özelliklerin ölçümünde kullanılan araçların geçerlilik ve güvenilirliği daha da kritik hale gelir.

Bu bağlamda ölçme, yalnızca belirli amaca yönelik olarak özelliklerin gözlemlenmesi ve sayı ya da sembole ifade edilmesi süreci iken; değerlendirme, ölçme sonuçlarının ölçüte vurularak anlamlandırılması ve buna dayalı karar alma sürecini ifade eder. Dolayısıyla ölçme ile değerlendirme birbiriyle ilişkili ancak farklı kavramlardır (Kızlık, 2012). Ancak ölçme, değerlendirme için ön koşuldur.



Değerlendirme

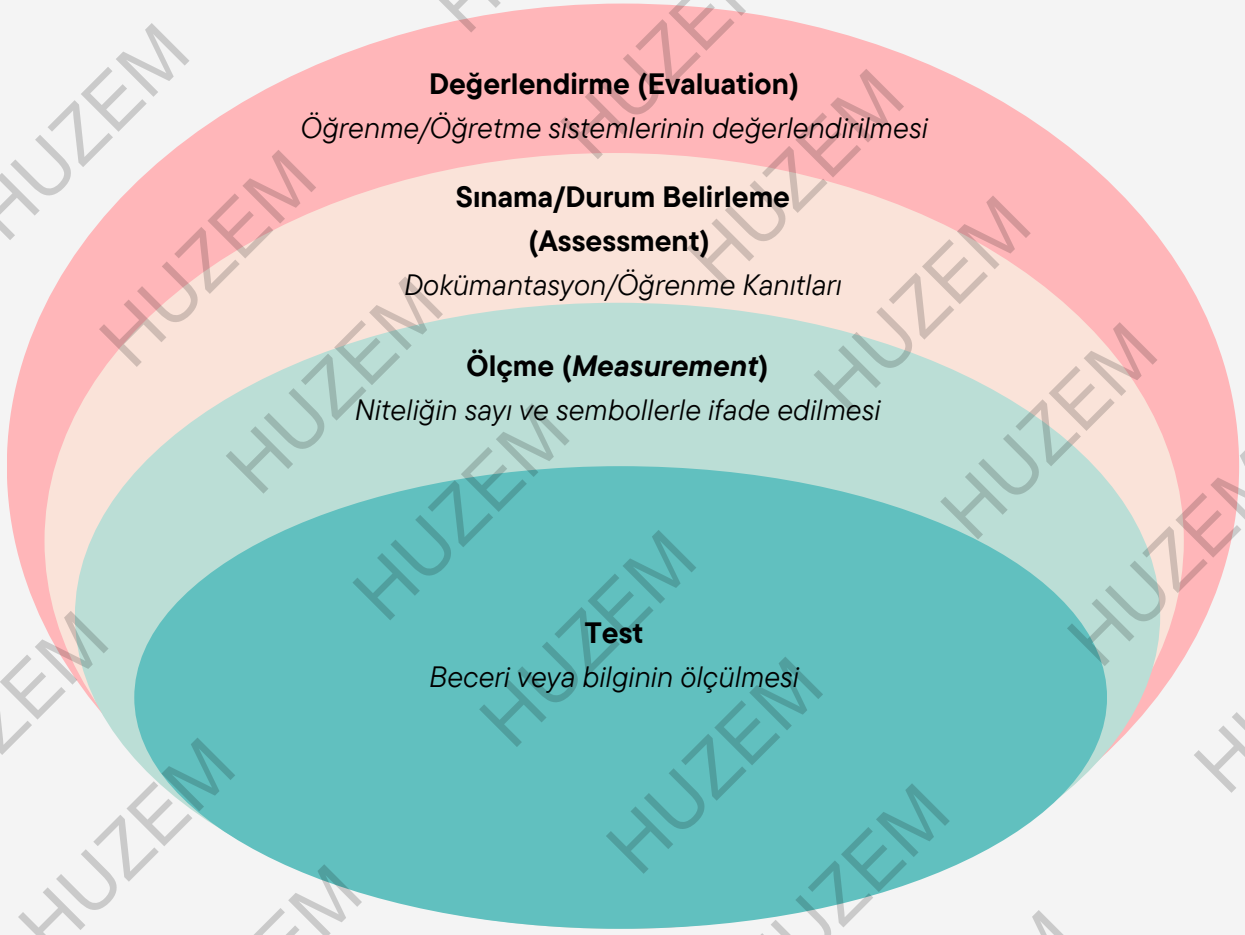
Değerlendirme, en genel anlamıyla belirli amaçlar doğrultusunda önceden belirlenmiş ölçütler dikkate alınarak ürün, hizmet veya performans hakkında bilinçli bir yargıya varma ya da karar alma süreci olarak tanımlanır (Karaağaçlı, 2002). Eğitim bağlamında değerlendirme; programların, derslerin ve hedeflerin ne ölçüde gerçekleştirildiğini ortaya koyan, öğretim sürecinin başında, süreç içinde ve sonunda yürütülebilen sistematik bir etkinliktir.

Değerlendirme sürecinde bireylerin bilgi, beceri ve tutumları farklı yöntemlerle ortaya konulabilir. Örneğin; akademik beceriler sınavlar veya standart testlerle doğrudan ölçülebilen, kavrayış ve anlayış gibi daha soyut nitelikler dolaylı ölçümlere ve çıkarıma dayalı değerlendirmelere ihtiyaç duyar (Lacy & Williams, 2018). Bu nedenle değerlendirme yalnızca gözlenebilir performansların değil, aynı zamanda öğrenme çıktılarının derinliğini anlamaya ve kıymetlendirmeye yönelik de bir süreçtir.

Değerlendirme çoğu zaman test ve ölçme ile karıştırılsa da, çok daha kapsamlı bir süreci ifade eder. Öğrencilerin yalnızca sınavlardan aldığı puanlar değil, bu puanların önceden belirlenmiş ölçütlere, kriterlere veya akran gruplarına göre yorumlanması da değerlendirme kapsamına girer. Bu noktada Gronlund (1981) değerlendirmeyi şu denklemle açıklamaktadır:

- Değerlendirme = nicel ölçüm + karar verme
- Değerlendirme = nitel ölçüm + karar verme

Bu bağlamda değerlendirme; öğrenci kazanımlarını test etmeyi, ölçmeyi, bunları nitel olarak açıklamayı, elde edilen sonuçların değerini belirlemeyi ve karar sürecini yönlendirmeyi kapsayan çok boyutlu bir süreçtir (Fitzpatrick vd., 2012; Ornstein & Hunkins, 2013).



Şekil 2. Eğitimde Değerlendirmede Temel Kavramlar (Aydın, 2023)

Değerlendirme ölçme sürecini kapsayan daha geniş bir çatı kavramdır. Ancak değerlendirme, yalnızca ölçme sürecini değil, aynı zamanda ölçüm sonuçlarının yorumlanmasını, ölçütle karşılaştırılmasını, değer biçilmesini ve karar verme süreçlerini de içerir. Başka bir ifadeyle, tüm ölçme türleri bir değerlendirme aracıdır fakat tüm değerlendirmeler ölçme aracı değildir.

Değerlendirme, öğrencilerin öğrenme düzeylerini ortaya koymanın yanı sıra, öğretim sürecinin niteliğini belirlemede de kritik rol oynar. Öğretim elemanı, değerlendirme yoluyla öğrencilerin gelişimlerini izler, davranışlarındaki değişimleri gözlemler ve öğretim yöntemlerinin etkililiğini değerlendirir (Fitzpatrick vd., 2012). Aynı zamanda, eğitim sisteminde alınan kararlar (örneğin sınıf tekrarı, program geliştirme veya politika düzenlemeleri) doğrudan değerlendirme sonuçlarına dayanır. Söz konusu sistemsel kararlar, genellikle anlık ve tek bir sınav sonucuna göre değil, öğrencilerin belirli bir süreç boyunca gösterdikleri kümülatif gelişime ve programın hedeflerine ulaşma düzeyine bakılarak alınır. Bu durum, değerlendirmenin yalnızca kısa vadeli sonuçlara değil, aynı zamanda uzun vadeli gelişim ve öğrenme süreçlerine dayalı olduğunu göstermektedir (Stukalina, 2012).

Ölçme ve değerlendirme süreçleri, birbirinden bağımsız unsurlar değildir. Gerçekleştirilen her ölçme eyleminin mutlaka bir amacı bulunmalıdır; Zira ölçme süreci sonunda ulaşılmak istenen nihai hedef değerlendirmedir (Başokçu vd., 2025).

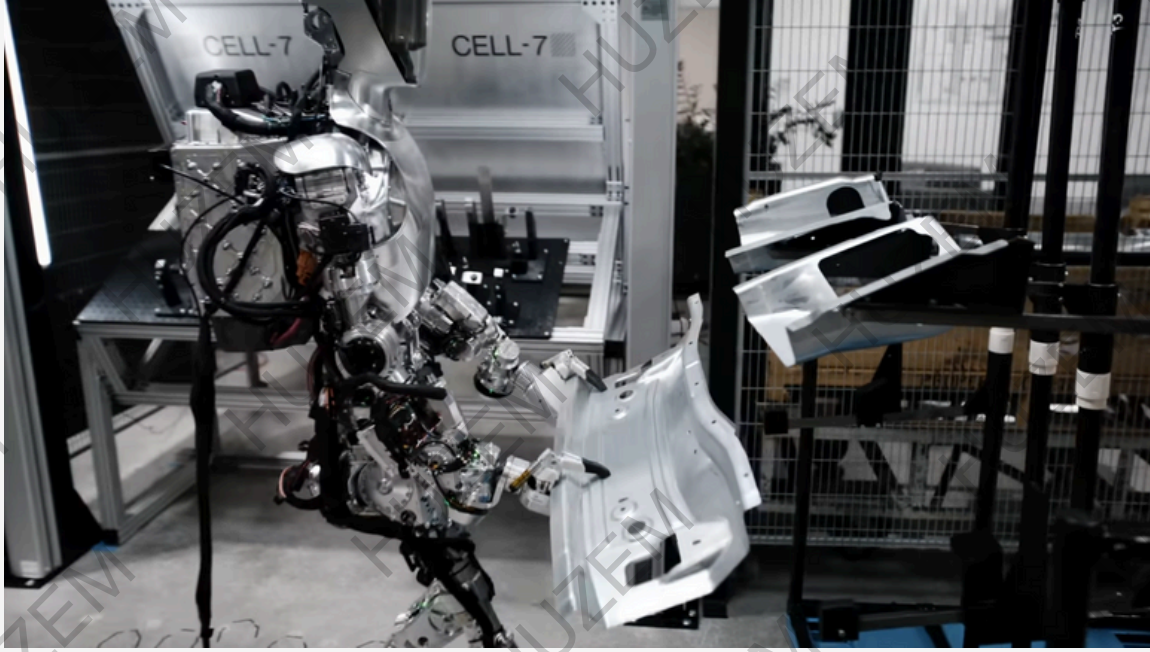
Dolayısıyla, değerlendirme aşamasından kopuk ve yalıtılmış bir ölçmenin herhangi bir anlam taşıması mümkün değildir. Bir ölçme işlemini, amacından veya sonuçlarına göre yapılacak olan değerlendirmeden ayrı olarak ele almak, bu süreç sonunda varılacak yargıların hatalı olmasına kaçınılmaz bir şekilde yol açacaktır. Buradaki en kritik nokta, ölçme süreci ile benimsenecek değerlendirme yaklaşımı arasındaki uyumun doğru bir şekilde tesis edilmesidir.



Microsoft
Copilot



Gemini



Şekil 3. Avrupalı otomotiv üreticileri, üretim sürecinde robotları kullanmaya başladı.

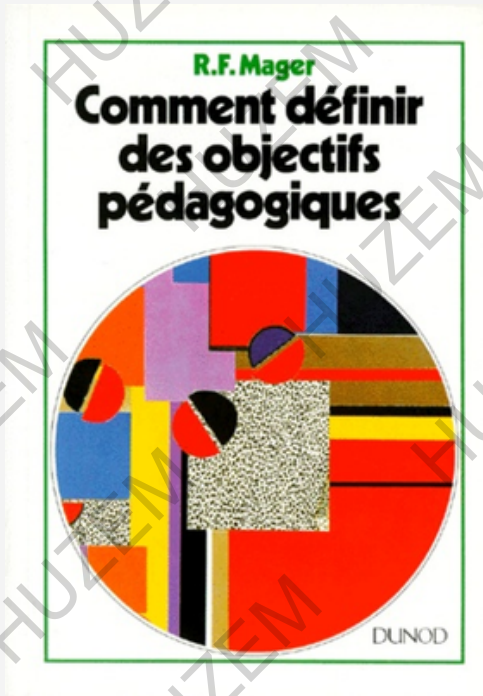
Bilmeye cesaret et! (Sapere Aude!) sloganıyla başlayan Aydınlanma Çağından günümüze kadar bilgi en büyük güçtü. Ancak içinde yaşadığımız yapay zekâ çağında bilgi artık güç değil. Doğru bilgiye ulaşma yollarını bulmak, bilginin kökenini, kıymetini, sınırlarını belirlemek ve bilgiyi yeni bilgiler inşa etmek için kullanmak günümüzde en büyük güçtür.

Geleneksel ölçme-değerlendirme anlayışı genellikle bilgiye sahip olma (ezber/hatırlama) düzeyine odaklanırken, her an erişilebilir olan bu bilgiyle ne yapabildiğimiz günümüzde daha önemli hale gelmiştir. Yeterlilik temelli değerlendirme yöntemleri, öğrencinin bilgiyi kullanma becerilerini (analiz etme , eleştirel düşünme , problem çözme, iletişim) ve yeni bilgi oluşturma (sentez, yaratıcılık, tasarım vb.) yetkinliklerini geliştirmeye odaklanır.

Değerlendirmenin kendisi de, uygulanma amacına bağlı olarak farklı türlere ayrılır. Bu bağlamda, temelde üç tür değerlendirmeden bahsedilebilir:

Tablo 1.2. Amacına Göre Değerlendirme Türleri

Tanılayıcı (Diagnostic) Değerlendirme	Biçimlendirici/Süreç (Formative) Değerlendirme	Düzy Belirleyici/Sonuç (Summative) Değerlendirme
Öğrenme sürecinin başında	Öğrenme süreci devam ederken	Öğrenme sürecinin sonunda
Öğrencinin hazırbulunuşluk düzeyini belirleme	Öğrencinin gelişimini izleme ve anında geri bildirim sağlama	Öğrencinin hedeflenen kazanımlara ne ölçüde ulaştığını belirleme
Öğrencinin eksik veya güçlü yönlerini belirleme	Öğrencinin hatalarını erken aşamada fark etme	Dönem sonu sınavları, proje değerlendirmeleri ve bitirme testleri vb.



Başarının kanıtı olarak neyi alacağınızı bile bilmiyorsanız, birisine “başarılıdır” damgasını vurmak oldukça şüpheli bir davranıştır...

R. F. Mager

Değerlendirme Yaklaşımları

Değerlendirme, eğitim programlarının etkililiğinin ve verimliliğinin ortaya konması ve programların yenilenmesi ya da iyileştirilmesi için kritik bir öneme sahiptir. Eğitim sistemlerine yön veren felsefeler değiştiğinde kullanılan değerlendirme anlayışları da buna bağlı olarak değişmektedir. Ölçme değerlendirmede geleneksel ve çağdaş olmak üzere iki farklı değerlendirme yaklaşımı vardır.

Tablo 1.3. Geleneksel ve Çağdaş Değerlendirme Yaklaşımlarının Karşılaştırılması

Geleneksel Değerlendirme	Çağdaş Değerlendirme
Daimicilik ve esasicilik eğitim felsefelerine dayanır.	İlerlemecilik ve yeniden kurmacılık eğitim felsefelerine dayanır.
Öğretim elemanı merkezlidir. Öğretim elemanı tek başına karar alır.	Öğrenci merkezlidir. Değerlendirme sürecinde öğretim elemanı, öğrenci, akran grupları, veli ve yöneticiler söz hakkına sahiptir.
Sonuç odaklıdır.	Süreç-ürün odaklıdır.
Tek bir alanı değerlendirmeye alır. Genelde bilişsel alanı değerlendirir.	Öğrencinin tüm gelişim alanlarını değerlendirir. Bilişsel, duyuşsal, psikomotor alanı değerlendirir.
Öğrenci hakkında kesin karar almayı amaçlar.	Öğrencilerin geliştirilmesine yönelik çalışır.
Objektif, nesnel ve tarafsızdır.	Subjektif, öznel ve yalındır.

Geleneksel Değerlendirme Yaklaşımları

Geleneksel değerlendirme, öğretim elemanı merkezli olup öğrencinin istenilen özelliklere sahip olup olmadığını belirlemek için yapılmaktadır. Değerlendirmede kullanılan ölçüt ve değerlendirmenin amacına göre iki ana başlıkta ele alınmaktadır.

Ölçüte Göre Değerlendirme Türleri

Değerlendirme, yalnızca amacına göre (*tanılayıcı, biçimlendirici, düzey belirleyici*) sınıflandırılmaz. Aynı zamanda, öğrenci performansını yorumlarken temel alınan referans noktasının, yani 'ölçütün' ne olduğuna göre de farklılaşır. Bir öğrencinin başarısı, ya önceden belirlenmiş, sabit bir standarda göre ya da içinde bulunduğu grubun performansına göre yorumlanabilir. Bu bağlamda değerlendirme, kullanılan ölçütün türüne göre mutlak değerlendirme ve bağıl değerlendirme olmak üzere iki temel biçimde incelenebilir.

Mutlak Değerlendirme

Değerlendirme sürecinde temele alınan ölçüt, değerlendirmenin türünü belirler. Mutlak ölçüt kullanılarak alınan kararlara mutlak değerlendirme denilmektedir.

- Öğretim programı (Hedefler)
- Derste ele alınan konular yani öğrenme eksiklikleri
- Testte kullanılan sorular
- Geçme puanı (baraj puanı)
- Öğretim elemanının belirlediği kesin kriterlere bakarak kararın alınmasına mutlak değerlendirme denilmektedir.

Örnekler:

- Dersten 100 üzerinden 70 ve üzeri puan alanların geçmesi
- 100 sorudan 50'sini yapanın başarılı kabul edilmesi
- Soruların %60'ını doğru cevaplayanların mülakata çağırılması
- Program hedeflerinin %70'ini kazanmış olanların başarılı sayılması
- Kapsamın %50'sini öğrenmiş olanların başarılı olması
- Muafiyet sınavından 80 ve üzeri alanların dersten muaf sayılması

Bağıl Değerlendirme

Değerlendirme sürecinde temele alınan ölçüt, değerlendirmenin türünü belirler. Bağıl ölçüt kullanılarak alınan kararlara bağıl değerlendirme denilmektedir.

- Çan eğrisine göre
- Öğrenciler karşılaştırılırken
- Belirli bir kontenjana öğrenci seçerken
- Öğrenciler arasında en yeteneklilerin seçiminde
- Ülke geneli normlara göre değerlendirme yapılırken kullanılan değerlendirmeler bağıl değerlendirmelerdir.

Bağıl değerlendirmeler, bireyin başarı düzeyini göreceli standartlara göre ölçerken; mutlak değerlendirmeler, önceden belirlenmiş ölçüt ve standartlara göre bireyin yeterliliğini değerlendirir. Bu iki yaklaşımın seçimi, testten geçen öğrenci sayısını ve başarı yorumlarını doğrudan etkileyebilir (Sternberg vd., 2022; Livingston, 1972; Cuhadar & Gelbal, 2021).

Örnekler

- Sınıf ortalaması üstünde puan alanların başarılı kabul edilmesi
- ALES puanı en yüksek olan ilk üç öğrencinin lisansüstü programa kabul edilmesi
- Resim yarışmasına katılan en iyi üç resmin belirlenmesi
- Sınava girenlerin (adayların) % 70'inden başarılı olanların mülakata çağırılması
- Sınıfta öğrenme düzeyi bakımından en geride kalan son beş kişinin belirlenmesi
- Sınavda en yüksek başarı gösteren ilk % 10'luk dilim içinde kalanların belirlenmesi
- Öğrenme düzeyi bakımından sınıfın son % 20'lik dilim içinde olanların belirlenmesi
- Uzun mesafe koşusunu ilk 10'da tamamlayan öğrencilerin okul koşu takımına alınması
- Öğrencilerin başarılarının Z ve T standart puanlarına göre değerlendirilmesi

Tablo 1.4. Ölçüte Göre Değerlendirme Türleri

Değerlendirme Türü	Amacı	Güçlü Yönleri	Zayıf Yönleri
Bağıl Değerlendirme	Öğrencinin başarısını diğer öğrencilerle kıyaslamak	Öğrencinin grup içindeki performansını ortaya koyar	Öğrencinin mutlak bilgi ve becerilerini göstermez
Mutlak Değerlendirme	Öğrencilerin bilgi ve beceri düzeylerinin belirli standartlara ulaşp ulaşmadığını saptamak	Öğrencilerin bilgi ve becerileri açısından mutlak performansını ortaya koyar	Ölçme-değerlendirme uygulaması daha zordur

Amaca Göre Değerlendirme Türleri

Eğitimde değerlendirmeler belli amaçlar doğrultusunda yapılmaktadır. Bu amaçlar; öğrencileri tanımak ve uygun bir programa yerleştirmek, öğrencinin süreç içerisinde gelişimini takip etmek ve öğrencilerin program sonunda öğrenme çıktılarına ne kadar ulaştığını belirlemek şeklinde üç ana başlık altında toplanabilir.



Şekil 4. Amacına Göre Değerlendirme Türleri

Stake'e göre çorbayı aşçı tadarsa biçimlendirici/süreç değerlendirme, müşteri tadarsa düzey belirleyici/sonuç değerlendirmeden söz edilir (Scriven, 1991). Çorba için gerekli malzemeler ve bunların kalitesi için yapılan değerlendirmeler de tanılayıcı/girdi değerlendirmeye örnektir.

1. Tanıma ve Yerleştirmeye Yönelik Değerlendirme

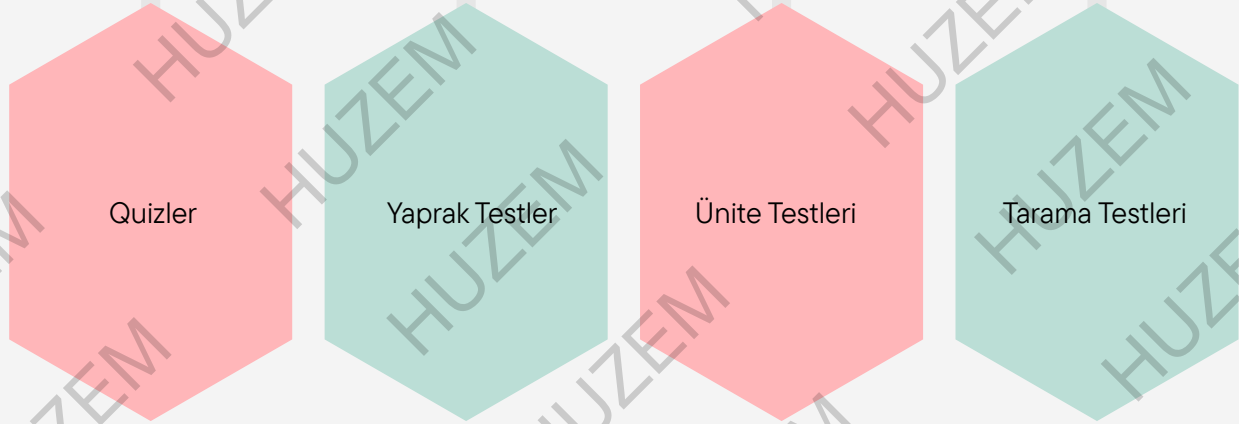
- Eğitim süreci başında öğrencilerin hazırbulunuşluk düzeylerini belirlemek, güçlü ve zayıf yanlarını belirleyip öğrencileri uygun bir programa yerleştirmek amacıyla yapılmaktadır.
- Öğretim hedefleriyle ilgili olarak öğrencilerin önceden neleri öğrenmiş olduklarını ve öğrenme ihtiyaçlarını ortaya çıkarmak için yapılan değerlendirmedir.
- Öğrencilere öğretilecek olan özelliklere sahip olup olmadıklarını, ön koşul davranışların öğrencilerde bulunup bulunmadığını belirlemek için yapılan değerlendirmedir.
- Tanıma ve yerleştirmeye yönelik değerlendirme öğrenciye not vermek amacıyla yapılmaz.



2. Biçimlendirme ve Yetiştirmeye Yönelik Değerlendirme

- Eğitim süreci devam ederken her konu sonunda öğrencilerin konuyla ilgili öğrenme eksikliklerini belirleyip gidermek amacıyla yapılan değerlendirmedir.
- Öğrencilerin bir konu ile ilgili öğrenip öğrenmediğini, eksikliği olup olmadığı hakkında yorum yapılır.
- Birbirine bağlı konuların yani birbirinin ön bilgisi şeklinde olan üniteler arası geçişi kolaylaştırır.
- İzlemeye yönelik değerlendirme öğretim elemanının ve öğretim faaliyetlerinin verimliliğini artırır.
- Formatif yapıda hazırlanan bir test üniteye işlenen bütün kritik kazanımları içermelidir. Bundan dolayı en önemli özelliği kapsam geçerliliğidir.
- Ünite ve konu sonunda uygulanan testlerin amacı not vermekten ziyade öğrencilerin gelişimlerini izlemektir.

Biçimlendirme ve yetiştirmeye yönelik değerlendirmede kullanılacak ölçme araçları:



3. Düzey Belirleme/Değer Biçmeye Yönelik Değerlendirme

- Öğretim süreci sonunda veya birkaç ünite sonunda yapılan değerlendirmedir.
- Öğrencilerin hedef davranışları kazanıp kazanmadığını kazanmışsa ne düzeyde kazanmış diye bakılan değerlendirme anlayışıdır.
- Öğrenciler hakkında geçti - kaldı, başarılı - başarısız kararları alınmaktadır. Bundan dolayı bu değerlendirme anlayışında amaç not vermek olabilir.
- Öğretim süreci sonunda yapılırken ara ara bir kaç ünite sonunda da yapılabilir.
- Düzey belirleyici değerlendirmeler öğrencilerin belge, sertifika, diploma alıp alamayacaklarını belirleyen değerlendirmelerdir.

Biçimlendirme ve yetiştirmeye yönelik değerlendirmede kullanılabilir ölçme araçları:

Vize/Final
Sınavları

Eriş Testleri

Mezuniyet
Sınavları

Sertifika
Sınavları

Geçerlilik

Geçerlilik, bir ölçme aracının amaca hizmet etme derecesidir. Geçerlilik bir ölçme aracında bulunması gereken en önemli özelliktir. Ölçme aracı seçilirken ilk bakılması gereken özellik geçerliliğidir. Diğer bir ifadeyle, geçerlilik bir ölçme aracının ölçmeye konu olan özellik dışında başka bir özelliği ölçmemesidir. Örneğin, öğrencilerin Türkçe becerilerini ölçmek için sınav uygulayan bir öğretim elemanı sınavları okurken öğrencinin güzel yazısına puan verirse sınavın geçerliliği olumsuz yönde etkilenir. Ölçme sonuçlarının kullanım amacına göre geçerlilik farklılık göstermektedir. Öğrencilerin sınav başarılarını ölçmek için geçerli olan bir araç öğrencilerin gelecekteki başarılarını yorumlama konusunda geçerli olmayabilir. Bundan dolayı ölçme aracının amacına göre geçerlik türleri ortaya çıkmaktadır.

Geçerliliği 4 başlıkta ele almak mümkündür:

Kapsam Geçerliliği

Testteki soruların konuları ve hedef davranışları temsil etme gücüdür. Kapsam, testte ölçülmek istenen içeriktir. Eğer ki sorular kritik olan içeriği temsil edebiliyorsa kapsam geçerliliği yüksek kabul edilmektedir. Kapsam geçerliliğini artırmak için yapılabilecek en etkili uygulama soruların hazırlanırken belirtke tablosuna uygun olarak hazırlamaktır. Kapsam geçerliliğini anlayabilmek için belirtke tablosunun iyi bir şekilde bilinmesi gerekmektedir.

Belirtke Tablosu

Belirtke tablosu bir derste kazanımlarla içerik arasındaki bağlantıyı gösteren iki boyutlu tablodur. Hedeflerin niteliksel ve niceliksel boyutu ile ilgili bilgi verir. Ünitelerin içeriğinin hedef davranışlarla nasıl ilişkilendirileceğini gösterir.

	Bilgi	Kavrama	Uygulama	Analiz	Sentez	Değerlendirme
1. Ünite	✓	✓				
2. Ünite			%20	%40		%65

Tablo 1.5. Eğitimde Araştırma Yöntemleri Dersi Yarıyıl Sonu Sınavı Belirtke Tablosu
Örneği

Bilişsel beceri düzeyi *	Öğrenme Çıktıları	Soru Sayısı
Bilgi	ÖÇ1: Bilimsel araştırma ile ilgili temel kavramları, terimleri, yöntemleri, etik ilkeleri ve araştırma sürecinin genel yapısını tanımlar.	4
Kavrama	ÖÇ2: Bilimsel bir araştırma raporunun (makale, bildiri, tez vb.) bölümlerini okuyarak elde ettiği sonuçları yorumlar.	4
Uygulama	ÖÇ3: Bilimsel araştırma sürecinin aşamalarını sırasıyla uygular. ÖÇ4: Bilimsel araştırma sürecini etik ilkelere uygun yürütür.	2
Analiz	ÖÇ4: Farklı araştırma yöntem ve desenlerini araştırma amacı, verilerin toplanması, örneklem seçimi ile verilerin analiz ve yorumlanması boyutları açısından karşılaştırır.	2
Sentez	ÖÇ6: Belirlediği bir araştırma problemine dayalı olarak, APA 7 kurallarına uygun biçimde özgün küçük kapsamlı bir araştırma önerisi hazırlar.	1
Değerlendirme	ÖÇ7: Verilen araştırma raporunu (makale, tez vb.) bilimsel yayın etiği ilkelerine uygunluk ve yöntemsel uygunluk kriterleri açısından değerlendirerek çalışmanın güçlü ve zayıf yönlerine ilişkin gerekçeli yargılarda bulunur.	1

*Klasik Bloom taksonomisi temel alınmıştır.

Görünüş Geçerliliği

Bir testin neyi ölçüyorsa onun gibi görünmesidir. Görünüş geçerliliği testin dış dünyasıdır. Dışarıdan bakan kişilerin testin hangi derse ait olduğunu bilmesi veya tahmin edebilmesidir. Türkçe testi Türkçe becerisini ölçen teste benzemesi görünüş geçerliliğinin yüksek olduğu anlamına gelir. Bundan dolayı öğrenci ve velilerin hakkında görüş bildirebileceği tek geçerlik türüdür. Görünüş geçerliliğini bazı durumlarda yükseltmek bazı durumlarda ise gizlemek gerekir. Örneğin psikolojik özellikleri ölçen testlerde görünüş geçerliliği aranmaz.

Ölçüt Geçerliliği

Bir testin geçerliliği çalışılırken başka bir testin ölçüt olarak alınmasıdır. Ancak her test ölçüt olarak kabul edilmemektedir. Ölçüt alınacak olan test ulusal düzeyde geçerliliği ve güvenilirliği yüksek kabul edilen bir test olmalıdır. Ölçüt olarak alınacak olan test uzmanlar tarafından hazırlanan bir test olmalıdır. Hazırlanan bir zeka testinin ölçütü başka bir zeka testi olacaktır. Ancak ölçüt alınan zekâ testi rehberlik ve araştırma merkezleri tarafından hazırlanan test olmalıdır. Ölçüt geçerliliği iki başlıkta ele alınabilir:

Tablo 1.6. Ölçüt Geçerliliği Türleri

Uygunluk Geçerliliği	Yordama Geçerliliği
- Bir testin geçerliliği hakkında yorum yapılırken öğrencilerle ilgi daha önceden elde edilen aynı özelliği yansıtan birtakım ölçütler belirlenir. - Bir testin geçerliliği hakkında yorum yapılırken öğrencilerin testten önceki göstermiş oldukları becerilerin test sonuçları ile arasındaki korelasyona bakılmaktadır. Örneğin öğrencilerin üniversite giriş sınavı sonuçlarının uygunluk geçerliliği belirlenmek isteniyorsa lise başarı puanları ölçüt olarak kabul edilip puanlar arasındaki ilişkiye bakılmalıdır. Öğrencilerin lise başarı puanları yüksek, üniversite başarı puanları da yüksekse veya öğrencilerin lise başarı puanları düşük, üniversite giriş sınavı başarıları da düşükse testin uygunluk geçerliliği yüksek demektir.	- Öncelikle yordama kavramının tanımına bakmak gerekmektedir. Yordama billinenlerden hareketle bilinmeyen tahmin etmek demektir. Yordama geçerliliği öğrencilerin sınav sonuçlarından hareketle gelecekteki performanslarını yani süreci tahmin etmektir. Örneğin üniversite giriş sınavının yordama geçerliliği çalışılırken öğrencilerin sınavdan almış oldukları puan ile üniversite sürecinde göstermiş oldukları performans karşılaştırılır. Üniversite giriş sınavında yüksek puan alan öğrenciler üniversitede başarılı düşük puan alan öğrenciler başarısızsa sınavın yordama geçerliliği yüksek kabul edilir. - Yordama geçerliliği özellikle seçme amaçlı kullanılan testlerin en önemli özelliğidir.

Yapı Geçerliliği

Yapı birbiriyle ilişkili olduğu düşünölen ögeler arası ilişki olarak tanımlanmaktadır. Yapı geçerliliği bir testin psikolojik - teorik özelliği olarak tanımlanmaktadır. Yapı geçerliliği psikolojik özelliklerin ölçümünde kullanılan testlerin en önemli özelliğidir. Psikolojik özelliklerin bir çok alt boyutu vardır. Bu alt boyutların tamamı birbiriyle uyumlu bir şekilde ölçölmelidir. Bundan dolayı yapı geçerliliği testin içerisinde olan uyumu temsil etmektedir. Bir testin bilenle bilmeyeni ayırt edip bu ayrımı tam olarak yansıtmısına yapı geçerliliği denir. Öğrencilerin bildiği konulardan yüksek bilmediği konulardan düşük puan almasıdır. Bu durumu olumsuz etkileyen her faktör yapı geçerliliğini düşürmektedir. Olumlu etkileyen her faktör yapı geçerliliğini artırmaktadır.

Yapı geçerliliği belirlemede birçok yol vardır. Ölçölen özellik bakımından farklı olduğu bilinen iki gruba uygulanır ve gerçekten ayırıp ayırmadığına bakılabilir. Geçerliliği güvenilirliği yüksek bir testten başarılı olan bireylere uygulanır ve alınan sonuçlar karşılaştırılabilir. Öğrenci yapıları kısa zamanda değişmemektedir. Bundan dolayı uygulanan testin uzun zamanda kararlı olup olmadığına bakılmaktadır. Sonuçlar arası bir kararlılık varsa testin yapı geçerliliği yüksek kabul edilmektedir.

Tablo 1.7. Geçerliliği Etkileyen Faktörler

Geçerliliği Etkileyen Faktörler

- Bir sorunun bilenle bilmeyeni ayırt etmesi geçerliliği artırır.
- Testte yer alan soru sayısını artırma geçerliliği artırır. Ancak bir yere kadar artıracağı unutulmamalıdır bir yerden sonra olumsuz etkileyecektir.
- Testte yer alan soruların yıllara göre değiştirilmesi geçerliliği artırır.
- Doğru cevabın seçeneklere dengeli dağıtılması geçerliliği artırır. Örneğin 80 soruluk olarak hazırlanan 5 seçenekli bir testte her seçenekten 16 tane olması geçerliliği artırır.
- Soruların bilgisayar ortamında çoğaltılması geçerliliği artırır çünkü bazı yanlış anlamaların önüne geçecektir.
- Aynı seçeneğin üst üste doğru cevap olarak gelmemesi geçerliliği artırır. Aksi bir durum geçerliliği olumsuz etkiler.
- Soruların kaynaklardan olduğu gibi alınması, kaynak kitaplar açık sınav yapmak, soruların verilmesi geçerliliği düşürür.
- Şans başarısı ve kopya çekmek geçerliliği düşürür.
- Sabit ve sistematik hatalar geçerliliği düşürür.

Güvenirlik

Güvenirlik, bir ölçme işleminin hatalardan arınık olmasıdır. Güvenirliğin en doğru tanımı bir ölçme aracının tesadüfi hatalardan arınık olmasıdır. Sabit ve sistematik hataların güvenilirlik üzerinde bir etkisi yoktur. Güvenirlik ayrıca ölçme sonuçlarının kararlı, tutarlı, duyarlı ve objektif sonuçlar vermesi olarak tanımlanmaktadır. Bu kavramları tek tek açıklayarak ele aldığımızda güvenilirlik kavramının en geniş tanımı yapılmaktadır.

Kararlılık

Ölçme aracının geniş zaman diliminde benzer sonuçlar vermesidir. İstatistiksel olarak bu zaman aralığı 3-6 hafta olarak tanımlanmaktadır. Bir başka tanımla kararlılık aynı ölçme aracıyla yapılan ölçme sonuçlarının birbirine yakın, benzer sonuçlar vermesi olarak tanımlanmaktadır.

Tutarlılık

Ölçme aracının yakın zaman diliminde benzer sonuçlar vermesidir. İstatistiksel olarak bu zaman aralığı 0-3 hafta olarak tanımlanmaktadır. Bir başka tanımla tutarlılık aynı özelliği ölçen farklı araçların birbirine yakın sonuçlar vermesidir. Tutarlılığın bir alt kavramı olan iç tutarlılık ise ölçme aracındaki birimlerin birbiriyle uyumlu olmasıdır. Bir testte yer alan soruların tek boyutlu uyum içerisinde olması olarak tanımlanmaktadır.

Duyarlılık

Duyarlılık ölçme aracındaki birimlerin birbirine yaklaşması olarak tanımlanmaktadır. Yani birimler arası mesafenin küçülmesidir. Örneğin, gram ölçümleri yapan bir terazi ile kilogram ölçümleri yapan bir terazi karşılaştırıldığında, gram ölçümleri yapan terazi daha duyarlıdır. Eğitimde kullanılan testler üzerinden örnek vermek gerekirse, bir konu alanıyla ilgili hazırlanmış 20 soruluk bir test ile aynı konu alanıyla ilgili hazırlanmış 100 soruluk bir test karşılaştırıldığında, 100 soruluk bir test daha hassas yani daha duyarlı sonuçlar verecektir.

Objektiflik

Objektiflik, ölçme sonuçlarının tarafsız, nesnel, yansız sonuçlar vermesidir. Ölçme işlemini kim yaparsa yapsın, ölçmeyi yapan kişiyle aynı (veya çok benzer) sonuca ulaşması olarak tanımlanmaktadır.

Puanlama güvenirliği olarak da bilinen objektiflik, özellikle açık uçlu sorular, projeler, portfolyolar ve sunumlar gibi yeterlilik temelli değerlendirme yöntemlerinde kritik önem taşır. Puanlama güvenirliği temel olarak iki açıdan incelenir:

Puanlayıcılar Arası Güvenirlik	Farklı puanlayıcıların, aynı öğrenci performansına (örn. aynı proje veya sınav kâğıdı) ne kadar benzer puanlar verdiğidir. Bu güvenirlik, örneğin bir performansı öğrencileri tanımayan farklı bir uzmanın da puanlaması ve bu puanlar arasındaki korelasyona bakılmasıyla test edilebilir.
Puanlayıcı İçi Güvenirlik	Bir puanlayıcının, aynı öğrenci performansını farklı zamanlarda puanlaması durumunda verdiği puanlar arasındaki kararlılığıdır. Bu, puanlayıcının kendi içinde tutarlı ve tarafsız davranmasını ifade eder. Bunu sağlamanın en iyi yolu, puanlayıcının puanlama yaparken bir cevap anahtarı veya rubrik kullanmasıdır.

Puanlama Güvenirliđi ve Artırma Yolları

Yeterlilik temelli deęerlendirme yöntemlerinde ve yazılı sınavlarda puanlama güvenirliđini (objektifliđi) saęlamak ve yanlılıđı en aza indirmek için ařağıdaki somut stratejiler uygulanmalıdır:

Tablo 1.8. Puanlama Güvenirliđi ve Artırma Yolları

1. Genel İzlenimle Puanlama	Puanlayıcının, öğrenci kâğıdını bir bütün olarak okuyup, edindiđi genel kanaate göre (örn. "bu kâğıt iyi," "bu zayıf") bir not vermesidir. Puanlama güvenirliđi en düşük yöntemdir, çünkü tamamen subjektif yargılara ve puanlayıcının anlık ruh haline açıktır.	Puanlama güvenirliđi en düşük ↑ ↓ Puanlama güvenirliđi en yüksek
2. Sınıflama Yoluyla Puanlama	Puanlayıcının, öğrenci kâğıtlarını "Çok İyi", "İyi", "Orta", "Zayıf" gibi önceden belirlenmiř kategorilere ayırması ve ardından bu kategorilere not atmasıdır.	
3. Sıralama Yoluyla Puanlama	Puanlayıcının, tüm öğrenci kâğıtlarını en iyiden en kötüye doęru fiziksel olarak sıralaması ve bu sıraya göre notları dağıtmasıdır.	
4. Ayrıntılı Puanlama Anahtarı ile Puanlama	Puanlayıcının, her soru veya kriter için beklenen cevapları ve bunlara verilecek puanları net bir şekilde tanımlayan bir cevap anahtarı, kontrol listesi veya rubrik kullanmasıdır. Puanlama güvenirliđi en yüksek yöntem budur.	

Güvenirliği Belirleme Yöntemleri

Güvenirliği belirlemek için tek uygulamaya ya da çift uygulamaya dayalı yöntemler tercih edilebilir.

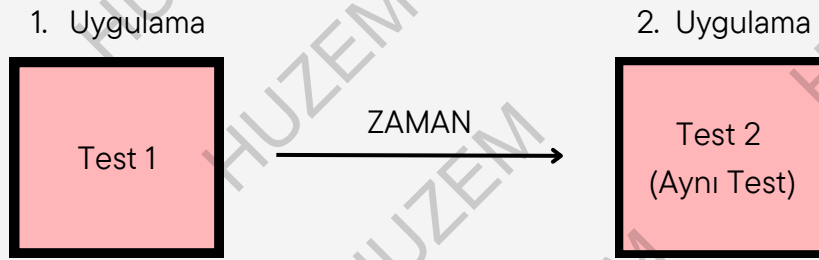
Tablo 1.9. Güvenirliği Belirleme Yöntemleri

İki Uygulamalı Yöntemler	Tek Uygulamalı Yöntemler
- Test - Tekrar - Test Yöntemi - Paralel Testler Yöntemi	- Testi Yarılama Yöntemi - Kuder Richardson (KR- 20) - Kuder Richardson (KR- 21) - Cronbach Alfa

İki Uygulamalı Yöntemler

Test - Tekrar - Test Yöntemi

Bir test uygulandıktan belli bir süre sonra aynı test hiçbir değişiklik yapılmadan aynı öğrencilere tekrardan uygulanır. İki uygulama arasındaki korelasyona bakılır. Test tekrar test yönteminde korelasyon katsayısı yüksek çıkarsa sonuçların kararlılığı yüksektir, düşük çıkarsa sonuçların kararlılığı düşüktür şeklinde yorum yapılır.



Asıl Amaç Kararlılık Katsayısı

Test tekrar test yönteminde sonuçları etkileyen en önemli faktör iki uygulama arasında geçecek olan zamandır. Zaman uzun tutulmamalı uzun tutulursa öğrenciler yeni öğrenmeler gerçekleştirebilir ve ilişki düşük çıkar. Zaman kısa tutulursa iki uygulama arasına ilişki yüksek çıkar. Bundan dolayı zaman ayarlanırken öğrencilerin soruları unutacağı kadar, bilgiyi unutmayacakları kadar bir zaman geçmelidir.

Test tekrar test yönde zaman içerisinde hızlı değişim gösteren bilişsel becerileri ölçen testlerin güvenilirliği hesaplanırken çok fazla kullanılmamaktadır. Daha çok zaman içerisinde değişime direnç gösteren psikolojik özelliklerin ölçüldüğü testlerin güvenilirliğini belirlemek için kullanılır. (Zekâ, yetenek, inanç, kişilik vb.)

Paralel Testler Yöntemi (Eş Değer Testler)

Birbirine eş değer olan iki farklı test farklı zamanlarda öğrencilere uygulanır ve iki uygulama arasındaki korelasyona bakılır. Paralel testler yönteminde iki uygulama arasında korelasyon katsayısı yüksek çıkarsa sonuçların tutarlılığı yüksektir, düşük çıkarsa sonuçların tutarlılığı düşüktür şeklinde yorum yapılır.



Asıl Amaç Tutarlılık Katsayısı

Paralel testler yönteminde iki uygulama arasından geçen zamanın bir önemi yoktur. Test tekrar test yönteminde yer alan zaman dezavantajını ortadan kaldırmıştır. Paralel testler yönteminde iki uygulama arasında sonuçlarda bir tutarsızlık çıkarsa bunun asıl sebebi ikinci testte yer alan sorulardır. Paralel testler yönteminde ikinci uygulama testinde sorular farklı ancak soru gücü, soru tipi, soruların kapsamı aynı olmak zorundadır. Paralel testler yönteminin dezavantajı ikinci uygulama testini birinci uygulama testine eş değer olarak hazırlanmasının zor olmasıdır.

Testi Yarılama Yöntemi

Testi yarılama yöntemi kullanılırken öncelikle bir test öğrencilere uygulanır. Uygulama sonunda test maddeleri tekler - çiftler olmak üzere iki eşit parçaya ayrılır. Test öğrencilere uygulandıktan sonra iki ayrı yarı, ayrı ayrı puanlanır. İki yarı arasındaki puanların korelasyonuna bakılarak testin iç tutarlılığı hakkında bilgi elde edilir. İki yarı arasındaki korelasyon yüksekse testin iç tutarlılığı yüksek, düşükse iç tutarlılık düşük diye yorum yapılır.

Testi yarılama yöntemi testi iki eşit yarıya bölerek güvenilirlik hesaplaması yaptığı için testin tamamının güvenilirliği hakkında bilgi veremez. Testin tamamının güvenilirliğini hesaplamak için Spearman - Brown formülüne ihtiyaç vardır:

$$\text{Testin Tamamının Güvenirliği} = \frac{2. \text{ testin yarısının güvenirliliği}}{1+ \text{ testin yarısının güvenirliliği}}$$

1
2
3
4
5
6
7
8
9
10

1.Uygulama

Tekler
1
3
5
7
9

1. Yarı

Çiftler
2
4
6
8
10

2. Yarı

Asıl Amaç İç Tutarlılık Kat Sayısı

Cronbach - Alfa Yöntemi

KR-20 - KR-21 yöntemlerinin kullanılmadığı çoklu puanlanabilen testlerde kullanılan bir yöntemdir. Alfa yöntemi derecelendirme ölçekleri (Likert ölçekler) ile hazırlanan testlerde kullanılan bir yöntemdir. Bir maddeden alınan puanlar 1-0 şeklinde dönüştürülemezse farklı kategorilerde ele alınıyorsa kullanılan yöntemdir. Alfa yöntemi testin iç tutarlılığını belirlemek için kullanılır. KR yöntemleri ile ilgili yapılan yorumların aynısı Alfa yöntemi için de geçerlidir.

Kuder - Richardson 20-21 Yöntemleri

KR 20-21 formüllerinin kullanılabilmesi için testin öncelikle bazı ön koşulları sağlaması gerekmektedir. Bu yöntemler nesnel özelliğe sahip olan çoktan seçmeli, doğru - yanlış vb. testlerde ve her sorunun aynı puana sahip olduğu testlerde kullanılabilir. Soruyu doğru yapan öğrenci sorunun puanını yanlış yapan öğrenci ise sıfır puan alacak şekilde düzenlenen testlerde hesaplanabilir. Formüller kullanılırken öğrencinin yaptığı her doğru soru için 1 puan yanlış yaptığı her soru için 0 puan uygulaması yapılır.

KR 20-21 yöntemleri kullanılırken test öğrencilere bir kez uygulanır. Test maddeleri analiz edilerek öğrencilerin maddelere vermiş oldukları cevapların birbiri ile uyumu ve testin bütünü ile olan uyumuna bakılır. Yani KR 20-21 formüllerinin asıl amacı testin iç tutarlılığı hakkında bilgi toplamaktır. KR 20-21 yöntemlerinin farklarını belirleyebilmek için bu iki yöntemin formüllerinden hareket etmek gerekmektedir.

Tablo 1.10. KR-20 & KR-21 Güvenirlik Katsayıları

KR-20 (Kuder - Richardson 20)	KR-21 (Kuder - Richardson 21)
• Testin maddelerin farklı güçlükte olduğu durumlarda kullanılır. • Madde güçlük indeksinden hareketle hesaplama yapılmaktadır.	• Testi oluşturan madde güçlük indeksleri aynı veya birbirine çok yakınsa kullanılır. • Hesaplanabilmesi için testin aritmetik ortalama değerine ihtiyaç vardır.
KR-20 katsayısı ve KR-21 katsayısı formüllerine bakıldığında KR-20 madde istatistiklerinden hareket ettiği için daha duyarlı sonuçlar verir. Bundan dolayı KR-20 katsayısı KR-21 katsayısından daha büyük çıkmaktadır. Eğer ki bir testte madde güçlük indeksleri birbirine eşitse KR-20 ve KR-21 güvenirlilik katsayıları birbirine eşit çıkmaktadır. KR-20 ve KR-21 güvenirlilik katsayıları yüksek çıktığında yapılacak yorumlar; • Testin iç tutarlılığı yüksektir. • Testteki sorular birbiriyle uyumludur. • Testteki sorular homojendir. • Test sonuçları hatalardan arındırılır.	

Kullanışlılık

Kullanışlılık, ölçme aracının özellikleri arasında önem sırasına konulduğunda en son sırada yer alır. Bazı durumlarda göz ardı edilebilir. Örneğin, ölçme aracının amacı öğrencilerin sentez düzeyindeki hedef davranışları kazanıp kazanamadığını ölçmek olsun. Sentez düzeyindeki hedef davranışları en iyi ölçen araç yazılı yoklamalardır. Ancak sınıfın kalabalık olması nedeniyle yazılı yoklamanın kullanışlı olmadığı düşünülüp bu aracın kullanılmaması doğru değildir.

Kullanışlılık, bir ölçme aracının zaman, emek ve maliyet açısından verimlilik sağlaması anlamına gelir. Bir testin daha kolay hazırlanması, uygulanması ve puanlanması onun kullanışlı olduğunu gösterir.

KAYNAKÇA

Aydın, M. K. (2023). Değerlendirmeye giriş: Temel kavramlar. M. K. Aydın & M. Kuş (Ed.), *Eğitimde değerlendirme* (s. 2-16). Vizetek.

Başokçu, T. O., Sullivan, A. B., Toprak, E., & Kalkan, Ö. K. (2025). *Sınıf İçi Ölçme ve Değerlendirme Süreçlerinde Öğretmenlerin Güçlendirilmesi: Ölçme ve Değerlendirme Kılavuz Kitabı*. Millî Eğitim Bakanlığı.

Cuhadar, I., & Gelbal, S. (2021). An evaluation of pass/fail decisions through norm- and criterion-referenced assessments. *International Journal of Assessment Tools in Education*, 8(1), 9-20. <https://doi.org/10.21449/ijate.747105>

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2012). *Program evaluation: Alternative approaches and practical guidelines*. Boston: Pearson.

Gronlund, N. E. (1981). *Measurement and evaluation in teaching*. Macmillan Publishing Company.

Karaağaçlı, M. (2002). Değerlendirme süreci. *Mesleki Eğitim ve Teknoloji Eğitimde Özel Öğretim Yöntemleri*, 31-47. Ankara: Nobel.

Kizlik, B. (2012). *Measurement, assessment, and evaluation in education*. Retrieved October 10, 2015.

Lacy, A. C., & Williams, S. M. (2018). *Measurement and evaluation in physical education and exercise science*. Routledge.

Sternberg, R., Chowkase, A., Parra-Martinez, F., & Landy, J. (2022). Criterion-referenced assessment of intelligence as adaptation to the environment: Is it possible, plausible, or practical? *Journal of Intelligence*, 10(3), 57. <https://doi.org/10.3390/jintelligence10030057>

Stukalina, Y. (2012). Addressing service quality issues in higher education: The educational environment evaluation from the students' perspective. *Technological and Economic Development of Economy*, 18(1), 84-98. <https://doi.org/10.3846/20294913.2012.658099>