

Uluslararası Test Standartlarının ve Yönergelerinin Sistematik ve Karşılařtırımlı İncelenmesi

DOI: 10.26466/opus.889770

*

Ezgi Mor Dirlik* Özge Altıntaş** Seval Kula Kartal ***

* Dr. Öğr. Üyesi, Kastamonu Üniversitesi, Eğitim Fakültesi, Kastamonu/Türkiye

E-Posta: emor@kastamonu.edu.tr

ORCID: [0000-0003-0250-327X](https://orcid.org/0000-0003-0250-327X)

** Öğr. Gör. Dr., Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Ankara/Türkiye

E-Posta: oaltintas@ankara.edu.tr

ORCID: [0000-0001-5779-855X](https://orcid.org/0000-0001-5779-855X)

*** Arař. Gör. Dr., Pamukkale Üniversitesi, Eğitim Fakültesi Denizli/Türkiye

E-Posta: sevalk@pau.edu.tr

ORCID: [0000-0002-3018-6972](https://orcid.org/0000-0002-3018-6972)

Öz

Bu çalışmanın temel amacı, dünya genelinde psikolojik testler ve ölçmeyle ilgilenen kurum ve kuruluşlar tarafından psikolojik testlere ve ölçme süreçlerine ilişkin önerilen standartları ve yönergeleri incelemek ve bu konuyla ilgili kaynakları psikometri bilim alanında çalışan test geliştiricilere, uyarlayıcılara ve testlerle ilgilenen arařtırmacılara tanıtmaktır. Bu amaç doğrultusunda, psikolojik testlere ilişkin standartlar geliřtiren kurumların öncüleri olan Amerikan Eğitim Arařtırmaları Birlięi, Amerikan Psikologlar Birlięi ve Eğitimde Ölçme Ulusal Konseyi tarafından önerilen ve güncelleme çalışmaları yapılan test standartları ayrıntılı olarak incelenmiştir. Bunlara ek olarak, test uyarlama ve kültürlerarası test kullanımına ilişkin kapsamlı yönergeler sunan Uluslararası Test Komisyonu tarafından geliřtirilmiř test uyarlama yönergelerine çalışma kapsamında yer verilmiştir. Avrupa Psikologlar Birlięi tarafından geliřtirilen ve kullanıma sunulan Test İnceleme Modeli'nin de tanıtıldıęı çalışmada son olarak, Uluslararası Standartlar Organizasyonu tarafından özellikle iş yaşamında kullanılan psikolojik testlere ve ölçme süreçlerine ilişkin oluşturulan standartlar incelenmiştir. İncelenen standartlara ve yönergelere ilişkin genel deęerlendirmeler yapılmıř, psikolojik ölçme araçları ile ilgilenen tüm arařtırmacılara bu standartların ve yönergelerin kapsamı detaylı biçimde tanıtılmıştır.

Anahtar Kelimeler: *Psikolojik testler, test standartları, test geliřtirme, test uyarlama.*

A Systematic and Comparative Review of International Test Standards and Guidelines

*

Abstract

The purpose of this study is to examine the standards and guidelines recommended by institutions and organizations interested in psychological testing and measurement worldwide and to introduce the relevant literature to those concerned. Within the scope of this purpose, the content of updated test standards recommended by the American Educational Research Association, the American Psychological Association and the National Council on Educational Measurement, which are the pioneers of institutions that develop standards for psychological testing, were examined. Besides, test adaptation guidelines developed by the International Testing Commission were included, which provide comprehensive guidelines on test adaptation and usage of intercultural testing. The Test Analysis Model proposed by the European Federation of Psychologists' Associations was also presented generally, and the standards specifically regarding psychological tests and measurement processes used in employment processes established by the International Standards Organization were examined. General evaluations regarding the listed guidelines and standards were made, and the contents of the guidelines and standards were introduced to the all researchers interested in psychological measurement.

Keywords: *Psychological tests, test standards, test development, test adaptation.*

Giriş

Bireylerin zekâ, kişilik, ilgi, tutum ve güdü gibi örtük özelliklerinin ölçülmesinde, bu özelliklerin göstergeleri olduğu varsayılan maddelerden oluşan psikolojik testler bir diğer adıyla psikolojik ölçme araçları kullanılmaktadır. Psikolojik ölçme aracı, belirli bir psikolojik değişkenin nesnel ve standardize ölçüsü olarak tanımlanmaktadır (Anastasi, 1988). Psikolojik ölçme araçları, bireylerin bilişsel ve duyuşsal özellikleriyle ilgili davranışları hakkında çıkarım yapmak amacıyla; davranışların örneklendirilerek bu örneklerin standartlara göre puanlanıp değerlendirildiği sistematik bir işlem süreci olarak tanımlanmaktadır (Urbina, 2004). Psikolojik ölçme araçlarının geleneksel işlevi, bireyler arasındaki farklılıkları ya da aynı bireyin farklı koşullar altındaki tepki farklılıklarının ölçülmesidir. Eğer farklı bireylerden elde edilen puanlar karşılaştırılabilir olacaksa, test koşulları mutlaka herkes için aynı olmalıdır (Anastasi ve Urbina, 1997). Psikolojik ölçme araçlarının tanımına ilişkin ortak vurgu, maddelerin, yönergelerin ve ölçme aracını yanıtlamak için gereken sürenin ölçme aracını alan her birey için aynı olması anlamına gelen “standart” olma vurgusudur. Psikolojik ölçme araçları için standart uygulama koşulları sağlanamadığında, bu araçlardan elde edilen puanlar, puanlara dayalı olarak bireyler arası yapılan karşılaştırmalar ve onlar hakkında alınan kararlar hatalı olmaktadır. Cronbach’a (1990) göre, test uygulayıcısının ifadeleri ve eylemleri, kullandığı materyaller ve puanlama kuralları sabitlendiğinde testin standartlaştırılmış olduğu kabul edilir; bu durumda o testin farklı zamanlarda ve mekânlarda toplanan puanları tamamen karşılaştırılabilir.

Bireyler hakkında doğru kararlar vermek, bir başka anlatımla, geçerliği ve güvenilirliği yüksek ölçmeler yapmak, ölçme aracı geliştirme/uyarlama aşamasından itibaren uluslararası alan yazında kabul görmüş belirli bir sistematığın takip edilmesini gerektirmektedir. Bu nedenle, birden çok uluslararası kuruluş eğitimde ve psikolojide kullanılan psikolojik ölçme araçlarının geliştirilmesi, uyarlanması, uygulanması ve sonuçların yorumlanıp karara varılması konusunda ciddi çalışmalar yapmaktadır. Bu kuruluşlar, ölçme süreçlerine ilişkin çeşitli standartlar ve yönergeler yayınlamaya çalışarak araştırmacılara bu konuda bilgiler sunmaktadır. Bu amaçla geliştirilmiş ve yaygın kabul görmüş standartlar incelendiğinde, Amerikan

Eğitim Araştırmaları Birliği (American Educational Research Association-AERA), Amerikan Psikologlar Birliği (American Psychological Association-APA), Eğitimde Ölçme Ulusal Konseyi (National Council on Measurement in Education-NCME), Uluslararası Test Komisyonu (International Test Commission-ITC) ve Avrupa Psikologlar Dernekleri Federasyonu (European Federation of Psychologists' Associations-EFPA) gibi kuruluşlar tarafından geliştirilmiş standartlar öne çıkmaktadır. Bu standartlar içerisinde AERA, APA ve NCME tarafından geliştirilen Eğitimde ve Psikolojide Test Standartları (1966, 1974, 1985, 1999, 2014) ise en yaygın olarak bilinen ve kullanılan standartlardır.

Eğitimde ve Psikolojide Test Standartları'nın amacı doğru test uygulamalarını teşvik etmek ve bu uygulamaların niteliklerinin değerlendirilmesinde bir temel oluşturmaktır. Bu nedenle, AERA, APA ve NCME tarafından testleri geliştiren, seçen, test sonuçlarının teknik niteliklerini yorumlayan ya da değerlendiren bireylerin kullanımı için test standartları geliştirilmektedir. Standartların geliştirilmesinin ardındaki temel amaç ilgili taraflara yasal bir yaptırım uygulamak için bir temel oluşturmak değil; eğitimde ve psikolojide bireylerin özelliklerinin ölçülmesi sürecine dahil olan tüm taraflar için ortak ölçütler oluşturmaktır. Yetenek, başarı, tutum, ilgi ve kişilik gibi özellikleri ölçmenin amaçlandığı süreçlerde, işe alım için yapılan görüşmelerde ve çeşitli klinik uygulamalarda test standartlarının uygulanabileceği ifade edilmektedir. Bunlara ek olarak, okul performans değerlendirmeleri gibi standardize edilmemiş ölçme süreçlerinde yapılacak incelemelerde de standartlardan yararlanılabileceği belirtilmektedir (AERA, APA ve NCME, 2014).

Psikolojik testler ve ölçmeler için belli standartlar geliştirip ve bu standartların evrensel olarak kullanımını sağlayarak yaygınlaştıran kuruluşlardan bir diğeri ITC'dir. ITC, eğitimde ve psikolojide kullanılan ölçme araçlarının uygun şekilde geliştirilmesini, değerlendirilmesini ve kullanılmasını teşvik etmektedir. Bünyesinde psikolojik testlerle ilgili oluşum ve dernekleri, test komisyonlarını, test yayıncılarını barındırmaktadır. ITC tarafından geliştirilmiş test yönergelerinin amacı, test kullanımını geliştirmek, test uygulamalarının etkili ve doğru bir şekilde yapılmasını sağlamaktır. Bu amaç doğrultusunda test uygulamaları üzerine, ITC (2001, 2006, 2012, 2014, 2015, 2017, 2018) tarafından günümüze kadar yedi yönerge yayınlamıştır. Bu yönergelerden altısı, uluslararası düzeyde

kabul görmüş projelerin sonuçlarına dayalı olarak oluşturulmuştur. İlgili kuruluş, hali hazırda Test Yayıncıları Birliği (Association of Test Publishers - ATP-) ile ortak olarak “Teknoloji Temelli Değerlendirmeler” üzerine yeni bir yönergenin geliştirilmesi üzerinde çalışmalarına devam etmektedir (ITC, 2020). ITC tarafından geliştirilmiş ve intestcom.org adlı internet sitelerinde erişime açık olarak sundukları yönergeler geliştirildikleri ve revize edildikleri yıllara göre şöyledir:

Test Kullanımına İlişkin Yönergeler (ITC, 2001), Testlerin Çevirisi ve Uyarlanmasına İlişkin Yönergeler (İlk kez, 2005 ile 2015 yılları arasında hazırlanmış, test teknolojisi ve uygulamalardaki gelişmeler doğrultusunda 2017 yılında revize edilerek ikinci baskısı yayınlanmıştır), Bilgisayar Tabanlı ve İnternet Tarafından Sağlanan Testlere İlişkin Yönergeler (ITC, 2006), Puanlamanın Kalite Kontrolü, Test Analizleri ve Test Puanlarının Raporlanmasına İlişkin Yönergeler (ITC, 2012), Testlerin, Sınavların ve Diğer Değerlendirmelerin Güvenliğine İlişkin Yönergeler (ITC, 2014), Uygulayıcıların Testlerin Güncel ve Eski Versiyonlarını Kullanmalarına ve Testlerin İmhasına İlişkin Yönergeler (ITC, 2015), Dilbilimsel ve Kültürel Açından Farklı Örneklerin Geniş Ölçekli Değerlendirilmesine İlişkin Yönergeler (ITC, 2018).

Psikolojik testlere ve ölçme işlemlerine ilişkin çeşitli çalışmalar yaparak Avrupa ülkelerinde kullanılmak üzere model üreten bir diğer uluslararası kuruluş EFPA’dır. EFPA tarafından testlerin nitelikleri ve kullanımının incelenmesini sağlayacak bir Test İnceleme Modeli önerilmiş ve ortaya çıktığı tarihten itibaren güncellenerek geliştirilmiştir. Son güncellemesi 2013’te yapılan EFPA modelinin diğer kuruluşlar tarafından geliştirilen projelere göre daha kapsamlı ve güncel olduğunu ifade etmek mümkündür. Bu model, eğitimde, sağlıkta, psikolojide ve diğer tüm bağlamlarda kullanılan psikolojik testlerin, anketlerin ve ölçeklerin ilgililere ayrıntılı bir şekilde tanıtılmasını amaçlamaktadır.

Standart geliştiren kuruluşların başında gelen Uluslararası Standartlar Teşkilatı (International Organization for Standardization-ISO) da test kullanımına ilişkin standartlar önermiştir. ISO tarafından 2011 yılında önerilen 10667 numaralı standart, iş ortamlarında yapılan bireysel ölçme ve değerlendirmeleri düzenleyen bir çerçeve sunmaktadır. Bu standart kapsamında, ITC ve EFPA tarafından geliştirilmiş standartlar da

bulunmaktadır. 10667 numaralı standart, tüm değerlendirme aşamaları için geniş bir çerçeve ve yönerge sunmaktadır (ISO, 2011).

Testler için oluşturulmuş standartlar incelendiğinde, psikolojik testlerin geliştirilmesi, uygulanması ve sonuçların yorumlanıp karara varılması konusunda birden çok uluslararası kuruluş tarafından önemli çalışmalar yapıldığı ortaya çıkmaktadır. Farklı kuruluşlar tarafından ölçme süreçlerine ilişkin standartlara ek olarak, ölçme süreçleriyle ilişkisi olan tüm taraflar için yol gösterici nitelikte çeşitli yönergelerin de yayınlandığı görülmektedir. Türkiye’de bu alanda yapılan çalışmalar incelendiğindeyse, test standartları ve yönergelere ilişkin az sayıda çalışmanın yapıldığı ve bu çalışmaların çoğunlukla bireysel çabalarla sınırlı kaldığı görülmektedir. Bu alandaki ilk çalışma, APA tarafından 1985 yılında yayınlanan Eğitimde ve Psikolojide Ölçme Standartlarının Türk Psikologlar Derneği öncülüğünde Türkçe’ye çevrilmesidir. Bu standartlar, Hovardaoğlu ve Sezgin (1998) tarafından Türkçe’ye çevrilmiş ve Türk Psikologlar Birliği tarafından kitap olarak basılmıştır. Test standartları konusunda bir kuruluş öncülüğünde yapılan bu ilk çalışmanın ardından, bu konudaki çalışmalar çoğunlukla bireysel araştırmacılar tarafından yürütülmüştür. Mor Dirlik ve Koç (2013) araştırmalarında, 1999’da güncellenen APA ve ITC standartlarını kullanarak bir kontrol listesi oluşturmuş ve eğitim kurumlarında sıklıkla kullanılan psikolojik testlerden dördünü bu standartlara göre incelemiştir. Çüm ve Koç (2013) tarafından yapılan çalışmada, ITC ve APA standartları kullanılmıştır. Mor Dirlik (2014), APA’nın test geliştirme standartlarını kullanarak tezler kapsamında geliştirilen ölçeklerin, geliştirilme süreçlerinin söz konusu standartlara uygunluğunu incelemiştir. İncelenen çalışmalarda eksik noktalar olduğu, test standartlarına belirtilen özellikle geçerlik standartların testler için yeterince incelenemediği sonucuna ulaşılmıştır.

Test standartlarına ilişkin yapılan çalışmalar incelendiğinde, ölçme sürecine dahil olan tüm taraflar için standartların ve yönergelerin oluşturulduğu görülmektedir. Örneğin, testi alanlar için ITC tarafından geliştirilen yönergeler varken, test kullanıcıları için İngiliz Psikologlar Birliği tarafından geliştirilen standartlar bulunmaktadır. Uluslararası standartlara ek olarak, Almanya, Norveç ve Hollanda gibi bazı ülkeler tarafından ulusal düzeyde geliştirilen standartlar da bulunmaktadır (Evers, Sijtsma, Lucassen ve Meijer, 2011). Standartlara ilişkin Türkiye’de yapılan çalışmalar incelendiğinde, daha önce de vurgulandığı gibi yönerge ya da standart

geliştirmeyi, standartların Türkçe'ye çevirisini yaparak kullanımını arttırmayı ve ilgililere tanıtmayı amaçlayan çalışmaların oldukça sınırlı olduğu görülmüştür.

Test kullanıcılarına, araştırmacılara ve yanıtlayıcılara yol gösterici olacak bir modelin, standartların ya da yönergelerin geliştirilmesinin geçerliği ve güvenilirliği yüksek ölçme sonuçları ile tekrarlanabilir ölçme süreçleri sağlayacağı açıktır. Tekrarlanabilir ölçme işlemleri tüm süreç ve sonuçlar açısından nesnelliği arttıracaktır. Böylece, testi yanıtlayan bireylerin de teste ilişkin adalet algıları daha olumlu yönde olacaktır. Dünya genelinde kabul görmüş ölçme standartlarının ülkemizde takip edilmesinin ve bu standartlar temelinde ülkemizin kültürel özelliklerini de gözetilen standartlar geliştirilmesinin, bu alanda atılması gereken önemli adımlar arasında yer aldığı ifade edilebilir. Bu nedenle, yönergelerin ve standartların geliştirilmesi için dünya genelinde yapılmış çalışmaların incelenmesinin, benzerlik ve farklılıkların ortaya koyulmasının bu konuda çalışan araştırmacılara yol gösterici olacağı düşünülmektedir. Bu doğrultuda, çalışmanın amacı, dünya genelinde psikolojik ölçmelerin nitelikli bir biçimde gerçekleştirilmesini sağlamak için çeşitli kuruluşlar tarafından önerilen ölçme standartlarının ve yönergelerinin tanıtılmasıdır. Psikolojik testlere ve ölçmelere ilişkin standartlar konusundaki güncel araştırmaları ve gelişmeleri derleyen bu çalışmanın, test geliştirici, kullanıcı ve her düzeyde ilgili açısından test standartlarına ilişkin farkındalığı arttıracığı düşünülmektedir.

Yöntem

Bu çalışma kapsamında, eğitsel ve psikolojik ölçme alanlarında dünyanın önde gelen kuruluşları tarafından oluşturulmuş test standartlarını ve yönergelerini Türkiye'de ölçme süreciyle ilgilenen tüm taraflara tanıtmak amaçlanmıştır. Bu amaç doğrultusunda, önemli kuruluşlar tarafından hazırlanmış ve dünya genelinde yaygın biçimde kullanılan test standartları ve yönergelerinin neler olduğu belirlenmiştir. Belirlenen test standartları ve yönergeler için hazırlanmış ilgili tüm dokümanlara ulaşılmaya çalışılmıştır. Bu alanda yapılan önemli çalışmaları ve gelişmeleri ortaya koymak amacıyla bu dokümanlar derinlemesine incelenmiştir. Doküman incelemesi türündeki bu çalışma, betimsel yaklaşıma uygun biçimde yürütülmüştür.

Çalışma kapsamında, çeşitli uluslararası kuruluşlar tarafından geliştirilmiş dört farklı test standardına ilişkin detaylı bilgi sunulmuştur. Çalışma grubuna dâhil edilen standartlardan biri AERA, APA ve NCME (2014) tarafından geliştirilmiş test standartlarıdır. İncelemek üzere bu test standartlarının seçilmesinin iki nedeni bulunmaktadır. Biri, bu standartların ülkemizde en fazla tanınan ve en yaygın kullanılan standartlar olmasıdır. 1985 yılında yayınlanan standartların Türkçeye çevrilmesinden sonra bu standartlar iki kez güncellenmiştir. Bu güncellemelerle, test standartlarında çok önemli değişiklikler ve geliştirmeler yapılmıştır. Ancak, Türkiye’de test standartlarındaki güncellemelerin yeterli biçimde izlenmediği düşünülmektedir. Bu nedenle bu standartlar çalışma grubu kapsamına dâhil edilmiştir. AERA, APA ve NCME (2014) standartları test geliştirme, geçerlik, güvenilirlik, klinik ölçmeler vb. gibi daha çeşitli kapsam ve uygulama alanları için geliştirilmiştir. Ancak, ITC’nin çalışmaları incelendiğinde testlerin çeviri süreçlerine ilişkin yönergelerin ağırlıkta olduğu görülmüştür (Brennan, 2006; ITC, 2017). Testlerin uyarlanması sürecinde daha yol gösterici olabileceği düşünülerek incelemede ITC standartlarına da yer verilmiştir. ISO (2011) tarafından önerilen 10667 numaralı standardın tüm değerlendirme aşamaları için geniş bir çerçeve ve yönerge sunduğu düşünülmektedir. Bu nedenle, çalışma kapsamında incelenen standartlar arasında yer almaktadır. Son olarak, EFPA (2013) tarafından oluşturulan Test İnceleme Modeli’nin, diğer kuruluşlar tarafından geliştirilen standartlara ve yönergelere göre daha kapsamlı ve güncel olduğu düşünüldüğünden, bu model de çalışma grubuna dâhil edilmiştir.

İşlem

Çalışma kapsamında dört farklı kuruluşun test standardına ilişkin inceleme yapılmıştır. Bu standartlardan ilki, AERA, NCME ve APA (2014) tarafından geliştirilen standartlardır. Test standartlarının incelenmesi sırasında özellikle Türkçe’ye çevirisi yapılan 1985 yılı standartlarından sonra gerçekleşen güncellemelere odaklanılmıştır. Test standartlarında yapılan güncellemelere ilişkin detaylı bilgi içeren AERA, NCME ve APA (1999, 2014) kaynaklarına ulaşılarak, bu kaynakların özellikle daha önceki yıllarda yayınlanan standartlara göre değişiklik içeren bölümlerinin neler olduğu

ortaya koyulmuştur. Ayrıca, içeriğine ilişkin bilgiler elde etmek amacıyla 2014 yılında yayınlanan güncel standartlar daha detaylı incelenmiştir. Benzer biçimde, EFPA ve ISO standartlarına ilişkin en detaylı bilgiler içeren kaynaklara ulaşıldıktan sonra, bu kaynaklara dayalı olarak ilgili standartların ve yönergelerin temel niteliklerini ve içeriklerini ortaya koymak amacıyla bir inceleme yapılmıştır. ITC standartlarına ilişkin bilgiler incelendiğinde, özellikle test uyarlama sürecinde araştırmacılar için yol gösterici nitelikte yönergeler içerdiği görülmüştür. Bu nedenle, bu yönergelerin Türkçe'ye çevrilmesine karar verilmiştir. ITC tarafından test uyarlamaya ilişkin geliştirilen yönergeler yazarlar tarafından ayrı ayrı Türkçe'ye çevrilmiştir. Daha sonra, ortaya çıkan üç çeviri üzerinde birlikte çalışılarak ortak bir çeviri formu oluşturulmuştur. Bu form, hem Türkiye hem de Amerika Birleşik Devletleri'nde yaşamış, her iki dile ve kültüre ilişkin bilgi ve deneyimleri olan, ayrıca test standartları konusunda da bilgi sahibi olan bir dil uzmanı tarafından da incelenmiştir. Dil uzmanından gelen geri bildirimlere dayalı olarak yazarlar tarafından çevirinin son formu oluşturulmuştur.

Bulgular

Çalışma kapsamında, AERA, NCME ve APA (2014), EFPA (2013) ITC (2017) ve ISO (2011, 2020) tarafından yayınlanan test standartları ve yönergeleri incelenmiştir. Bu incelemeler başlıklar altında aşağıda sunulmuştur

AERA, NCME ve APA: Eğitimde ve Psikolojide Test Standartları

AERA, NCME ve APA tarafından önerilen Eğitimde ve Psikolojide Test Standartları 1966, 1974, 1985, 1999 ve 2014 olmak üzere beş kez güncellenmiştir. Daha önce ifade edildiği gibi, 1985 yılında yayınlanan standartlar Hovardaoğlu ve Sezgin (1998) tarafından Türkçe'ye çevrilmiştir. Bu nedenle bu çalışmada, 1985 yılından sonra standartlarda yapılan değişikliklere ve güncel standartların içeriğine odaklanılmıştır. 1985 standartlarında, her standart kullanımdan önce tüm testler tarafından karşılanması gereken standartlar, karşılanması istenen ancak her durum için uygun olmayan standartlar ve önemi uygulama koşullarına bağlı olan standartlar biçiminde tanımlanmıştır. Böylece standartlar arasında birincil,

ikincil ve koşullu standartlar olarak adlandırılan bir ayrıma gidilmiştir. AERA, NCME ve APA (1999) tarafından yapılan güncellemedeyse, tüm standartların ilgili oldukları bağlamlarda önemli olduğu vurgulanarak standartların önem düzeyi gösterir biçimde sınıflanmasının doğru olmadığı ifade edilmektedir. Bu tür sınıflamaların, belirli durumlarda ele alınması gereken standartların göz ardı edilmesine neden olabileceği öne sürülmektedir. Dolayısıyla, bu güncellemede 1985 standartlarında kullanıldığı gibi birincil, ikincil ya da koşullu gibi başlıklar kullanılmamaktadır. Bunun yerine, belirli bir kullanım durumuyla ilgisiz, o durum için önemsiz ya da uygulanabilir olmayan tüm standartların karşılanması gerektiği ve ilgili durum için tüm standartların birincil standartlar olduğu vurgulanmaktadır. Ayrıca, bu güncelleme kapsamında standartların sayısı da arttırılmıştır. Bunun ilk ve en önemli nedeni, yeni test türlerinin ya da var olan testlerin yeni kullanımları sonucunda yaşanan gelişmelerdir. Yine bu güncellemede tüm yanıtlayıcılara eşit bir yaklaşım sağlamak amacıyla bu konularla ilgili standartlar da eklenmiştir.

Eğitimde ve Psikolojide Test Standartları (1999) gözden geçirilerek AERA, NCME ve APA tarafından 2014 yılında yeniden güncellenmiştir. Bu güncellenmenin temelde eğitim politikalarında ve test kullanımında hesap verebilirlikle ilgili konuların göz önüne alınması, testlerin erişilebilirliği kavramının genişletilmesi, işyerinde testlerin rolünün daha kapsamlı biçimde ele alınması, teknolojinin test geliştirme ve uygulama sürecindeki rolünün genişletilmesi, standartların daha iyi biçimde iletilebilmesi için daha iyi bir yapının oluşturulması gibi amaçlarla gerçekleştirildiği ifade edilmektedir. Bu gereksinimlere karşılık verebilmek amacıyla, bu güncellemede bazı önemli adımlar atılmıştır. Bu adımlardan biri, 1999 güncellemesinde “Eğitimde Testler ve Durum Belirleme” ve “Program Değerlendirme ve Kamu Politikalarında Testlerin Kullanımı” başlıklı bölümlerin testlerin eğitimde hesap verebilirlik için kullanımıyla ilgili konulara yer vermek amacıyla yeniden yazılmasıdır. Ayrıca, “İşyerinde Ölçme Değerlendirme ve Yetkilendirme” bölümü bir standardın ne zaman istihdam ya da yetkilendirmeye ilgili olduğunun daha açık tanımlanabilmesi amacıyla yeniden düzenlenmiştir. Bu güncellemede, teknolojinin etkisi tüm bölümlerde göz önünde bulundurulmuştur. Teknolojiyle ilgili olarak açık uçlu madde yanıtlarının otomatik puanlanması, yenilikçi madde türlerinin uygulanması ve puanlanması,

bilgisayar tabanlı test uygulamaları gibi konulara ilişkin standartlar “Test Planlama ve Geliştirme” bölümlerinde ele alınmıştır.

Eğitimde ve Psikolojide Test Standartları'nın 2014 yılındaki güncellemeninde yapılan en önemli gelişmelerden biri “yanlılığın” kavramlaştırılması olmuştur. 1999 standartlarında, bu konuya “Test Kullanımı ve Uygulamalarında Yansızlık”, “Farklı Dil Geçmiş Olan Bireylerin Ölçülmesi”, “Engelli Bireylerin Ölçülmesi” başlıklı ayrı bölümlerde yer verilmektedir. 2014 standartlarında belirtilen bölümlerde ele alınan konular daha kapsamlı tek bir bölüm altında toplanarak, erişilebilirliğin ve yanlılığın, test geliştirme ve uygulama alanındaki temel konulardan biri olduğunu vurgulamak amacıyla “Ölçmede Yansızlık” başlıklı yeni bir bölüm yazılmıştır. Bu değişiklik, tüm test yanıtlayıcılarına aynı biçimde davranmanın gerekliliğini ve önemini vurgulamak amacıyla yapılmıştır. İlgili bölümde, çoğunlukla farklı dil ya da kültür geçmişi ve engeli olan bireylerle ilgili örnekler bulunmaktadır. Ancak, bu bölümde tüm yanıtlayıcılar için tarafsız ve eşit bir uygulama önündeki olası engelleri yansıtılabilmek amacıyla, cinsiyetle, yaş gruplarıyla, farklı etnik ve ırksal kökene sahip bireylerle ilgili örneklerle de yer verilmektedir. Eğitimde ve Psikolojide Test Standartlarının içeriği incelendiğinde, üç genel başlık altında 13 ayrı bölüm olarak düzenlendiği görülmektedir. AERA, NCME ve APA (2014) standartlarının içeriğini daha detaylı biçimde ortaya koymak amacıyla üç genel başlık altında yer alan bölümlere ve alt bölümlere ilişkin bilgiler Tablo 1’de verilmiştir.

Tablo 1. AERA, NCME ve APA (2014) Test Standartlarının İçeriği

Temel Bilgiler	
1. Geçerlik	Testin amacı ve teste dayalı yapılacak yorumlar, geçerlik çalışmasında yapılan uygulamalar ve örnekleme, geçerlik kanıt türleriyle ilgili üç küme altında toplanmış 25 standart bulunmaktadır.
2. Güvenirlik	Testin farklı oturum ve koşullarda uygulanması durumunda güvenilirlik kanutlarının sunulması, ölçme kesinliği, genellenebilirlik katsayıları, ölçme kesinliğini etkileyen faktörler, ölçmenin standart hatası, kararlılık sürekliliği ve kesinliği, grup ortalamalarının güvenilirliği, güvenirlüğün belgelenmesiyle ilgili sekiz küme altında toplanmış 20 standart bulunmaktadır.
3. Ölçmede Yansızlık	Test planlama, geliştirme, uygulama ve puanlamada yanlılığı azaltma, hedef grupta belirli kullanım amacı için yapılan test puanı yorumlarının geçerliği, ölçülmek istenen yapıyla ilişkisiz değişkenliğe engel olmak için izlenebilecek yollar, testin kullanım amacıyla uyumsuz puan yorumlarına karşı önlemler ile ilgili dört küme altında toplanmış 20 standart bulunmaktadır.

İşlemler	
4. Test Planlama ve Geliştirme	Test planlaması ve belirlemeleri, madde yazımı ve redaksiyonu, test geliştirme, uygulama, puanlama süreçleri ve test materyalleri, test revizyonu ile ilgili dört küme altında toplanmış 25 standart bulunmaktadır.
5. Puanlar, Ölçekler, Normlar, Eşitleme ve Kesme Puanları	Puanların yorumlanması, norm puanları, test eşitleme ve kesme puanları ile ilgili dört kümede toplanmış 23 standart bulunmaktadır.
6. Test Uygulama, Puanlama, Raporlama ve Yorumlama	Testin uygulanması, puanlanması, raporlanması ve yorumlanması ile ilgili üç kümede toplanmış 16 standart bulunmaktadır.
7. Testler için Destekleyici Belgeler	Test belgelerinin içeriğine ilişkin, testin uygun kullanımı, geliştirilmesi, yönetimi ve puanlanması ile teste ilişkin belgelerin zamanında teslim edilmesini içeren dört kümede toplanmış 14 standart bulunmaktadır.
8. Test Yanıtlayıcılarının Hak ve Sorumlulukları	Test yanıtlayıcılarının testi almadan önceki bilgi edinme hakları, test sonuçlarına erişme ve test sonuçlarının izinsiz kullanımına karşı korunma hakları, test puanlarının adil ve doğru raporlanması ile test yönetim süreci boyunca testi yanıtlayanların davranışlarına ilişkin sorumluluklarını içeren dört kümede toplanmış 12 standart bulunmaktadır.
9. Test Kullanıcılarının Hak ve Sorumlulukları	Yorumların geçerliliği, bilginin yayılması, test güvenliği ve telif haklarının korunması ile ilgili üç kümede toplanmış 23 standart bulunmaktadır.
Test Uygulamaları	
10. Psikolojik Testler ve Durum Belirleme	Test kullanıcısının nitelikleri, test seçimi, test yönetimi, testin yorumlanması ve test güvenliği ile ilgili beş kümeden oluşan 18 standart bulunmaktadır.
11. İşyeri Testlerinin Geliştirilmesi ve Belgelendirme	Hem istihdam testlerinin geliştirilmesi hem de belgelendirme için genel olarak geçerli standartlar ile yalnızca istihdam testlerinin geliştirilmesi ve yalnızca belgelendirme için standartların üç kümede toplanmasıyla oluşan 16 standart bulunmaktadır.
12. Eğitimsel Testler ve Durum Belirleme	Eğitsel ölçmelerin tasarımı ve geliştirilmesi, eğitsel testlerin kullanımı ve yorumlanması ile eğitsel değerlendirmelerin yönetimi, puanlanması ve raporlanmasıyla ilgili üç kümeden oluşan 19 standart bulunmaktadır.
13. Testlerin Program Değerlendirme, Politika Çalışmaları ve Hesap Verebilirlik için Kullanımı	Program değerlendirme, politika çalışmaları ve hesap verebilirlik sistemleri için test programlarının ve endekslerinin tasarımı ve geliştirilmesi ile yine bu sistemlerde kullanılan testlerden elde edilen bilgilerin yorumlanması ve kullanımıyla ilgili iki kümeden oluşan dokuz standart bulunmaktadır.

EFPA: Test İnceleme Modeli

EFPA tarafından geliştirilen Test İnceleme Modeli'nin temel amacı, psikolojide, eğitimde, sağlıkta ve diğer alanlarda kullanılan tüm psikolojik testleri, ölçekleri, anketleri, hatta projektif teknikleri tanımlamak ve bu ölçme araçlarını ayrıntılı ve titiz bir biçimde değerlendirmeyi sağlamaktır. Test İnceleme Modeli kullanılarak ölçme araçlarına ilişkin toplanan bilgiler,

EFPA'nın bilgilendirme stratejisine dahil edilmekte ve tüm ilgililerle paylaşılmaktadır (Evers vd., 2011; Muñiz ve Bartram, 2007). Test İnceleme Modeli, farklı Avrupa ülkeleri tarafından da benimsenmiştir. Norveç, İspanya, İsveç, İngiltere, Çek Cumhuriyeti, Macaristan ve Litvanya gibi ülkelerde modelin çevirisi yapılarak kullanıma sunulmuştur.

Test İnceleme Modeli, İngiliz Psikologlar Birliği tarafından geliştirilen Test İnceleme ve Değerlendirme Formu ile Test Niteliğinin Değerlendirilmesinde Hollanda Modeli gibi birçok test inceleme modeli temel alınarak oluşturulmuştur. Son olarak 2013 yılında güncellenen modelin içeriği incelendiğinde, modelin üç temel kısımdan oluştuğu görülmektedir. İlk bölümde ölçme aracına ilişkin tanımlayıcı bilgiler yer almakta ve incelenen test tüm ayrıntıları ile sunulmaktadır. İkinci kısımdaysa, materyaller, normlar, güvenilirlik, geçerlik ve bilgisayarda üretilen raporlar gibi temel nitelikler açısından test incelemeye alınmaktadır. Son kısımda ise değerlendirme yapılırken kullanılan kaynakça verilmektedir. Modelin içeriği kadar önemli olan bir diğer durum, modelin uygulamaya konulmasının gerekliliğidir. İncelemeye alınan test, iki bağımsız değerlendirici tarafından tıpkı bir bilimsel proje ya da makale değerlendirme süreçlerinde olduğu gibi modele dayalı olarak değerlendirilmektedir. İki değerlendirici-hakemden gelen raporlar danışman editör tarafından incelenmektedir. Değerlendirmeler arasında önemli farklılıkların olması durumunda, test gerekirse üçüncü bir değerlendiriciye gönderilmektedir. EFPA, hakemlerin nitelikli test kullanıcıları arasından seçilmesini önermekle beraber akademisyenlerin, test yazarlarının ve psikometri ya da psikolojik ölçmelerdeki uzmanların da hakem olarak değerlendirmeye alınabileceğini belirtmektedir. Testlerin değerlendirilmesinde bir diğer önemli nokta, test değerlendirme sonuçlarının tüm araştırmacılara ve test kullanıcılarına açık olmasıdır. Testler, kitap ya da dergi olarak basılabileceği gibi Ulusal Psikoloji Birliği tarafından internet üzerinden de yayınlanabilmektedir (EFPA, 2013).

Modelin yaygınlaştırılmasındaki temel amaç, test inceleme sürecini ve kriterlerini Avrupa'da genelleştirmektir. Buna ek olarak, kendi test inceleme modeline sahip olmayan ülkelerde de bu modelin kullanımının sağlanması amaçlanmaktadır. Modelin farklı ülkelerde kullanılması durumunda çeşitli revize işlemlerinin yapılması gerekeceği ifade edilmektedir. Bu amaçla, Test İnceleme Modeli üzerinde yerel düzeyde yapılacak değişikliklerin ilgili

kültüre daha uyumlu bir inceleme çerçevesi sunacağı belirtilmektedir. Bu nedenle, test inceleme süreci için oluşturulan bu çerçeve EFPA tarafından “model” olarak isimlendirilmektedir (EFPA, 2013).

Test İnceleme Modeli'nin içeriği incelendiğinde, modelin üç temel bölümden oluştuğu görülmektedir. İlk bölümde, herhangi bir yargısal sürece başvurulmadan, yalnızca testin temel özellikleri ortaya koyulmaktadır. Teste ilişkin betimsel bilgilerin yer aldığı “Ölçme aracının tanımı” olarak adlandırılan bölümde genel bilgiler, sınıflama, ölçme ve puanlama, bilgisayarda raporlama, destek ve maliyet başlıkları yer almaktadır. Genel bilgiler başlığı altında, ölçme aracının adı, yayıncısı/dağıtıcısı, yayın tarihi, form bilgileri ve revize edilme durumu ve araştırmacıların bilgileri verilmektedir. Sınıflama başlığı altında, testin içeriği, önerilen kullanım alanları, yanıtlanma yolu, testi alanlardan beklenenler, testte kullanılan madde türü, testin yanıtlanma süresi, testin farklı formları varsa bu formlara ilişkin bilgiler verilmektedir. Ölçme ve puanlama başlığı altında puanlama sürecinde izlenen ilkelere ve norm puan dönüşümlerine ilişkin bilgiler verilmektedir. Bilgisayarda raporlama başlığı altında, oluşturulan raporların niteliğine, medya kullanımına ilişkin izin durumuna ve testin farklı koşullara uyarlanabilirliğine ilişkin bilgiler yer almaktadır. Destek ve maliyet başlığında ise testin yayıncılarına, maliyetine ve hedef grubuna ilişkin bilgiler verilmektedir.

Test İnceleme Modeli'nin ikinci bölümünde, teste ilişkin değerlendirmeler yapılmaktadır. Dolayısıyla, model kapsamında değerlendirmede temel alınacak ölçütler bu bölümde yer almaktadır. Bu bölümde yer alan ölçütlere dayalı olarak test, 0 ile 4 puan arasında değişen bir dereceleme temelinde değerlendirilmektedir. Bu değerlendirmede verilen 0 puan söz konusu ölçütün incelenen test için geçerli olmadığı anlamına gelmektedir. 1 puan testin ilgili ölçütü karşılama açısından yetersiz olduğunu, 2 puan kısmen yeterli olduğunu, 3 puan test tarafından ölçütün iyi düzeyde karşılandığını, 4 puan ise mükemmel düzeyde karşılandığını göstermektedir. EFPA (2013) tarafından, test için 0 ve 1 olarak puanlanan ölçütler test için kritik düzeyde önemli olduğunda, kullanıcıların bu testi tercih etmemesi ya da testin yalnızca alan uzmanları tarafından dikkatli bir biçimde ve çok gereklilik arz eden durumlarda kullanılması önerilmektedir. Testin değerlendirilmesi sürecinde hakemler tarafından, makaleler, el kitapları, raporlar, açık kaynaklar, yayıncılar tarafından

oluşturulmuş fakat resmi olarak basılmamış dokümanlar ve yayıncılar tarafından ifşa edilmemiş anlaşmalar gibi testle ilgili tüm kaynaklar kullanılır. Bu kaynaklara dayalı olarak, ilgili testin altı temel alandaki nitelikleri hakemler tarafından değerlendirilmektedir. Testin değerlendirildiği temel alanlar ve alt başlıklara ilişkin daha detaylı bilgi Tablo 2’de verilmiştir. Tablo 2’de Test İnceleme Modeli kapsamında psikolojik testlerin değerlendirilmesinde kullanılan temel alanlar ve bu alanların altında yer alan göstergeler genel olarak verilmiştir. Test İnceleme Modeli, oldukça fazla sayıda gösterge ve alt alan içermektedir. Bu nedenle, Tablo 2’de verilen bilgiler modelin birincil ve ikincil düzeydeki başlıklarını içermektedir.

Tablo 2. Test İnceleme Modeli Değerlendirme Bölümünün İçeriği

Testin Kuramsal Temellerinin Açıklanması
<ul style="list-style-type: none">• Ölçülen yapı/yapıların kuramsal temelleri• Test geliştirme/çevirme/uyarlama süreci• Madde analizlerinin ve madde analiz modelinin detaylandırılması• İçerik analizinin sunumu• İlgili çalışmaların incelenmesi ve sunulması
Test Materyallerinin Niteliği
<ul style="list-style-type: none">• Kalem- kâğıt, bilgisayar ve internet tabanlı testler için kullanılan materyallerin niteliği• Yönergelerin niteliği• Test materyallerinin (kitapçıklar, yanıt kağıtları, testte kullanılan nesnelere gibi) niteliği
Test Normları
<ul style="list-style-type: none">• Norm dayanaklı testlere ilişkin ölçütlerin uygunluğu• Ölçüt dayanaklı testlere ilişkin ölçütlerin uygunluğu• Örneklemin ölçülen özelliğin doğasına, evrenin niteliğine göre belirlenmesi• Kesme puanlarının belirlenme süreci• Sürekli norm dağılımı sürecine ilişkin belirlenme
Güvenirlilik
<ul style="list-style-type: none">• Güvenirliğe ilişkin kanıt toplama yöntemleri• İç tutarlılık anlamında güvenirlilik• Test tekrar test güvenirliliği• Madde tepki kuramı yaklaşımı• Puanlayıcılar arası güvenirlilik• Güvenirlilik katsayısının büyüklüğü• Örnekleme büyüklüğünün yeterliliği• Örneklemin heterojenlik düzeyi• Testin kullanım amacına uygun güvenirlilik inceleme yönteminin seçilmiş olması• Birden fazla güvenirlilik incelemesinin yapılmış olması• Yapılan işlemlerin ayrıntılı raporlanmış olması
Geçerlik
<ul style="list-style-type: none">• Birleştirilmiş geçerlik tanımına göre oluşturulan geçerlik kanıtlama yöntemlerinden, hangi/hangilerinin uygulandığının belirlenmesi

-
- Yapı geçerliği inceleme yöntemleri
 - Ölçüt geçerliğini inceleyen yöntemler
 - Testle verilecek kararlara uygun geçerlik kanıtlama yöntemlerinin seçimi
 - Test için uygun olan geçerlik kanıtlama yöntemlerine dayalı olarak ilgili geçerlik kanıtlarının toplanması
 - Test puanlarının geçerliğinin düşük ve yüksek olduğu grupların tanımlanmış olması
 - Test puanlarının geçerliğine ilişkin çalışma yapılmamış grupların tanımlanmış olması
 - Geçerliği incelenen test puanlarının elde edildiği grubun yaşına ilişkin betimsel bilgilerin verilmesi
 - Ölçüt dayanaklı geçerlik kanıtları için, özellikle klinik durumlarda ROC (Receiver Operating Characteristics- Alıcı İşletim Karakteristiği) eğrisinin sunulması
-

Bilgisayarda Üretilen Raporların Niteliği

- Raporların ayrıntı içerme durumu
 - Raporların geçerli olduğu test kullanıcı grupları
 - Kişiye ya da gruba özel raporların sunulma durumu
 - Raporların yanıtların tutarlılığını/tutarsızlığını belirtme durumu
 - Raporların, yanıtlara ilişkin yanlılık-merkeze yığılma gibi istatistikleri içerme durumu
 - Raporların, bireysel puanların yorumlanacağı güven aralıklarını içerme durumu
-

Nihai Değerlendirme

- Testin bütününe ilişkin açık, net ve kısa bir yargıda bulunulması
 - Testin üstün ve eksik yönlerinin ifade edilmesi
 - Test kullanıcılarının özelliklerine ilişkin bilgi verilmesi
 - Testin nasıl ve ne zaman kullanılacağına ilişkin öneriler sunulması
-

ITC: Test Uyarlama Yönergeleri

Psikolojik testlerin farklı kültürlerle uyarlanması konusunda önemli ve kapsamlı çalışmalar yürüten kuruluşların başında gelen ITC tarafından test uyarlama sürecine ilişkin yönergeler geliştirilmiştir. Bu yönergelerin, test uyarlamayı düşünen ve bu alanla ilgilenen tüm araştırmacılara izlenmesi gereken aşamaları adım adım anlatan ayrıntılı bir rehber olduğu ifade edilebilir. İlk kez 2005'te geliştirilen yönergeler, uzun ve ayrıntılı çalışmalar sonucunda güncellenmiş ve 2017'de daha güncel ve detaylı bir şekilde yenilenen yönergeler yayınlanmıştır. ITC'nin (2017) yönergeleri incelendiğinde, test uyarlama sürecine ilişkin altı bölüm içerisinde on sekiz farklı yönerge geliştirildiği görülmektedir. Test uyarlama yönergelerinden üçü, uyarlama işlemine başlamadan önce gözetilmesi ve karşılanması gereken koşullara ilişkin bilgiler içermektedir. Bu yönergelerde, ölçülen yapının uyarlanmak istenen kültürle ve hedef kitleyle örtüşme düzeyinin incelenmesi gerektiği vurgulanmaktadır. Ayrıca, testin hedef dile çeviri sürecinde göz önünde bulundurulması gereken durumlara da yer verilmektedir. Özgün ve uyarlanan formların psikometrik nitelikleri açısından eşitliğinin incelenmesine ilişkin gerekliler sıralanmaktadır.

Uygulama yönergeleri kapsamında, testin uygulanması sırasında, hedef dil ve kültür ögelerinin dikkate alınarak gerekli yönergelerin hazırlanması ve tüm uygulamalarda uyulması gereken test koşullarının belirlenmesi gerekliliği ifade edilmektedir. Puanlama aşamasındaysa, eşitliğin sağlandığı durumlarda puan farklılıklarının yorumlanması konusuna ilişkin uyarılara yer verilmektedir. Belgelendirme işlemlerine ilişkin yönergelerde, elde edilen kanıtların tümünü kapsayan ve hedef kültürde kullanımını destekleyen belgelerin sağlanması gerektiği ifade edilmektedir. İçeriğine ilişkin bilgi verilen ITC test uyarlama yönergeleri Tablo 3'te sunulmuştur.

Tablo 3. Test Uyarlama Yönergeleri

Önkoşul Yönergeler
1. Herhangi bir uyarlama çalışmasına başlamadan önce, testin fikri mülkiyet haklarına sahip olan kişi/kişilerden gerekli izinler alınmalıdır.
2. Testle ölçülen psikolojik yapının tanımının ve içeriğinin, hedef kitlelerde test puanlarının amaçlanan kullanımları için yeterli düzeyde örtüşme sağladığı belirlenmelidir.
3. Testin ölçtüğü yapının tanımının ve içeriğinin test puanlarının amaçlanan kullanımları için hedef kitleyle sağladığı örtüşme düzeyi belirlenmelidir. Hedef kitle için testin amaçlanan kullanımlarıyla ilgili olmayan herhangi bir kültürel ve dilsel farklılığın etkisi en aza indirilmelidir.
Test Geliştirme Yönergeleri
4. Uyarlama sürecine, hedef kitlenin dilsel, psikolojik ve kültürel farklılıkları konusunda deneyim ve bilgi sahibi uzmanlar dâhil edilmelidir.
5. Test uyarlama sürecinin hedef kitleye uygunluğunu en üst düzeye çıkarmak için uygun çeviri deseni ve yöntemleri kullanılmalıdır.
6. Test yönergelerinin ve madde içeriklerinin tüm hedef kitleler için benzer bir anlama sahip olduğuna dair kanıtlar sağlanmalıdır.
7. Madde tiplerinin, dereceleme ölçeklerinin, puanlama kategorilerinin, test düzenlemelerinin, uygulama yollarının ve diğer süreçlerin tüm hedef kitlelere uygun olduğuna ilişkin kanıt sağlanmalıdır.
8. Madde analizlerini, güvenilirlik incelemelerini ve diğer küçük ölçekli geçerlik çalışmalarını yapabilmek için uyarlanmış teste ilişkin veri toplanarak uyarlanmış test üzerinde gerekli düzenlemeler yapılabilir.
Doğrulama Yönergeleri
9. Testin amaçlanan kullanımıyla ilgili deneysel çözümlenmeler yapılmasına olanak sağlayacak nitelikte ve büyüklükte örneklem seçilmelidir.
10. Tüm hedef kitleler için yapısal, yöntemsel eşdeğerlik ve madde eşdeğerliğine ilişkin istatistiksel kanıtlar sağlanmalıdır.
11. Hedefle kitle için uyarlanan testin normlarını, güvenilirliğini ve geçerliğini destekleyen kanıtlar sağlanmalıdır.
12. Testin farklı dil formlarından elde edilen puan ölçeklerini ilişkilendirirken uygun eşitleme deseni ve veri çözümleme süreçleri kullanılmalıdır.
Uygulama Yönergeleri
13. Test puanlarından elde edilen çıkarımların geçerliğini etkileyebilecek uygulama süreçleri ve yanıtlama biçimleri nedeniyle oluşan kültür ve dille ilgili sorunları en aza indirecek uygulama materyalleri ve yönergeleri hazırlanmalıdır.
14. Tüm hedef kitlelerde yakından izlenmesi gereken test koşulları belirtilmelidir.

Puanlama ölçekleri ve Test Puanlarını Yorumlama Yönergeleri

15. Grup puanı farklılıkları var olan tüm ilgili bilgiler ölçüt alınarak yorumlanmalıdır.
16. Puanların rapor edildiği ölçekte değişmezlik düzeyi sağlandığında kitleler arasındaki puanlar karşılaştırılmalıdır.

Belgelendirme Yönergeleri

17. Test farklı bir hedef kitlede kullanılmak üzere uyarlandığında, yapılan tüm değişiklikler denkliliği desteklemek için elde edilen kanıtların dökümünü de içerecek biçimde belgelendirilmelidir.
 18. Test kullanıcılarına, uyarlanmış testin hedef kitlede kullanımının iyi uygulamalarını destekleyecek nitelikte belgeler sağlanmalıdır.
-

ISO: Test Standardı

Farklı alanlara yönelik standartlar öneren kurumların başında gelen Uluslararası Standartlar Teşkilatının test kullanımına ilişkin geliştirdiği standartlar da bu çalışma kapsamında incelemeye alınan standartlar arasında yer almaktadır. ISO (2011) tarafından önerilen 10667 numaralı standart, 2020 yılında güncellenmiştir. Bu standart, işyerlerinde gerek müşteriye gerekse personele ilişkin bilgi toplamak ve işle ilgili kararlar vermek için kullanılan testler ve ölçme süreçlerini kapsamaktadır. Standart kapsamında, testi alan müşterilerin şu alanlardaki gereksinimleri dile getirilmektedir (ISO, 2020):

- Ölçme işlemine neden ihtiyaç duyulduğu ve ölçmenin mantığı açıklanmalı,
- Ölçmenin hangi koşullar altında yapıldığı belirtilmeli.
- Değerlendirme yöntemleri, testin uygulanması ve ölçme yaklaşımına ilişkin alınan kararlar belirlenmeli.
- Ölçme sürecinde görev alacak personelin sahip olması gereken yeterlikler ve uzmanlık alanları tanımlanmalı.
- Ölçme sonuçlarının ulaşılabilirliğine, kullanımına, saklanmasına ve sonradan üretilecek raporlara ilişkin alınan kararlar belirtilmeli.
- Ölçme işlemlerinin nasıl yürütüleceğine ilişkin örgütsel düzenlemeler yapılmalı.

Bu standart kapsamında genellikle iş/mesleğe yönelik ölçmelerin temel alındığı ifade edilmektedir. Bu nedenle standart kapsamındaki ölçme işlemlerinin ve yaklaşımlarının bu bağlam için geçerli olduğu belirtilmektedir. Bu durumlara ek olarak, istihdamla (örneğin, işe alma, seçme, geliştirme, terfi, işe yerleştirme, işten çıkarma ve yeniden görevlendirme gibi), kariyerle (örneğin, mesleki rehberlik ve rehabilitasyon,

işten çıkarma danışmanlığı), grubun bütünüyle (örneğin, eğitim girişimleri, ekip oluşturma) ve örgütle (örneğin, yeniden yapılandırma, ahlaki ve kültürel girişimler) ilgili kararlarda bu standart çerçevesinden yararlanılabilmektedir.

Sonuç ve Öneriler

Psikometrik nitelikleri açısından yeterli bir test geliştirme, test sürecinde yer alan tüm tarafların (testi geliştirenler, testin basımını ve satışını yapanlar, testi uygulayan ve puanlayanlar, test sonuçlarını yorumlayanlar, karar vermek için test sonuçlarını kullananlar, testi alanlar, test sponsorları ve testi gözden geçiren ve eleştirenler) test geliştirmeye ilişkin bilgi, beceri ve yeterliklere sahip olmasını gerektirmektedir (AERA, NCME ve APA, 2014). Test standartları, istenen nitelikte bir testin elde edilebilmesi için yürütülmesi gereken sürecin tüm taraflar için açık bir biçimde tanımlanmasını sağlamaktadır. Bu amaçla dünyanın önde gelen kuruluşları tarafından farklı bağlamlar için çeşitli yönerge ve standartlar geliştirilmektedir. Bu çalışma kapsamında, dünya genelinde yaygın kabul gören ve kullanılan dört farklı test standardının temel niteliklerinin ve içeriklerinin incelenerek tanıtılması amaçlanmıştır.

Çalışmada incelenen standartlardan biri AERA, NCME ve APA (2014) tarafından geliştirilmiş Eğitimde ve Psikolojide Test Standartlarıdır. Bu standartların içeriği incelendiğinde, temel bilgiler, işlemler ve test uygulamaları başlıklarına odaklanıldığı görülmüştür. Temel bilgiler içerisinde geçerlik, güvenirlik ve yanlılıkla ilgili standartlara yer verilmektedir. İşlemler başlığı altında, test planlama ve geliştirme, puanlar, ölçekler, normlar, eşitleme ve kesme puanları, test uygulama, puanlama, raporlama ve yorumlama, testler için destekleyici dokümanlar, test alanların hak ve sorumlulukları, test kullanıcılarının hak ve sorumluluklarıyla ilgili standartlara yer verilmektedir. Test uygulamaları başlığı altında, psikolojik testler ve değerlendirme, iş yaşamında kullanılan testler ve yetkilendirme, eğitsel testler ve değerlendirme, program değerlendirme, politika çalışmaları ve hesap verebilirlik için testlerin kullanımıyla ilgili standartlara yer verilmektedir. Ayrıca, temel bilgiler ve işlemler başlıkları altındaki tüm bölümler, tüm testlere ve test kullanımlarına uygulanabilecek ilkeleri özetleyen bir kapsayıcı standartla tanıtılmaktadır.

Çalışma kapsamında incelenen bir diğer test standardı EFPA tarafından geliştirilen Test İnceleme Modeli'dir. EFPA, Test İnceleme Modeli kapsamında psikolojik testlerin değerlendirilmesinde kullanılan temel alanlar ve bu alanların altında yer alan göstergelere yer vermektedir. Model kapsamında ele alınan göstergeler incelendiğinde, yapılan tüm işlemlerin ayrıntılı olarak betimlenmesi, veri toplamak için kullanılan grupların yaş düzeylerine kadar detaylı bir biçimde sunulması gerektiği ortaya çıkmaktadır. Bu modelde, en fazla göstergenin geçerlik ve güvenilirlik bölümlerinde yer aldığı görülmüştür. Bu niteliklerin testlerin temel psikometrik nitelikleri olduğu göz önünde bulundurulduğunda, bunun beklenen bir durum olduğu düşünülmektedir. Modelde, geçerliğe ve güvenilirliğe ilişkin kanıtların toplandığı durum ve gruplara ilişkin bilgi verilmesi gerektiği ifade edilmektedir. Göstergeler incelendiğinde, test geliştirme sürecinde üzerinde çalışılmayan ya da test puanlarının geçerlenmesinin uygun olmadığı durum ve grupların belirtilmesine de önem verildiği ortaya çıkmaktadır. Ayrıca, modelde yer alan göstergelerde, testin kullanım amaçlarıyla uyumlu geçerlik ve güvenilirlik kanıtlarının sunulması üzerinde de durulmaktadır.

Test uyarlama sürecine odaklanması nedeniyle çalışma kapsamında incelenmesine karar verilen bir diğer standart ITC (2017) tarafından geliştirilmiş yönergelerdir. Yönergeler genel olarak incelendiğinde, özellikle eşdeğerlik kavramına vurgu yapıldığı ortaya çıkmaktadır. Kültürel ve dilsel denkliğin sağlanması bir önkoşul olarak kabul edilmektedir. Teste ilişkin yapılan tüm değişikliklerin kullanıcılara ve ilgilere ayrıntılı bir şekilde sunulmasına vurgu yapılmaktadır. Bunun yanında, testlerin geçerlik ve güvenilirlik gibi psikometrik özelliklerine ek olarak, testlerden elde edilen puanların anlamları ve normları gibi konuların da ele alınması önerilmektedir.

Çalışma kapsamında ele alınan bir diğer standart ISO (2011) tarafından önerilen 10667 numaralı standarttır. Bu standart kapsamında, işe/mesleğe yönelik ölçmeler temel alınmaktadır. İstihdamla, kariyerle, grubun bütünüyle ve örgütle ilgili kararlarda yararlanılabilecek bir çerçeve sunmaktadır. Bu çerçeve içerisinde, işyerinde çeşitli amaçlarla yapılan ölçmelerde test alan bireylerin gereksinimlerine odaklanılmaktadır. Bu bağlamda, ölçmenin amacı, uygulama koşulları, değerlendirme yolları ve

raporlama sürecine ilişkin alınan kararların belirtilmesine önem verilmektedir.

Bu çalışmada, psikolojik testlerin geliştirilmesi, uyarlanması, uygulanması ve değerlendirilmesi süreçlerinde yol gösterici nitelikteki standartlara ilişkin bilgiler verilmiştir. Psikolojik testler, çok çeşitli amaçlarla ve farklı bağlamlarda bireyler hakkında bilgi toplanmak istendiğinde yaygın olarak başvurulan yollardan biridir. Bu nedenle, psikolojik testlerin amaçları doğrultusunda doğru biçimde geliştirilmesi ve kullanılması psikometrinin en önemli konuları arasında yer almaktadır. Bunu destekleyecek biçimde, Hambleton (2004; 2006), psikometri alanında ortaya çıkabilecek gelişmeleri ele aldığı çalışmasında, çoğunlukla psikolojik testlerin geliştirilmesi ve kullanılması süreçlerinde gerçekleşmesi beklenen gelişmelere değinmektedir. Bu gelişmeler içerisinde, testlerin uluslararası düzeyde kullanımı ve bundan kaynaklanabilecek uyarlama sürecine ilişkin sorunlar, yeni psikometrik modellerin geliştirilmesi ya da test analizlerinin yapılabileceği yeni teknolojilerin üretilmesi, bilgisayar teknolojilerindeki gelişmelere paralel olarak yeni madde türlerinin geliştirilmesi, bilgisayar ve internet ortamındaki testlerin yaygınlaşması, geri bildirim verme sisteminin değişmesi, farklı alanlarda çalışan bireylere ölçme uzmanlığının verilmesi konularına yer verilmektedir. Bu gelişmelerden özellikle test kullanımının uluslararası boyuta taşınması ve uyarlama sorunlarının günümüzde tüm ülkeler için geçerli bir durum olduğu düşünülmektedir. Çeşitli modellerin, standartların ve yönergelerin takip edilmesinin, ilgili süreçleri açık bir biçimde tanımlayarak bu konuda yaşanabilecek sorunları en aza indirgeyeceği düşünülmektedir. Bu nedenle, bu konuda çalışan araştırmacılara dünya genelinde test standartları konusunda gerçekleşen gelişmeleri izlemeleri, bu gelişmelerin Türkiye’de yapılan psikometrik çalışmalara daha fazla yansımaları sağlamak için ilgili standartları kültürel özellikleri de göz önünde bulunduracak biçimde uyarlayan çalışmalar yapmaları önerilmektedir.

EXTENDED ABSTRACT

**A Systematic and Comparative Review of
International Test Standards and Guidelines**

*

Ezgi Mor Dirlik – Özge Altıntaş – Seval Kula Kartal
Kastamonu University – Ankara University – Pamukkale University

In social sciences, researchers generally deal with the latent constructs which are hardly described. Hence, the measurement of these structures contains many difficulties because they cannot be observed directly. Only the behaviors of the humans display and imply these constructs, so in order to measure these constructs, humans' behaviors are observed. Due to their abstract qualities, researchers have to study these features meticulously. Since the 1800s, efforts to the development of psychological measurement tools have been launched with Webers' and Fechner's works on psychophysics. Thanks to the rapid developments in the social sciences with Wundt's studies on psychology laboratory, psychology has become a science field that includes academic and formal studies (Schultz and Schultz, 2011).

Psychological tests are used to measure the individuals' differences on these constructs under standard and objective conditions. In these tools, a sample of humans' behaviors are considered and individuals' reactions to these behaviors are detected. There are many descriptions of the psychological tests and the common point of these descriptions is the standardization of the process of test administration. If the standard conditions are not provided throughout the test administration process, the scores obtained from these tools may be not valid and reliable, which are the most crucial and desired features of the scores. So as to make the right decisions based on the scores of the psychological tests, the tests should be used under standard conditions for all test-takers. To obtain valid and reliable scores, the systematical process should be followed both in the test development and administration process.

To get valid and reliable scores with the psychological tests, the systematical and hierarchical process of test development should be followed. This process and standardization requirements of these tests have been proposed by several international associations such as the American Educational Research Association (AERA), the American Psychological Association (APA) and the National Council on Educational Measurement (NCME), the International Test

Commission (ITC), the International Standards Organization (ISO) and European Federation of Psychologists (EFPA). These institutions have offered several documents that describe and emphasize the process of test development, adaption, test usage and interpretation of the scores. Especially by the APA, the educational and psychological testing have been described in a very detailed perspective, from the development phase to the investigation of psychometric qualities. The same situation is valid for the ITC. The ITC (2017) released the updated version of test standards. Both the APA and ITC have recently updated the international guidelines for test use. The EFPA published a test review model in 2013 and this model has been translated into many European languages and accepted by many nations. The ISO suggested test administration standards, especially for vocational issues.

A model that guides test users, researchers and respondents will provide more valid, reliable results. It is clear that it will provide reproducible measurement processes. Repeatable measuring processes will increase the reliability of the results. As for Turkey, there is a limited number of studies focused on these issues and only individual efforts have occurred in order to adapt testing guidelines to Turkish. In 1998, Hovardaoğlu and Sezgin translated Standards for Educational and Psychological Testing, published in 1986, into Turkish. The translation study was organized by the Turkish Psychological Association. Since then, there has not been any attempt in Turkish literature to revise or update these standards. Several researchers have studied these standards and guidelines, however, there are no national accepted testing standards for Turkey. It is clear that psychological testing is affected by the cultural settings, hence rather than adapting the standards to Turkish, the development of our original standards will be more suitable and effective. The very first step of developing our original testing guidelines and standards is to examine the works done around the world. The investigation of the similarity and revealing the differences among the worldwide testing standards will lead researchers studying on this issue. In line with this intention, the researchers aimed to revise and analyze the documents on psychological testing and measurements that have been proposed by the listed institutions. By introducing these standards to the researchers, it is expected that it will be possible to increase the awareness of test users and all parties interested in the testing process regarding the test standards.

Methodology

Within the scope of this research, it is aimed to introduce the test standards and guidelines established by the worlds' leading institutions to all parties interested in the testing process in our country. For this purpose, the most important worldwide standards and guidelines have been determined. All relevant documents were tried to be reached via online resources. In order to reveal important studies and developments, these documents have been investigated. Regarding the main purpose of the study, this research, conducted on the basis of document review, is based on the descriptive approach.

The documents that were analyzed are the revised version of APA's standards, EFPA's test review model, ISO professional testing standards and ITC's updated test adaptation guidelines. First, these materials were analyzed by the authors and then ITC's test adaptation guidelines were translated into Turkish by the authors. The translated form was presented to an English language specialist who has the PhD degree and upon his recommendations, the form was revised.

Results

In this research, the development, adaptation and application processes of universal testing standards were analyzed and all relevant information was provided to the researchers. It is a clear fact that psychological tests are so essential part of measurement and assessment in education and psychology that it is possible to gather information about individuals for a wide variety of purposes and in different contexts with the help of psychological tests. Therefore, the process of the psychological test development and usage of the scores for appropriate purposes are the most important issues about the testing. In parallel with the developments in computer technologies, new test types, the spread of tests in computer and online environment, changes in the feedback system have become current issues. These developments, especially the use of tests in the international fields, and adaptation problems have gained importance for all countries. Turkey is one of these countries that has no original test standards and guidelines. Also, none of the models has been adopted in Turkish officially. Hence by investigating the various test review models, testing standards and guidelines, by defining the processes clearly, the problems that may be experienced in the development process will be minimized. Therefore,

the researchers studying these subjects are to follow the developments in test standards thoroughly in order to ensure more relevant reflection of psychometric studies in Turkey. Researchers are recommended to adapt standards that were published by the pioneer international institutions to our culture to consider the local characteristics of our country.

Kaynakça / References

- American Psychological Association, American Educational Research Association, & National Council on Measurement in Education (1966). *Standards for educational and psychological tests and manuals*. Washington, DC: American Psychological Association.
- American Psychological Association, American Educational Research Association, & National Council on Measurement in Education (1974). *Standards for educational & psychological tests*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anastasi, A. (1988). *Psychological testing*. (6th ed.). New York: Macmillan.
- Anastasi, A., and Urbina, S. (1997). *Psychological testing* (7th ed.). New Jersey: Prentice-Hall.
- Brennan, R. L. (2006). *Educational measurement*. Westport, CT: ACE/Praeger.
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper Collins.
- Çüm, S. ve Koç, N. (2013). Türkiye’de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarlama çalışmalarının incelenmesi. *Eğitim Bilimleri ve Uygulama Dergisi*, 12(24), 115-135.

- EFPA Board of Assessment (2013). EFPA Review Model for the Description and Evaluation of Psychological and Educational Tests, Version 4.2.6. <http://www.efpa.eu/professional-development/assessment>
- Evers, A., Sijtsma, K., Lucassen, W., and Meijer, R. R. (2011). The Dutch review process for evaluating the quality of psychological tests: History, procedure and results. *International Journal of Testing*, 10, 295-317.
- Hambleton, R. K. (2004). Theory, methods, and practices in testing for the 21st century. *Psicothema*, 16, 696-701.
- Hambleton, R. K. (2006, March). *Testing practices in the 21st century*. Key Note Address, University of Oviedo, Spain.
- Hovardaoğlu, S. ve Sezgin, N. (1998). *Eğitimde psikolojide ölçme standartları*. Ankara: Türk Psikologlar Birliği Yayınları.
- International Test Commission (2001). International guidelines for test use. *International Journal of Testing*, 1(2), 93-114. https://doi.org/10.1207/S15327574IJT0102_1
- International Test Commission (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, 6(2), 143-171. https://doi.org/10.1207/s15327574ijt0602_4
- International Test Commission (2012). *International guidelines on quality control in scoring, test analysis, and reporting of test scores*. https://www.intestcom.org/files/guideline_quality_control.pdf
- International Test Commission. (2014). *International guidelines on the security of tests, examinations, and other assessments*. https://www.intestcom.org/files/guideline_test_security.pdf
- International Test Commission (2015). *International guidelines for practitioner use of test revisions, obsolete tests, and test disposal*. https://www.intestcom.org/files/guideline_test_disposal.pdf
- International Test Commission. (2017). *The ITC guidelines for translating and adapting tests* (2nd ed.). https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf
- International Test Commission. (2018). *ITC guidelines for the large-scale assessment of linguistically and culturally diverse populations*. https://www.intestcom.org/files/guideline_diverse_populations.pdf
- International Test Commission. (2020). *New guidelines on technology-based assessments, in collaboration with the Association of Test Publishers (ATP)*. <https://www.intestcom.org/page/28>

- ISO (2011). *Procedures and methods to assess people in work and organizational settings* (Part 1 and 2). Technical committee: Human Resource Management.
- ISO (2020). *Procedures and methods to assess people in work and organizational settings*. Technical committee: Human Resource Management.
- Mor Dirlik, E. (2014). Ölçek geliştirme konulu doktora tezlerinin test ve ölçek geliştirme standartlarına uygunluğunun incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 5(2), 62-78.
- Mor Dirlik, E. ve Koç, N. (2017). The analysis of the psychological tests using in educational institutions according to the testing standards. *Journal of Measurement and Evaluation in Education and Psychology*, 8(4), 453-468.
- Muñiz, J., and Bartram, D. (2007). Improving international tests and testing. *European Psychologist*, 12, 206-219.
- Schultz, D. P., and Schultz, S. E. (2011). *A history of modern psychology* (10th Ed.). Belmont, US: Wadsworth, Cengage Learning.
- Urbina, S. (2004). *Essentials of psychological testing*. Hoboken, New Jersey: John Wiley and Sons, Inc.

Kaynakça Bilgisi / Citation Information

- Mor Dirlik, E., Altıntaş, Ö. ve Kula Kartal, S. (2021). Uluslararası test standartlarının ve yönergelerinin sistematik ve karşılaştırmalı incelenmesi. *OPUS-Uluslararası Toplum Araştırmaları Dergisi*, 18(39), 804-830. DOI: 10.26466/opus.889770.