5. BÖLÜM

TEST GELİŞTİRME SÜRECİ

TEST DEVELOPMENT PROCESS

Prof. Dr. Melek Gülşah ŞAHİN

TEST GELİŞTİRME SÜRECİ

Prof. Dr. Melek Gülşah ŞAHİN
Gazi Üniversitesi

Öz

Eğitimin pek çok aşamasında testler kritik öneme sahiptir. Testlere dayalı olarak alınan kararların geçerliğinin sağlanabilmesi her şeyden önce test geliştirme sürecinin uygun şekilde yürütülmesine bağlıdır. Testler geliştirilirken esnek ve dinamik yapıya sahip bir süreç takip edilir. Bu bölümde geniş ölçekli testler dikkate alınarak test geliştirme aşamaları birbiri ile ilişkili sekiz başlıkta sunulmuştur. Testin amacının belirlenmesiyle başlayan test geliştirme sürecinin ikinci aşamasında testin planlaması gerçekleştirilmekte; bu doğrultuda testin içerik, biçimsel, psikometrik, puanlama, uygulama ve zaman özelliklerine dair planlamalar yapılmaktadır. Plana dayalı olarak maddelerin yazılması, puanlama anahtarlarının oluşturulması ve uzman görüşlerinin alınması ise sürecin üçüncü aşamasını oluşturmaktadır. Uzman görüşlerinden hareketle gerekli düzenlemelerin gerçekleştirilmesinden sonra deneme formu oluşturulup deneme uygulaması yürütülmektedir. Bölüm kapsamında deneme uygulaması yapılırken uyulması gereken kurallar ile olası hata kaynaklarına karşı alınabilecek önlemlere de değinilmiştir. Bu aşamada sınav güvenliğine yönelik alınacak tedbirler de önem arz etmektedir. Dolayısıyla bölümde bu tedbirlerden de bahsedilmiştir. Deneme uygulamasının ardından madde ayırt ediciliği, madde güçlüğü, madde karakteristik eğrileri ve değişen madde fonksiyonu gibi madde istatistiklerinin incelenmesi aşaması gelmektedir. Madde istatistiklerine dayalı olarak seçilen maddeler bir sonraki aşamada belli bilgiler ile birlikte madde bankasına seçilerek testlerde kullanılmak üzere saklanabilmektedir. Nihai formun oluşturulmasıyla test geliştirme süreci sonlandırılmaktadır. Bununla beraber bölümde testin geçme puanlarının ve standartlarının belirlenmesi, testin uygulanması, test sonuçlarının ilanı ve testin teknik raporunun hazırlanması gibi test geliştirildikten sonra gerçekleştirilebilecek uygulamalara da yer verilmiştir.

Anahtar Kelimeler

Test Geliştirme Süreci, Test Özellikleri, Test Şeması, Testin Uygulanması, Test Teknik Raporu

TEST DEVELOPMENT PROCESS

Prof. Dr. Melek Gülşah ŞAHİN
Gazi University

Abstract

Tests are of critical importance at many stages of education. Ensuring the validity of decisions derived from test results depends primarily on the rigorous execution of the test development process. A flexible and dynamic process is followed when developing tests. In this section, the stages of test development are presented as eight interrelated phases, with a particular focus on large-scale tests. The test development process begins with defining the purpose of the test. The second phase involves planning the test, including its content, format, psychometric properties, scoring procedures, administration, and timing. The third phase of the process involves writing test items, creating scoring keys, and obtaining expert opinions in line with the plan. the necessary revisions have been made based on expert opinions, a pilot test is created and administered. This section also addresses the rules to be followed during pilot testing and the precautions needed to minimize potential sources of error. At this stage, measures to ensure test security are also of great importance. Accordingly, these measures are discussed in this section. Following the pilot test, the next phase involves analyzing item statistics, including item discrimination, item difficulty, item characteristic curves, and the differential item function (DIF). Based on these analyses, suitable items are added to the item bank together with their associated metadata for future test development. The test development process culminates in the creation of the final version. In addition, the chapter describes post-development procedures, including establishing passing scores and performance standards, administering the test, reporting the results, and preparing the technical report.

Keywords

Test Development Process, Test Specifications, Test Blueprint, Test Administration, Test Technical Report

1. GİRİŞ

Bireylerin hayatında alınan kararlarda önemli yeri olan testlerin varlığı çok eski zamanlara dayanmaktadır. Bugün olduğu gibi geçmişte de toplumlarda önemli kabul edilen bilgi ve becerilerin değerlendirilmesinde testler kullanılıyordu. Örneğin, Antik Yunanlılar spor ve edebiyat gibi alanlarda en iyi kişileri belirlemek için testlerden yararlanıyordu. Benzer şekilde Çin'de yaklaşık M.Ö 2. yüzyıla kadar dayanan testler memuriyete geçiş için kullanılmakta idi. 1800'lü yıllara gelindiğinde ise testler kişilik özelliklerini ve diğer gizil (örtük) yapıları belirlemek amacıyla kullanılmaya başlanmıştır (Cohen & Swerdlik, 2018; Kane, 2006). Matbaanın icadı, bilgisayarların ortaya çıkışı, internetin kullanımı, yapay zekâ teknolojilerindeki gelişmeler gibi insanlık tarihinde önemli rol oynayan faktörler; testlerin türünü, uygulanma ve değerlendirilme biçimini farklılaştırsa da testlere olan ihtiyaç yüzyıllardır değişmemiştir. Günümüzde sınıf-içi öğrenmelerin değerlendirilmesinde, kademeler arası geçiş sınavlarında, kamu personeli alım süreçlerinde, sertifika sınavlarında ve öğretim programlarının etkililiğinin tespit edilmesi gibi pek çok alanda testlerden yararlanılmaktadır.

Haladyna ve Rodriguez (2013), testleri standart koşullar altında bir özelliğin niteliğini veya derecesini saptamak için kullanılan araçlar olarak tanımlayarak standartlaştırma kavramının test puanlarının yorumlanması açısından önemli olduğunu vurgulamıştır. Standart koşullar altında uygulanması, testlerin geçerliği ve güvenirliği yüksek ölçümler üretebilmesi için kritik öneme sahip olmakla birlikte, tek başına yeterli değildir. Geçerliği ve güvenirliği yüksek ölçümlerin elde edilmesi, her şeyden önce test geliştirme sürecinin titizlikle yürütülmesine bağlıdır.

Test geliştirme süreci, bilimsel bir araştırma projesi hazırlamaya benzemektedir. Baykul (2000) test geliştirmeyi, özellikleri önceden kestirilmesi gereken bir süreç olarak tanımlamaktadır. Bu tanımda geçen "özelliklerini önceden kestirebilme" ifadesiyle iyi bir test planı hazırlanması kastedilmektedir. Dolayısıyla test geliştirmeye, bu süreçte görev alacak kişileri ve izlenecek adımları içeren ayrıntılı bir plan oluşturarak başlanmalıdır. İşlem adımlarının detaylı olarak planlanması ve her bir aşamada izlenecek hususların belirlenmesi, test geliştiricilerinin bu süreçte daha objektif kararlar almasını ve amacına uygun ölçme araçları geliştirmesini sağlamaktadır. Bu nedenle test geliştirme aşamaları ve her bir aşamada dikkat edilmesi gereken hususları bilmek son derece önemlidir.

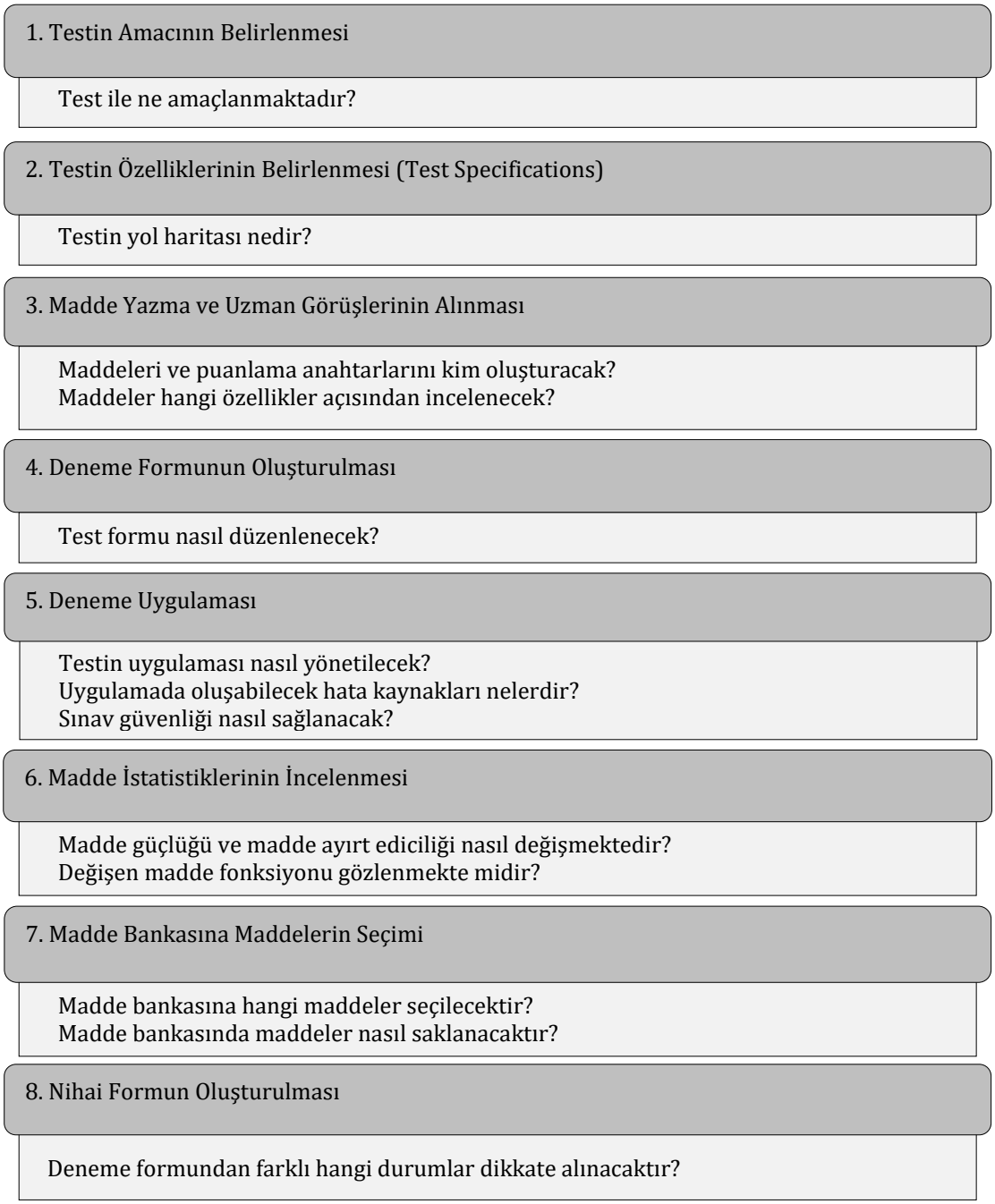
Kitabın bu bölümünde test geliştirme sürecinin esnek ve dinamik bir yapıya sahip olduğu göz önünde bulundurularak test geliştirme aşamaları sunulmuştur. Her bir aşama ayrı başlıklarda detaylandırılmıştır. Buna ek olarak; test geliştirme sürecinden sonra gerçekleştirilebilecek uygulamalara da bölümde kısaca değinilmiştir.

2. TEST GELİŞTİRME AŞAMALARI

Test geliştirme süreci birbirleriyle ilişkili aşamalardan oluşur. Alanyazın incelendiğinde (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), 2014; Baykul, 2000; Cohen & Swerdlik, 2018; Crocker & Algina, 2008; Çetin, 2019; Doğan, 2019; Downing, 2006; Murphy & Davidshofer, 2001; Özçelik, 1992; Turgut, 1992) bu aşamaların farklı araştırmacılar tarafından değişik şekillerde ele alındığı, genel hatlar aynı kalmakla beraber detaylarda birtakım farklılıklar bulunduğu görülmektedir. Bu bölümde ilgili kaynaklardan hareketle test geliştirme aşamaları organize edilmiştir. Sınıf-içi testler ile geniş ölçekli testlerde test geliştirme sürecinin farklılık gösterebileceği göz önünde bulundurulmuş, bölümde test geliştirme aşamaları geniş ölçekli testler temele alınarak yapılandırılmıştır. Test geliştirme süreci Şekil 1'de, sürecin her bir aşamasına ilişkin anahtar ifadeler ile birlikte özetlenmiştir.

Şekil 1

Test Geliştirme Süreci



2.1. Testin Amacının Belirlenmesi

Test geliştirme sürecindeki ilk basamak testin amacının belirlenmesidir. Testin amacına karar verirken; *i)* Test hangi amaç(lar) için kullanılacak, *ii)* Test ile hangi kararlar alınacak, *iii)* Test hangi yorumları getirecek ve *iv)* Test hangi gruba yönelik olarak geliştirilecek? sorularına cevap aranır. Test geliştirmenin diğer aşamaları testin amacına göre şekillenir. Bu anlamda testin amacını net bir biçimde ortaya koymak, süreci sağlam bir zemine oturtmak anlamına gelir. Baykul (2000) test geliştirmenin genel amaçlarını şu şekilde açıklamaktadır:

- **Psikolojik yapıların betimlenmesi, ilişkilerin incelenmesi:** Testler ilgi, kaygı, motivasyon, öz-yeterlik gibi psikolojik yapıları tanımlamak veya bu yapılar arasındaki ilişkileri ortaya koymak için geliştirilebilir.
- **Seçme ve yerleştirme**: Yükseköğretime geçiş, liselere öğrenci seçimi, bir mesleğe giriş veya bireylerin yabancı dil seviyelerinin tespiti amacıyla gerçekleştirilen testler örnek olarak verilebilir.
- **Öğrenciler hakkında eğitim kararlarının verilmesi:** Öğrencilerin öğrenme eksikleri ile güçlüklerini tespit etme, öğrenme düzeyini saptama ve öğrencileri yönlendirme amacıyla kullanılan testler bu gruba girer.
- **Öğretim ve öğretim programı hakkında kararların alınması**: Öğretim programının etkililiğini tespit etmek ve öğretim hizmetinin niteliğini belirlemek amacıyla gerçekleştirilen testler bu kapsamda yer alır.

Baykul'a (2000) benzer şekilde Grounlund da (1977) testlerin tanıma-yerleştirme, biçimlendirme veya düzey belirleme amacıyla hazırlanabileceğini belirtmiştir. Tanıma-yerleştirmeye yönelik testlerde öğrencinin ön bilgi ve hazırbulunuşluk düzeylerinin belirlenmesi hedeflenir. Bu amaca yönelik hazırlanan testte öncelikle öğrencilerde bulunması beklenen ön koşul davranışların temsil edilmesi önemlidir. Biçimlendirici değerlendirmeye yönelik testlerde amaç, öğrenme eksiklerini ve hatalı öğrenmeleri tespit ederek öğretmen ve öğrencilere öğrenmenin izlenmesine ilişkin geri bildirim vermektir. Dolayısıyla mümkünse bu tür testlerde öğrenme çıktılarının tamamına yer verilmelidir. Düzey belirlemeye dönük testlerde ise not veya sertifika verilmesi esastır. Bu testlerde her bir öğrenme çıktısıyla ilgili madde bulunması genelde mümkün olmaz. Bu nedenle, test öğrenme çıktılarını temsil eden bir madde örneklemi alınarak oluşturulur.

Görüldüğü üzere farklı amaçlara yönelik, çeşitli testler geliştirilebilmektedir. Testin tek ve spesifik bir amaca yönelik geliştirilmesi testten elde edilen sonuçların geçerliği için önemli bir husustur. Geçerlik kanıtları testin amacına göre belirleneceğinden amacın açık biçimde ifade edilmesi geçerlik çalışmalarına da yön verir (AERA, APA & NCME, 2014). Örneğin, amaç bir işe/eğitim kurumuna seçilecek kişileri belirlemekse geçerliğe ilişkin yordayıcı kanıtlar ile ölçme sonuçlarına dayalı geçerlik kanıtlarını incelemek gerekir. Geçerlik kanıtlarına dair detaylı bilgiler kitabın 3. bölümünde sunulmuştur.

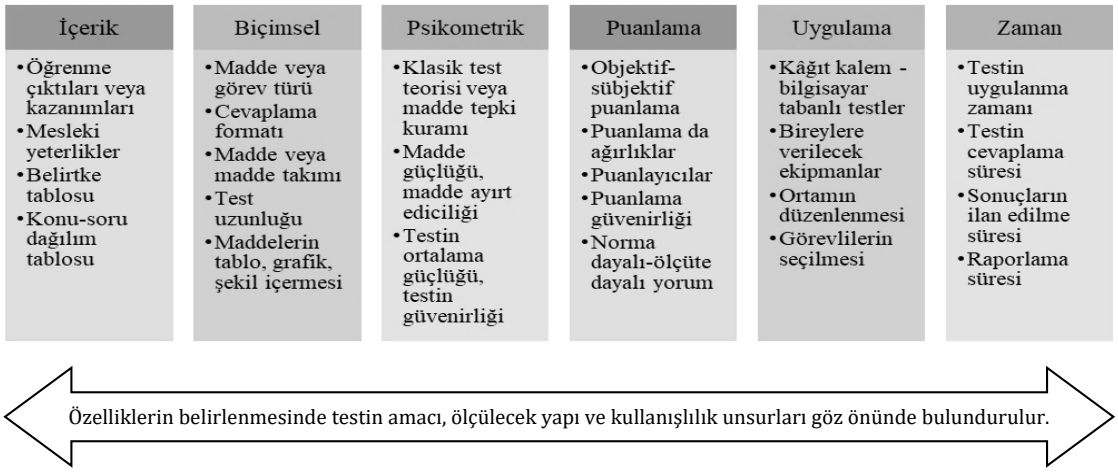
Testin amacı, testte yer alacak maddelerin türünü ve zorluk derecesini de etkiler (Crocker & Algina, 2008; Grounlund, 1977). Örneğin, öğrenme eksiklerinin tespitine yönelik hazırlanmış bir testte açık uçlu maddeler son derece işlevsel olabilirken doğru-yanlış türü maddeler amaca hizmet etmez. Yine çok sayıda başvuran kişi arasından az sayıda kişinin seçileceği bir sınavda, zor maddelerden oluşan bir test amaca hizmet ederken öğrenme güçlüğü yaşayan öğrencileri saptamaya yönelik bir sınavda zor maddeler testin amacı ile bağdaşmaz.

2.2. Testin Özelliklerinin Belirlenmesi (Test Specifications)

Testin özelliklerinin belirlenmesi aşaması, testin inşasına yönelik sistematik bir plan olarak kabul edilebilir. Bu aşamada kararlaştırılan bilgiler test geliştiriciler için kılavuz niteliği taşır. Hedef içeriğin testte nasıl temsil edileceği bu aşamada belirlenir. Test geliştirme süreçlerinin ele alındığı çalışmaların çoğunda bu aşama yer almamakta ya da yüzeysel biçimde sunulmaktadır. Crocker ve Algina (2008), test özelliklerinin belirlenmesini belirtke tablosunun oluşturulması ile sınırlamıştır. Diğer taraftan Downing (2006) test özelliklerinin belirlenmesini, testin özelliklerine yönelik operasyonel (işe vuruk) tanımları içeren daha ayrıntılı bir aşama olarak nitelendirmiştir. AERA, APA ve NCME (2014) tarafından yayınlanan Eğitsel ve Psikolojik Test Standartlarında da testin özelliklerinin belirlenmesi oldukça ayrıntılı tanımlanmış olup bu aşamada pek çok önemli karar alınmasını gerektiğine vurgu yapılmış ve ilgili kararlar amaç, içerik, biçimsel, psikometrik, puanlama ve uygulama (test yönetimi) başlıklarında sunulmuştur. Turgut (1992), testin özelliklerini belirlerken zaman ile ilgili de çeşitli kararlar almak gerektiğine değinmiştir. Test geliştiricisinin sıralanan başlıklarda alması gereken kararlar Şekil 2'de özetlenmiş ve her bir başlık bölümde ele alınmıştır.

Şekil 2

Testin Özelliklerinin Belirlenmesi Aşamasında Üzerinde Durulması Gereken Başlıklar ve Kararlar (AERA, APA & NCME, 2014; Turgut, 1992)



2.2.1. İçeriğin Belirlenmesi

İçerik özelliklerinin belirlenmesinde, yapı veya alanın içerik, beceri, süreç ve tanısal özellikleri kararlaştırılır (AERA, APA & NCME, 2014). Örneğin, yabancı dil seviyesini tespit etmek amacıyla uygulanacak bir sınavın alt testlerinde öğrencilerin İngilizce okuma becerilerinin saptanması amaçlanabilir. Bu amaca ulaşmak için Avrupa Ortak Dil Çerçevesi düzeyleri dikkate alınarak farklı düzeyler için okuma parçaları oluşturulabilir. Başka bir örnek verilecek olursa 6. sınıf öğrencilerinin fen bilimleri dersinde öğrenme eksiklerinin ortaya konulmasının amaçlandığı bir testte içerik öğrenme kazanımlarına göre şekillenmelidir. Kısacası bu aşamada; "Test ile ne ölçülmek istenmektedir, testin kapsamı nedir?" gibi sorulara cevap aranır.

İçerik belirlenirken hedef yapıyla ilgili;

- Görüşler alınabilir ve içerik analizi gerçekleştirilebilir,
- Kaynaklar ve araştırmalar incelenebilir,
- Kritik durumlar belirlenebilir,
- Doğrudan gözlemlere başvurulabilir,
- Öğretim materyalleri, ders kitapları, öğretmen-öğrenci ders notları, öğretim programları ve kazanımları incelenebilir (Baykul, 2000; Crocker & Algina, 2008).

Eğitimde testlerin içeriği öğretim programına veya öğrenme çıktılarına dayalı olarak yapılandırılır. Öğretim programındaki kazanımların / öğrenme çıktılarının dağılımı ile örtüşen bir test geliştirebilmek için genelde ***belirtke tablosu*** veya ***konu-soru dağılım tablosundan*** yararlanılır. Belirtke tablosu, ölçülmesi amaçlanan yapı, davranış veya kazanımların konular ile ilişkilendirildiği bir matristir. Belirtke tablosunda bir boyutta konu/ünite/kazanımlar listelenirken diğer boyutta bilişsel düzeyler bulunur (Turgut, 1992). Tabloda konuların öğretimdeki önemlerine dayalı olarak her bir bilişsel düzeyde yer ve ağırlık verilmelidir (Özçelik, 1992). Bilişsel düzeylerin belirlenmesinde farklı taksonomilerden (Bloom, revize edilmiş Bloom, Barrett, Marzano, Popham, Oosterhof vb.) yararlanılabilir. Hatta öğretim programında o programa özgü bir sınıflandırma söz konusuysa belirtke tablosu hazırlanırken programda tanımlanan bu sınıflandırmadan da yararlanılabilir. Örneğin, Türkiye Yüzyılı Maarif Modeli'nde yer verilen bilişsel düzeyler referans alınarak da bir belirtke tablosu oluşturulabilir. Tablo 1'de ölçme ve değerlendirme dersi ara sınavı için revize edilmiş Bloom taksonomisine göre düzenlenmiş bir belirtke tablosu örneği sunulmuştur.

Tablo 1

Ölçme ve Değerlendirme Dersi Ara Sınav İçin Belirtke Tablosu Örneği

İçerik	Hatırlama	Anlama	Uygulama	Analiz Etme	Toplam
Temel Kavramlar	2 (%8)	3 (%12)	1 (%4)	-	6 (%24)
Ölçme Araçlarının ve Ölçümlerin Psikometrik Nitelikleri	2 (%8)	2 (%8)	2 (%8)	1 (%4)	7 (%28)
Test Geliştirme Süreci	1 (%4)	1 (%4)	-	-	2 (%8)
Geleneksel ve Tamamlayıcı Ölçme Araçları	3 (%12)	4 (%16)	3 (%12)	-	10 (%40)
Toplam	8 (%32)	10 (%40)	6 (%24)	1 (%4)	25 (%100)



- Geniş ölçekli sınavlarda öğretim programındaki tüm kazanımların ölçülmesi zaman ve maliyet açısından pek mümkün olmaz. Bu nedenle test geliştiricinin içerikteki hangi kazanımların daha önemli olduğunu belirlemesi ve içerik dengesini gözeterek maddeleri yapılandırılması gerekir (Crocker & Algina, 2008). Yani kritik davranışlara ve kazanımlara karar verilmelidir.
- Belirtke tablosu hazırlanırken test puanlarına dayalı olarak verilecek kararın önemi de göz önüne alınmalıdır. Test ile alınacak karar ne kadar kritikse belirtke tablosu o kadar ayrıntılı hazırlanmalıdır (Çetin, 2019; Downing, 2006).

Konu-soru dağılım tabloları, ölçmeye konu içeriğin hangi bileşeninin / öğrenme çıktısının testte hangi ağırlıkta temsil edildiğini ortaya koymaya yönelik bir tablodur. Ulusal ve uluslararası sınavlara yönelik kamuoyu ile paylaşılan konu-soru dağılım tabloları bunun tipik bir örneğidir**.** Örneğin Millî Eğitim Bakanlığı, Ölçme ve Değerlendirme Sınav Hizmetleri Genel Müdürlüğü (2025) gerçekleştireceği ortak sınavlar için her bir ders bağlamında oluşturulacak senaryolara ilişkin konu-soru dağılım tabloları yayımlamıştır. Aşağıda 8. sınıf fen bilimleri için hazırlanmış örnek konu-soru dağılım tablosu sunulmuştur.

Tablo 2

*8. Sınıf Fen Bilimleri Dersi Birinci Dönem İkinci Ortak Yazılı Sınavı İçin Örnek Konu-Soru Dağılım Tablosu**

Öğrenme Alanı	Kazanımlar	Madde Sayısı
Dünya ve Evren	F.8.1.1.1. Mevsimlerin oluşumuna yönelik tahminlerde bulunur.	1
Canlılar ve Yaşam	F.8.2.2.2. Tek karakter çaprazlamaları ile ilgili problemler çözerek sonuçlar hakkında yorum yapar.	2
	F.8.2.3.3. Mutasyon ve modifikasyon arasındaki farklar ile ilgili çıkarımda bulunur.	1
Fiziksel Olaylar	F.8.3.1.1. Katı basıncını etkileyen değişkenleri deneyerek keşfeder.	1
	F.8.3.1.2. Sıvı basıncını etkileyen değişkenleri tahmin eder ve tahminlerini test eder.	2
Madde ve Doğa	F.8.4.1.2. Elementleri periyodik tablo üzerinde metal, yarı-metal ve ametal olarak sınıflandırır.	1
	F.8.4.2.1. Fiziksel ve kimyasal değişim arasındaki farkları, çeşitli olayları gözlemleyerek açıklar.	1

**Fen Bilimleri 8. Sınıf Konu Soru Kitapçığı*, Millî Eğitim Bakanlığı (MEB), 2025, (URL-1) adresinden alınmıştır.

Testin içeriğini planlarken belirtke tablosu veya konu-soru dağılım tablosu oluşturmanın dışında daha farklı süreçler de işletilebilir. Örneğin bir test mesleki yeterlik düzeylerini saptamak veya işe alınacak kişileri seçmek amacıyla kullanılacaksa testin içeriğini netleştirirken iş analizine başvurulabilir (AERA, APA & NCME, 2014). Bu kapsamda program geliştirme çalışmalarında ihtiyaç analizi yaparken gerçekleştirilen meslek analizi ile progel - dacum tekniğine benzer adımlar izlenebilir. İş analizinde literatürdeki hazır ölçme ve değerlendirme araçlarından yararlanabileceği gibi her bir mesleğe özgü görev ve faaliyetlerin listesini içeren görev envanterleri de oluşturulabilir. Görev envanterleri meslekteki bireylere uygulanan, görevlerin sıklığı, zorluğu ve önemine yönelik dereceleri içerir (Raymond & Neustel, 2006). Burada başarı testlerinden farklı olarak işin gerektirdiği özellikler net olarak tanımlanmalıdır. Elde edilen veriler üzerinden gerçekleştirilecek içerik analizi mesleğin gereklilikleri ve mesleğe özgü kritik özellikleri saptamaya yardımcı olur.

Testin içeriğinin kim tarafından belirleneceği de önemli bir husustur. Bu aşamada testin türü de etkili olmaktadır. Testin yüksek veya düşük riskli olması önemli bir faktördür. Sınıf içi değerlendirmelerde içerik öğretmen tarafından daha basit ve resmî olmayan biçimde organize edilebilirken, yüksek risk arz eden sınavlarda testin içeriğini daha sistematik biçimde belirlemek gerekir. Bu bakımdan paydaşlardan en az biri için yüksek

risk taşıyan geniş ölçekli testlerde içerik, sınavı uygulayan kurum tarafından oluşturulan komisyonlarca resmî süreçlere dayalı olarak oluşturulmalıdır. Bu tür sınavlarda geçerlik kanıtları için güvenlik ön plana alınarak savunulabilir süreçler işletilmelidir (Çetin, 2019; Downing, 2006).

Belirlenen içeriğin testin amacıyla örtüştüğünü kanıtlayabilmek için geçerlik çalışmaları gerçekleştirilmelidir. Downing (2006) yüksek riskli sınavlarda içeriği oluştururken çok aşamalı bir uzman görüşü alma süreci takip etmeyi önerir. İlk olarak; öğretim programı ve materyalleri ile ders kitapları incelenerek, alan uzmanları ile öğretim üyelerinin, ölçme ve değerlendirme uzmanlarının görüşleri alınarak kapsamlı bir içerik (özellik/kazanım) listesi oluşturulmalıdır. İkinci aşamada ise başka bir uzman grubundan belirlenen test içeriğinin öğrencilerin yeni bilgileri öğrenmesinde veya mesleği icradaki önemini dereceleme ölçekleri aracılığıyla değerlendirmeleri istenebilir. Ardından değerlendirmeyi yapan uzmanların görüşlerinden yola çıkılarak içerik gözden geçirilir. Bu aşamada uzmanlar arasındaki uyumu da dikkate almak gerekir. Uzmanlar arasındaki uyumun hesaplanmasında, uzman sayısı göz önüne alınarak, basit uyum yüzdesinden, Kappa istatistiklerinden (Cohen'in kappası, Fleiss'in kappası, ağırlıklandırılmış kappa) ya da Krippendorf alfa katsayısından faydalanılabilir, genellenebilirlik kuramı ve/veya çok yüzeyli Rasch modeline dayalı analizler gerçekleştirilebilir.

İçeriğin belirlenmesinde geçerliği tehdit edebilecek iki tür sınırlılık söz konusudur: Kapsam eksikliği ve kapsam dışı unsurlar. İlgili yapının/içeriğin eksik temsil edilmesi kapsam eksikliği, yapı ile ilgisiz özelliklerin testin içeriğine karışması ise kapsam dışı unsurlar olarak tanımlanır. Örneğin sözel iletişim becerisini ölçmeye dönük bir test dinleme becerisiyle ilgili maddeler içermiyorsa bu test hedef kapsamı yeterince temsil edemiyor demektir. Bahse konu testte yazma becerisine yönelik maddelerin bulunması ise kapsam dışı özelliklerin teste dâhil edildiği anlamına gelir. Kapsama yönelik bu iki sınırlığın üstesinden gelebilmek için hedef yapının çerçevesi net olarak ortaya konmalı, ölçmeye dâhil edilecek ve hariç tutulacak özelliklerin listesi oluşturulmalıdır (Kline, 2005).

2.2.2. Biçimsel Özelliklerin Belirlenmesi

Biçimsel özelliklerin belirlenmesi aşamasında amaca en uygun madde veya görev türüne karar verilir. Tüm amaca uygun tek bir madde türünden söz etmek olanaksızdır. Alanyazında maddeler veya görevler özelliklerine göre çeşitlilik gösterir ve maddelerin sınıflandırmasına ilişkin birçok farklı yaklaşım vardır (Crocker & Algina, 2008; Osterlind, 2002; Rodriguez, 2002; Turgut, 1992). Genel anlamda maddeler; yanıtı seçmeyi gerektiren ve yanıtı katkı (yapılandırmayı) gerektiren maddeler şeklinde ikiye ayrılabilir. Çoktan seçmeli, doğru yanlış ve eşleştirmeli maddeler yanıtın seçimine dayalı madde türleridir. Şans başarısı içeren ve puanlanışı itibariyle objektif nitelik taşıyan bu maddelerde birey sunulan seçeneklerden birini ya da birkaçını işaretleyerek cevabını verir. Açık uçlu, kısa cevaplı ve boşluk doldurma tipinde maddeler ise yanıtı

yapılandırmayı gerektiren madde türleridir. Bu maddeler bireye cevaplama özgürlüğü tanır ve bu özgürlük yanıtın ne kadar sınırlandırıldığına bağlı olarak değişir. Ayrıca bu maddeler yanıtı seçmeyi gerektiren maddelerin aksine puanlanışı bakımdan subjektif özellik taşır ve şans başarısı içermez. Madde türleriyle ilgili detaylı bilgi edinmek için ilgili kaynaklar ve bu kitabın 6. bölümü incelenebilir.

Testin amacı ve içeriğine yönelik olarak testte tek bir madde türüne yer verilebilir veya test farklı türde maddeler içerecek şekilde hazırlanabilir. Öğrencilerin matematik dersinde problem çözme becerilerini değerlendirmeye yönelik bir test geliştirmek amaçlandığında, test çoktan seçmeli, kısa cevaplı veya açık uçlu maddelerden oluşturulabilir. Madde türü ile sayısına karar verirken hedef kitlenin özellikleri (yaşı, herhangi bir engel durumunun olup olmadığı vb.), grubun büyüklüğü, maliyet ve cevapların ilan edilme süresi gibi değişkenler göz önünde bulundurulmalıdır. Bir mesleğe yerleştirme ya da mesleki yeterlikleri saptama amacıyla yapılan bir sınavda ise mesleki becerileri değerlendirmeye hizmet edecek performans görevleri geliştirilebilir. Bu aşamada sınırlandırılmış ya da sınırlandırılmamış performans görevinin seçimine de karar verilmelidir. Becerinin türü, değerlendirmeye esas ölçütler, zaman ve maliyet unsurları bu seçimde kritik rol oynayacaktır.

Testin amacının veya yapıya ilişkin teorinin başlangıçta net olarak belirlenebildiği durumda, test özelliklerinin her bir aşaması planlanır ve süreçte bu plan ile uyumlu bir ilerleme sağlanabilir. Ancak test geliştirmenin geneli gibi bu süreç de değişmez değildir. Gerekli durumlarda değişikliklere gidilebileceği unutulmamalıdır. Daha açık ifade etmek gerekirse kimi zaman hedef yapının teorik çerçevesi önceden net olmaz, maddelerden edilen verilerden hareketle oluşturulması gerekir. Bu aşamada madde puanlarının ölçülecek yapıyla ilişkili başka özelliklere ait puanlarla korelasyonuna da bakılabilir (AERA, APA & NCME, 2014). Örneğin ölçme ve değerlendirme uzmanlığına yönelik bir mesleki yeterlik sınavında, bazı maddeler bireylerin istatistik özyeterlik ölçeğinden elde ettikleri puanlar ile korelasyonuna dayalı olarak seçilebilir.

2.2.3. Psikometrik Özelliklerin Belirlenmesi

Psikometrik özelliklerin belirlenmesi aşamasında madde ve test istatistiklerine yönelik kararlar alınır. Bu doğrultuda klasik test teorisi bağlamında madde ayırt ediciliği ve madde güçlüğü hesaplanabilir ya da madde tepki kuramı bağlamında madde karakteristik eğrileri incelenebilir. Açık uçlu maddelerden oluşan testlerde madde istatistikleri hesaplanmadan önce puanlamaların güvenirliği de kontrol edilmelidir. Örneğin dört dil becerisinin değerlendirildiği bir yabancı dil sınavında her bir beceride yer alan maddelerin/görevlerin geçerliğini belirlemek amacıyla madde/görev ayırt edicilik değerleri, madde-toplam test korelasyonları hesaplanabilir. Ayrıca konuşma ve yazma becerilerine yönelik puanlar için puanlayıcılar arası güvenirlik katsayıları hesaplanabilir.

2.2.4. Puanlama Özelliklerinin Belirlenmesi

Puanlama özelliklerinin belirlenmesi aşamasında maddelerin nasıl puanlanacağı kararlaştırılır. Elbette bu kararda belirleyici olan madde türüdür. Yanıtı seçmeyi gerektiren maddelerde çoğunlukla 0-1 şeklinde ikili puanlama yapılır. Yanıtı katkı gerektiren maddelerde ise genelde kısmen doğru yanıtların da değerlendirildiği çoklu bir puanlama benimsenir. Bu aşamada madde yazarları, hazırladıkları maddeler için doğru ve kısmen doğru kabul edilecek cevapları belirtmelidir. Becerilerin ölçülmesi söz konusu olduğunda ise performansa dayalı değerlendirme kapsamında puanlama anahtarına (kontrol listesi, dereceleme ölçekleri, rubrik) karar verilmelidir. Örneğin, bir mesleğe yerleştirme ya da mesleki yeterlikleri saptama amacıyla geliştirilen performans görevlerine uygun bütünsel ya da analitik rubrik, genel ya da göreve özel rubriklerden hangisinin kullanılacağına bu aşamada karar verilir. Ayrıca testten alınacak toplam puanda madde ağırlıklarının nasıl yer alacağı bu aşamada kararlaştırılmalıdır. Yine puanlama süreçleri detaylı biçimde planlanmalı; puanlayıcıların nasıl belirleneceği, puanlamanın kaç kişi tarafından, ne zaman ve nasıl gerçekleştirileceği, puanlayıcılar arası uyuşmanın hangi yöntemle hesaplanacağı, olası puanlayıcı yanlılıklarının nasıl önleneceği bu aşamada kararlaştırılmalıdır.

2.2.5. Uygulama Koşullarının Belirlenmesi

Uygulama koşullarının belirlenmesi aşamasında yanıt aranması beklenen sorular Şekil 3'te sunulmuştur.

Şekil 3

Testin Uygulama Koşullarını Belirlerken Yanıt Aranması Gereken Sorular

Testin uygulama formatı nedir?
Test ne kadar süre içerisinde uygulanacaktır?
Bireylere hangi ekipman / ekipmanlar verilecektir?
Sınav yerleşim planı nasıl olacaktır?
Bilgisayar ortamında uygulanacak testler için özel bir yazılım ya da donanım gerekmekte midir?

Uygulama koşullarının belirlenmesi aşamasında ilk olarak testin bireylere hangi şekilde (kâğıt kalem - bilgisayar tabanlı) uygulanacağına karar verilmelidir. Testin uygulamasına kalan süre ve testin uygulamasında bireylere verilecek ekipman / ekipmanlar da uygulama koşulları olarak kararlaştırılması beklenen hususlardır. Bilgisayar tabanlı

testlerin uygulanması söz konusu ise bir diğer husus ise internet bağlantısının gerekliliğinin sorgulanmasıdır. Son olarak sınav yerleşim planı ve görevlilerin tespitine ilişkin karar da bu aşamada alınabilir.

Burada yanıt aranmayan sorular dışında testin formatına dayalı farklı kararlar da söz konusu olabilir. Örneğin, bir yabancı dil seviye tespit sınavının katılımcılara bilgisayar tabanlı uygulanacağını varsayalım. Testin doğrusal mı yoksa bireye uyarlanmış mı olduğu önemlidir. Uyarlanmış testlerde madde seçim algoritmasına da karar vermek gerekir. Maddelerin uygulanmasında göz önünde bulundurulması gereken bir başka kritik husus bireylerin maddeye maruz kalma indeksleridir. Testin uzunluğu doğrusal testlerde sabitken uyarlanmış testlerde sabit veya değişen uzunlukta olabilir. Burada içerik dengelemeye dönük incelemeler de gerçekleştirilebilir. Bilgisayar ortamında bireye uyarlanmış testler ve özellikleri hakkında ayrıntılı bilgi için 15. bölüm incelenebilir.

2.2.6. Zaman Koşullarının Belirlenmesi

Zaman koşullarının belirlenmesi aşamasında testin hazırlanma takvimi, uygulanacağı tarih, uygulamanın kaç saat süreceği, sonuçların ne kadar sürede ilan edileceği ile ilgili planlamalar yapılır. Testin uygulama süresini belirlerken pek çok değişkeni hesaba katmak gerekir. Örneğin testin uygulama yönergeleri söz konusu ise bunun için geçecek zaman test süresine dâhil edilmemeli, ek süre ayarlanmalıdır. Ya da sınavda birden fazla kitapçık olduğu durumda, her bir kitapçık için ayrı bir süre belirlenmelidir. Bir kitapçık toplandıktan sonra diğer kitapçık dağıtılacaksa arada geçen kayıp zaman sınav bitiş saatine eklenmeli ve testin toplam süresinden düşülmemelidir. Özel gereksinimli adaylara ek süre tanınacaksa hangi engel grubundaki adaya fazladan ne kadar süre verileceği de bu aşamada belirlenmelidir.

Testin özelliklerinin belirlenmesi aşamasında, ulusal ve uluslararası test merkezlerinde çoğunlukla testin özelliklerinin bir tablo ile gösterildiği **test şeması (test blueprint)** kullanılır. Testler hazırlanırken test şeması referans alınır. Bazı kaynaklarda test şeması ile belirtke tabloları birbirinin yerine kullanılır. Ancak test şeması, belirtke tablosundan daha fazla bilgi (maddelerin bilişsel düzeyi, türü, sayısı, güçlük düzeyi, puanlamadaki ağırlıkları ile testin uygulanma biçimi [elektronik, basılı vs.], süre gibi) içerir. Test şemaları geniş ölçekli sınavlar yapan test merkezleri için madde ve performans görevlerinin geliştirilmesine, sonuçları raporlanmasına ve araştırmalara da yön verir. (URL-2). Tablo 3'te 20 maddeden oluşan bir test için örnek test şeması yer almaktadır.

Tablo 3

Örnek Test Şeması

İçerik	Öğrenme Hedefi	Bilişsel Düzeyi	Madde Türü	Madde Sayısı	Madde Ağırlığı	Güçlük Düzeyi (1-3)	Puanlama	Süre (dk.)
Konu 1	H1.1	Hatırlama	Kısa cevaplı	4	%20	1	4*3	4*1
	H1.2	Anlama	Çoktan seçmeli	2	%10	1-2	2*5	2*2
	H1.3	Uygulama	Açık uçlu	1	%5	3	1*10	1*5
Konu 2	H2.1	Anlama	Çoktan seçmeli	2	%10	2-3	2*5	2*2
	H2.2	Uygulama	Açık uçlu	2	%10	2-3	2*10	2*5
Konu 3	H3.1	Hatırlama	Doğru-yanlış	4	%20	1	4*3	4*1
	H3.2	Anlama	Kısa cevaplı	2	%10	1-2	2*3	2*1
	H3.3	Uygulama	Çoktan seçmeli	2	%10	1-3	2*5	2*2
	H3.4	Uygulama	Açık uçlu	1	%5	3	1*10	1*5
Toplam				**20**	**%100**		**100**	**42**

2.3. Maddelerin Yazılması ve Uzmanlardan Görüş Alınması

Testler her zaman sorulardan oluşmaz, bazen çözülmesi istenen bir problem veya yargıda bulunulması beklenilen önermelerden meydana gelir. Kitabın 6. bölümünde ifade edildiği üzere madde terimi bu uyarıcıları kapsar ve sorudan daha geniş bir anlama sahiptir (Vale, 2006). Dolayısıyla testin puanlanabilen en küçük birimi için soru yerine madde terimi tercih edilir (Baykul, 2000). Testin özelliklerinin belirlenmesi aşamasında madde türü ve sayısı ile puanlama anahtarı gibi özelliklere karar verilirken bu aşamada aşağıdaki süreçler izlenir.

2.3.1. Madde Yazımında ve Denetiminde Görev Alacak Kişilerin Belirlenmesi

Bir testte yer alan maddelerin kalitesi madde yazarlarının ne kadar nitelikli madde yazabildiği ile ilişkilidir. İyi yazılmamış bir madde için uygun gözden geçirme süreçleri işe koşulsa ve düzeltmeler gerçekleştirilse de nitelikli bir maddeye ulaşma ihtimali zayıftır. Bundan dolayı madde yazarları alanında uzman kişiler arasından titizlikle seçilmelidir. Madde yazarı maddeyi bilimsel hata içermeyecek, herkes tarafından aynı şekilde anlaşılacak ve ölçme ilkeleri ile uyumlu olacak şekilde hazırlamalı, maddenin doğru cevabını belirlemeli ve güçlük indeksine ilişkin tahminde bulunmalı, ayrıca meslektaşları tarafından hazırlanan maddelerin incelemesinde uzman değerlendirici

olarak görev alabilmelidir (Haladyna, 2004). Kısacası madde yazarı ilgili konu alanını veya psikolojik yapıyı iyi tanımalı, ölçme ve değerlendirmenin temel kavramlarını bilmelidir. Baykul (2000) yaratıcılık gerektiren bir iş olmasından dolayı madde yazmanın sanatsal bir yönü olduğunu da ifade etmektedir. Ek olarak madde yazarları; sözel iletişim becerileri gelişmiş, farklı madde türlerine hâkim ve testi alan grubu yeterince tanıyan kişiler olmalıdır. Dahası madde yazarları sabırlı olmalı; madde yazmanın zamanla gelişen bir beceri olduğunu unutulmamalıdır (Mehrens & Lehmann, 1991).

Bu aşamada madde yazarlarının sayısı da belirlenmelidir. Bu sayıyı belirlerken testin içeriği, belirtke tablosuna göre öngörülen madde sayısı, yazarların uzmanlık alanları vb. özellikler göz önünde bulundurulmalıdır. Madde yazarı sayısı kadar, ekipteki kişilerin ne kadar heterojen olduğu da önemlidir. Test geliştirme sürecinde madde yazarlarında olabildiğince çeşitliliğe (kıdem yılı, mezuniyet düzeyi, cinsiyet vb.) gidilmelidir. Böylece yazılan maddelerde çeşitlilik sağlanabilir.

Madde yazarlarının madde yazım kurallarını ne kadar işletebildiği bir diğer önemli hususdur. Madde yazarları madde türleri ve madde yazımında dikkate edilecek hususlarla ilgili hem teorik hem de uygulamalı eğitimlere alınmalıdır. Downing'e (2006) göre ideal olan test geliştirmede görev alan tüm madde yazarlarının madde yazma ve test geliştirme üzerine eğitimlere katılması ve uzmanlığının olmasıdır. Eğitimlerin hazırlanması ve uygulanması, bazı durumlarda zaman ve maliyet açısından ekonomik olmayabilir. Bu gibi durumlarda madde yazarlarına yönelik kılavuz bir kitapçık hazırlanması yoluna gidilebilir. Kitapçıkta; yönergeler, madde formatı, uygun ve uygun olmayan örnekler, madde yazma ve incelemeye ilişkin kurallar yer alabilir. Bu kitapçıklar yeni soru yazarlarının bir eğitimden geçmesine gerek kalmadan madde yazımına ilişkin bilgileri edinmesine hizmet edeceği gibi mevcut yazarlara istedikleri zaman madde yazımına ilişkin ilkeleri tekrar etme imkânı sunar (Haladyna, 2004).

2.3.2. Madde Yazım Kuralları Çerçevesinde Maddelerin Yazılması

Madde türüne göre maddelerin yazımında dikkat edilecek hususlar alanyazında net biçimde tanımlanmış olup konuyla ilgili kaynaklar incelenebilir (Baykul, 2000; Haladyna, 2004, Haladyna ve ark., 2002; Haladyna & Rodriguez, 2013; Mehrens & Lehmann, 1991; Osterlind, 2002; Russell & Airasian, 2012; Turgut, 1992). Ayrıca bu kitabın 6. Bölümüne bakılabilir. Burada testin biçimsel özelliklerini belirlemek amacıyla oluşturulan belirtke tablosu, konu-soru dağılım tablosu veya test şeması madde yazarları tarafından incelenir. Bu doğrultuda maddeler oluşturulur ve hazırlanan maddeler testin madde havuzunu meydana getirir.

Madde havuzuna genellikle test formunda yer alacak madde sayısından daha fazla sayıda madde dâhil edilir (AERA, APA & NCME, 2014). Deneme uygulaması henüz gerçekleştirilmediğinden testin nihai formunda yer alması planlanan madde sayısının en

az üç katı madde içeren bir madde havuzu hazırlanması önerilir. Tabi bu sayının madde yazarlarının ne kadar deneyimli olduğuna ve ne kadar nitelikli madde yazabildiğine göre değişebileceği unutulmamalıdır (Baykul, 2000). Şöyle ki madde yazarları deneyimli ise yazılan maddelerin görece az bir kısmının elenmesi beklenir, dolayısıyla madde havuzu daha az sayıda maddeden oluşabilir. Diğer taraftan madde yazarlarının tecrübesi düşükse yazılan maddelerin elenme ihtimali daha yüksek olur ki bu durumda madde havuzuna dâhil etmek üzere daha fazla sayıda madde yazılması gerekebilir.

Gelişen teknoloji ile madde yazımı bütünüyle insan gücüne bağlı bir süreç olmaktan çıkmıştır. Artık yapay zekâ tabanlı araçlar sayesinde otomatik madde üretimi gerçekleştirilebilmekte, bu da daha kısa sürede, daha az maliyetle daha büyük madde havuzlarının oluşturulmasına olanak tanımaktadır. Otomatik üretilen maddeler için de tıpkı insanlar tarafından yazılan maddelerde geçerli olan uzman görüşü alma süreçleri hayata geçirilmelidir. Otomatik madde üretimi hakkında daha detaylı bilgi edinmek isteyen okuyucular kitabın 15. bölümünü inceleyebilir.

2.3.3. Maddelerin Cevap/Puanlama Anahtarının Hazırlanması

Madde yazarları, hazırladıkları her bir madde için cevap ya da puanlama anahtarı hazırlamalıdır. Yanıtı seçmeyi gerektiren maddeler objektif biçimde ve çoğunlukla 0-1 puanlandığından yanıtları kolaylıkla hazırlanabilir. Yanıtı katkı gerektiren maddelerin puanlanmasında ise puanlama subjektif nitelik taşır ve 0-1 dışında puanlamalar da söz konusudur. Çoktan seçmeli maddeler, doğru-yanlış maddeler, eşleştirmeli maddeler gibi objektif puanlanan ölçme araçları için madde ve yanıtları uzman görüşlerine sunulmalı, görüş birliğinin sağlanması ile cevap anahtarının nihai hali oluşturulmalıdır. Açık uçlu madde ve performans görevi gibi subjektif puanlanan ölçme araçlarında ise ölçme aracının hazırlanması ile eş zamanlı olarak puanlanmada kullanılacak anahtarın da (kontrol, listesi, dereceleme ölçeği veya rubrik) hazırlanması gerekmektedir. Ayrıca puanlama sürecinde geçerliği ve güvenirliği yüksek sonuçlar elde edilmesi için ölçme aracından bağımsız olarak puanlama anahtarının da geçerlik ve güvenirliği incelenmelidir. Bu amaçla uzman görüşü alınabilir ve farklı ölçme kuramlarına göre puanlayıcı güvenirliğine yönelik çeşitli analizler gerçekleştirilebilir. Örneğin, İngilizce yazma becerisinin değerlendirilmesinde kullanılacak bir rubrik, uzman görüşlerine dayalı olarak incelenip revize edilebilir. Revize edilen rubrik küçük bir grupta sınanarak ilgili yapıya uygunluğu, kategori sayısının yeterli olup olmadığı, kategoriler arasındaki farkların belirginliği ve güvenilir puanlamalar üretip üretmediği incelenebilir.

Madde yazarlarının yazdıkları maddeleri cevap ya da puanlama anahtarları ile birlikte bir forma/sisteme aktarması, maddenin künyesini oluşturmak ve geçerlik denetimlerinde uzmanlar tarafından incelenmesini kolaylaştırmak için önemlidir. Formlarda madde yazarının bilgisi, dersin adı, konu ve öğrenme alanları, hedeflenen içerik (kazanım-davranış), tahmini madde güçlüğü, yararlanılan kaynaklar, madde ve

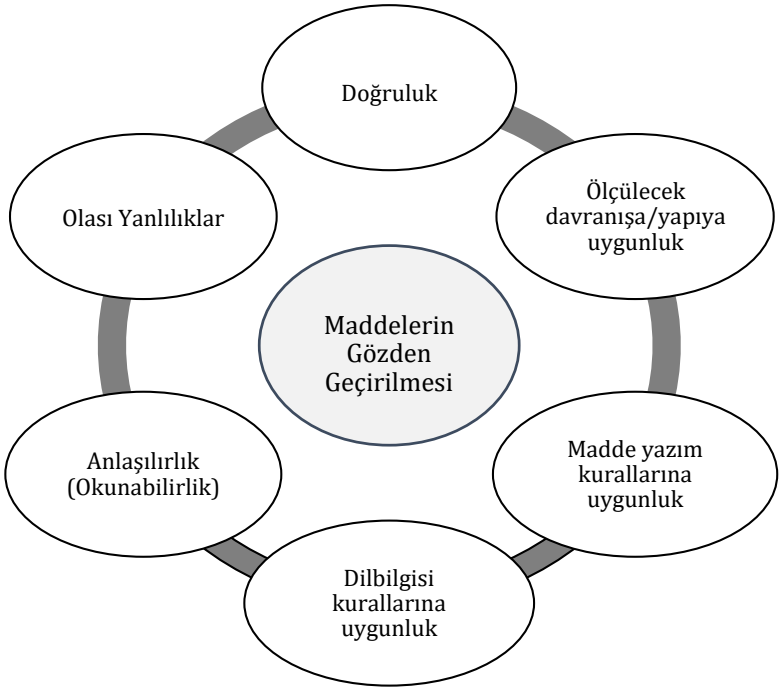
maddenin doğru yanıtı gibi bilgiler yer alabilir (Çetin, 2019). Formlar basılı olarak elde doldurulacağı gibi günümüzde çoğunlukla bilgisayar ortamında madde havuzuna entegre olarak doldurulabilir.

2.3.4. Maddelerin Gözden Geçirilmesi

Madde yazımından sonra maddeler Şekil 4 te yer alan özellikler açısından alan uzmanı, ölçme ve değerlendirme uzmanı ve dil uzmanı tarafından incelenmelidir (AERA, APA & NCME, 2014; Crocker & Algina, 2008).

Şekil 4

Maddeleri Gözden Geçirirken İncelemede Bulunması Gereken Özellikler



Alan uzmanları maddelerin bilimsel doğruluğu, hedef yapıya uygunluğu ve yanlılık içerip içermediği hakkında değerlendirme yapar. Ölçme ve değerlendirme uzmanları her bir maddeyi madde yazım ilkelerine uygunluk açısından denetler ve maddenin yanlılık içerip içermediğine dair görüş bildirir. Dil uzmanları ise maddelerin dilbilgisi, yazım-noktalama kurallarına uygunluğunu ve anlaşılır olup olmadığını kontrol eder. Maddeler uygulanmadan önce uzmanlar tarafından yanlılık incelemesine tabi tutulması sonuçların geçerliği açısından son derece önemlidir. Maddeleri gözden geçirme aşaması bir kalite kontrol süreci olarak düşünülebilir. Bu sayede madde yazarının gözden kaçırdığı noktalar uzmanlar tarafından tespit edilebilir, farklı bir bakış açısıyla maddeler yeniden değerlendirilebilir.

Sınıf içi ölçmelerde öğretmenler hazırladıkları maddeler için bir başka öğretmenden görüş isteyebilirler. Ancak geniş ölçekli sınavlarda maddelerin birden fazla uzmanın görüşüne sunulması gerekir. Uzmanlar inceledikleri maddenin uygun olduğu yönünde görüş bildirebilir, düzeltilerek kullanılmasına veya reddine karar verebilirler. Uzmanların düzeltme önerileri varsa madde, madde yazarları tarafından tekrar gözden geçirilir ve gerekli düzenlemeler gerçekleştirilir. Bu aşamada en az iki uzmandan görüş alınması tavsiye edilir. Böyle bir durumda uzmanlar arasında görüş ayrılığının olması halinde üçüncü bir uzmanın görüşüne başvurulur. Yüksek riskli sınavlarda ise en az üç uzmandan görüş alınması daha uygun bir seçenektir.

Uzmanlardan görüş alırken çoğunlukla her bir uzman diğer uzmanlardan bağımsız olarak maddeleri denetler. Ancak süreci daima bu şekilde ilerletme zorunluluğu yoktur. Duruma göre uzmanlar bir araya gelerek uzman paneli şeklinde de maddeler hakkında inceleme yapabilir.

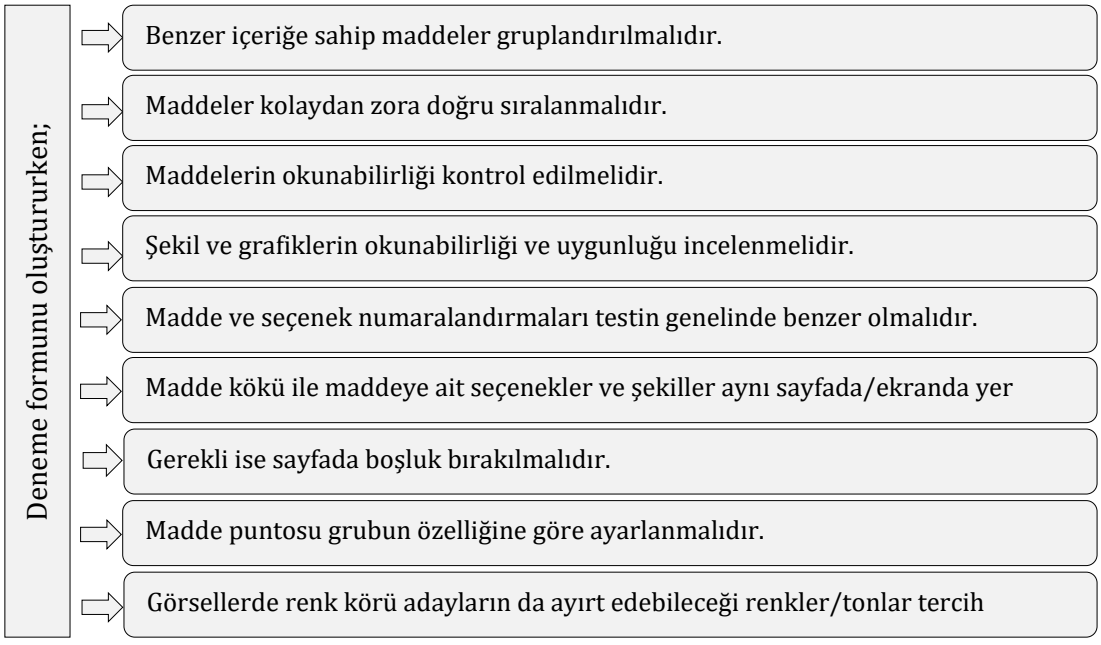
Madde yazarken ve maddelerle ilgili uzmanlardan görüş alırken gözetilmesi gereken bir diğer husus güvenliktir. Tüm sınavlarda madde yazma ve denetleme işlemleri daha önceden belirlenmiş güvenlik prosedürleri dâhilinde gerçekleştirilmelidir. Özellikle yüksek riskli sınavlarda madde yazımında ve denetiminde görev alan uzmanlar belirli gizlilik ve güvenlik protokolüne tabi tutulmalıdır. Örneğin madde yazarları ile gizlilik anlaşmasının imzalanması, madde yazımının ve denetlenmesinin üst düzeyde güvenlik önlemlerinin alındığı ortamlarda yapılması, madde yazarının birinci derece akrabalarının katılım sağlayacağı sınavlara ilişkin madde yazım süreçlerinden muaf tutulması, bir madde yazarı tarafından hazırlanan çok sayıda maddeye aynı sınavda yer verilmemesi gibi kurallar işletilebilir.

2.4. Deneme Formunun Oluşturulması

Bu aşamada test maddeleri denenmek üzere bir testte veya formda birleştirilir. Testin başlangıçta belirlenen amaca ne kadar uygun olduğunu saptamanın yollarından biri madde istatistiklerini incelemektir. Bunun için deneme uygulaması gerçekleştirilir. Deneme uygulamasının ardından bazı maddelerin öngörülen şekilde çalışmadığı tespit edilirse esas uygulama öncesinde bu maddelerin testten çıkarılması gerekir. Çıkarılan maddelerin ardından testin kapsam geçerliğinin korunması için deneme formuna nihai testte yer alacak madde sayısından daha fazla sayıda madde yerleştirilmelidir. Deneme formu, Şekil 5'te sunulan hususlar dikkate alınarak hazırlanmalıdır.

Şekil 5

Deneme Formu Hazırlanırken Dikkat Edilmesi Gereken Hususlar





- Güvenlik riskleri sebebiyle yüksek riskli sınavlarda deneme uygulaması yapılması mümkün olmayabilir, dolayısıyla deneme formunun oluşturulması aşaması atlanır ve doğrudan nihai uygulamaya geçilir. Deneme uygulaması yapılamadığından yüksek riskli sınavlarda maddeleri yazma ve gözden geçirme süreci çok daha titizlikle yürütülmelidir.
- Günümüzde makine öğrenmesi algoritmalarından yararlanılarak deneme uygulamasına gerek kalmadan maddelerin güçlüğü ve yanıt süresi gibi değişkenler kestirilebilmektedir (Fulari & Rusert, 2024; Razavi & Powers, 2025; Yousefpoori-Naeim ve ark., 2024). Bu anlamda deneme uygulamasının yapılamadığı durumlarda makine öğrenme modellerinden yararlanılması önerilebilir.
- Deneme uygulaması gerçekleştirmenin zor olduğu durumlarda başvurulabilecek bir diğer yol; bireylere testi uygularken test formunda belli bir oranda denenmek istenen maddelere yer vermektir. Daha açık bir anlatımla test formu sadece puanlamada dikkate alınacak maddelerden oluşmaz, sonraki uygulamalar için geliştirilen ve nitelikleri kontrol edilmek istenen maddeler de içerir (College Board, 2022). Tabi böyle bir uygulamaya başvurulduğunda denenmek üzere kullanılan maddeler katılımcılarda yorgunluğa neden olmamalı ve katılımcıların asıl maddelerdeki performansını farklılaştırmamalıdır. Bu nedenle puanlamaya dâhil edilmeyip sadece psikometrik niteliklerini belirlemek amacıyla test formuna eklenen maddeler sınırlı sayıda tutulmalıdır.

Yüksek riskli sınavlarda sıkça tercih edilen, çoktan seçmeli maddelerden oluşan testlerde yukarıda belirtilen hususların yanında, doğru cevabın seçeneklere dağılımının dengeli olmalıdır. Örneğin dört seçenekli bir testte her bir seçeneğin doğru yanıt olma olasılığı %25'tir. Maddeler bağımsız hazırlandığından bu adımda gerekli düzenlemeler yapılmazsa doğru cevap hep belli seçeneklere yerleşebilir. Şans başarısına yol açmaması için doğru cevap seçeneklerinin belli bir örüntü oluşturmamasına dikkat edilmelidir (Haladyna, 2004). Ek olarak kesin bir kural olmamakla birlikte arka arkaya 3'ten fazla maddenin doğru cevabının aynı seçenek olmamasına özen gösterilmelidir.

Deneme formu ya da formları oluşturulduktan sonra tüm formlara standart bir yönerge eklenmelidir. Yönergede testin amacı, madde sayısı, testin süresi, cevaplamada dikkate edilecek hususlar (örneğin cevapların optik forma nasıl kodlanması gerektiği) ve puanlamanın nasıl yapılacağı (örneğin düzeltme formülünün uygulanıp uygulanmayacağı) gibi bilgiler yer almalıdır. Unutulmamalıdır ki testi alan bireylerin testi geliştirenin amaçladığı şekilde cevap vermesi yönergelerin açık ve net olmasına bağlıdır. Bu bakımdan standart yönergeler test materyalinin bir parçası olmalıdır (AERA, APA & NCME, 2014).

Impara ve Foster (2006) test uygulamasında kopya çekilmesi, madde havuzundaki maddelerin çokça kullanımından dolayı adaylar tarafından bilinmesi (maddeye maruz kalma oranının artması) ve maddelerin (test veya havuzun) uygulama öncesinde izinsiz ele geçirilmesi gibi geçerliği ve güvenirliği tehdit edebilecek birçok olumsuz durumun, test formunun hazırlanma sürecinde işe koşulacak önlemler ile kontrol altına alınabileceğini belirtmiştir. Impara ve Foster'ın (2006) bu kapsamda önerdiği stratejiler ilgili kaynaktan yararlanılarak aşağıda özetlenmiştir:

- ***Testte yer alan cevap seçeneklerinin rastgele sıralanması:*** Cevapların dizin olarak ezberlenmesini önler. Kâğıt kalem testlerinde bu stratejinin kullanılması halinde birden fazla cevap anahtarı oluşturulmalıdır. Bilgisayar tabanlı sınavlarda yazılım sayesinde bu işlem kolaylaşmaktadır. Seçeneklerde sayılar varsa madde yazım kurallarına göre mantıksal bir sıra izlenmesi gerektiğinden bu strateji önerilmez.
- ***Testin farklı kitapçıklarının ya da formlarının kullanılması***: Testin farklı kitapçıklarının oluşturulması ile testteki maddelerin ve seçeneklerin yerlerinin değiştirildiği testlerin oluşturulması ifade edilmektedir. Testin farklı formları ile test maddelerinin belirli bir oranda (%60 ve %75 aralığında değişen) veya tamamının değiştiği ve ilk forma eşitliğinin sağlandığı formlarının oluşturulması belirtilmektedir. Her iki durumda da adayların soru sıralamasını ezberlemeleri, maddeleri ya da cevapları ezberlemeleri zorlaşacaktır. Kâğıt–kalem testlerinde farklı kitapçıkların oluşturulması maddeler aynı olduğundan dolayı daha ekonomiktir. Testin farklı formlarının geliştirilmesinde ise ekstra madde yazmak gerektiğinden maliyet artmaktadır. Ayrıca test formlarının eşitliğinin sağlanması da

ele alınacak diğer bir süreç olarak karşımıza çıkmaktadır. Bilgisayar tabanlı bireye uyarlanmış testlerde ise bireylerin yetenek düzeyleri göz önünde bulundurularak farklı maddelerden oluşan testler uygulanmaktadır.

- ***Bilgisayar tabanlı testlerin kullanılması***: Bilgisayar tabanlı testler doğrusal biçimde veya bireye uyarlanmış şekilde uygulanabilmektedir. Bireye uyarlanmış testlerde doğrusal testlere kıyasla daha az sayıda madde ile bireylerin yetenek düzeylerine ilişkin benzer kestirimler elde edebilmektedir. Ayrıca bireye uyarlanmış testlerde her aday farklı sorular ve testler alabileceğinden kopya çekilmesi büyük oranda önlenmektedir. Bireye uyarlanmış testlerdeki bazı maddeler, maddeye maruz kalma oranları artıkça adaylar tarafından tanınabilir, yani çok fazla karşılaşılan maddeler söz konusu olabilir. Bu da bir güvenlik riski oluşturup geçerlik ve güvenirliği olumsuz yönde etkileyecektir. Bununla birlikte, yazılımda gerçekleştirilecek ayarlamalar ve büyük madde havuzlarının kullanımı ile bu sınırlılık önemli ölçüde önlenebilir.
- ***Eşdeğer maddelerin kullanılması***: Testlerde madde veya cevap anahtarının ezberlenmesinin önüne geçmenin yollarından biri eşdeğer maddeler kullanmaktır. Bu sayede maddeler ezberlense dahi maddenin eşdeğerinin çözülmesi zorlaşmaktadır. Maddelerin eşdeğerini oluşturmak birçok alanda mümkün olup eşdeğer maddelerin yazılması paralel testlerin oluşturulmasında işlevsel bir rol üstlenmektedir. Sayısal sorularda sayıların değiştirilmesi veya sayıların aynı kalması ancak içeriğin farklılaştırılması ile eşdeğer maddeler oluşturulabilmektedir. Sözel sorularda ise ezberlenme ihtimali olan sözcük veya sözcük grupları değiştirilerek eşdeğer maddeler hazırlanabilmektedir.

2.5. Deneme Uygulaması

Pilot uygulama olarak da isimlendirilen bu aşamada, geliştirilen taslak form testin hedef grubuna benzer fakat daha az sayıda bireyden oluşan bir gruba uygulanır. Bu grubun asıl uygulamanın yapılacağı kitle ile iletişime geçme olasılığının bulunmadığından emin olunmalıdır. Testin deneme formları çoğaltılırken sıkı güvenlik önlemleri alınmalıdır. Buna göre test maddeleri baskı için güvenli olarak sisteme aktarılmalı, fazla veya hatalı basımlar güvenli şekilde imha edilmeli ve test formları güvenli olarak sınav ortamına taşınmalıdır. Son olarak herhangi bir güvenlik açığının oluşmaması için testin belli aşamaları belirli uzmanlara yalnızca sınırlı bir süre açılmalıdır (Downing, 2006). Test formu çoğaltılırken eksik veya hatalı basımların (örneğin sayfanın eksik olması, bazı maddelerin ve/veya görsellerin çıkmaması gibi) kontrolü sağlanmalıdır. Çok sayıda kitapçık olduğunda bu aşamada rastgele seçimler yapılabilir ve seçilen formların farklı uzmanlar tarafından incelemesi sağlanabilir. Ayrıca testin çeşitli formları mevcutsa bu formların denkliği de incelenmelidir. Olası hatalara karşı testi alacak kişi sayısından daha fazla form basılmalıdır.

Deneme uygulamasından önce uzmanların gözden geçirdiği maddeler için cevaplayıcılardan soru hakkında görüşlerini almak ve madde yazarı tarafından öngörülemeyen bir cevap ya da düşünme biçimi olup olmadığını anlamak amacıyla küçük bir grupta ön deneme uygulaması da gerçekleştirilebilir (Baykul, 2000). Ön denemede maddelerde anlaşılmayan noktaların bulunup bulunmadığına ve süre kullanımına ilişkin bilgiler elde edilebilir. Cevaplama sırasında uzun duraksamalar, karalamalar yapma ve cevap değişikliğine gitme durumları maddenin anlaşılırlığının düşük olduğuna dair ipucu sağlayabilir (Crocker & Algina, 2008). Erkuş (2019) ön deneme uygulamasının hedef grubun uç değerlerini (başarılı- başarısız, yüksek yetenek-düşük yetenek vb.) temsil eden 10-15 kişiden oluşan heterojen bir grupta gerçekleştirilmesini, uygulama sırasında öğrencilerin soruları ve çözüm yollarını sesli düşünüp yanıtlamalarının istenmesini ve uygulayıcının süreçte bizzat yer alarak yanıtlayıcının davranışlarını gözlemesini önermiştir. Ön deneme uygulamasına katılan öğrenciler soruları sesli okurken ve yanıtlarken ses/görüntü kaydı alınması tavsiye edilir. Böylece uygulayıcının dışında farklı uzmanların da uygulama süreçlerini incelemesi sağlanabilir ve bu uzmanlardan geribildirim alınarak geçerliğe dair cevaplama süreçlerine dayalı kanıtlara ulaşılabilir. Bilgisayar tabanlı yürütülen ön deneme uygulamalarında ise sesli düşünme protokolüne ilave olarak göz izleme (eye tracking) sistemleri kullanılabilir.

Ön deneme uygulamasından elde edilen bulgular doğrultusunda maddelerde gerekli düzenlemeler gerçekleştirildikten sonra, nihai uygulamaya katılacak gruba benzer nitelikteki bir başka grupta deneme uygulaması gerçekleştirilir. Deneme uygulamasının yapılacağı örneklemin büyüklüğü belirlenirken grubun özelliği, testin türü, madde sayısı, zaman sınırı, maliyet, gerçekleştirilecek analizler vb. değişkenler göz önünde bulundurulmalıdır. Deneme uygulamasında; testi uygulayan kişilerin ve öğrencilerin sınav kurallarına uymaları sağlanmalı, öğrenciler sınava güdülenmeli ve sınav ortamı uygun şekilde düzenlenmelidir. Sıralanan başlıklarla ilgili gerekli düzenlemeler yapılmadığı takdirde ölçümlerin psikometrik özelliklerini zedeleyen hata varyansları oluşur. McCallin (2006) geçerlik ve güvenirliği düşüren hata varyansına kaynaklık edebilecek faktörleri “grubu etkileyen” ve “bireyleri etkileyen” kaynaklar şeklinde ikiye ayırmıştır (bkz. Tablo 4).

Tablo 4

Testin Uygulanmasında Grubu ve Bireyi Etkileyen Hata Kaynakları[*]

	Hata Kaynağı	Örnekler
Grubu Etkileyen	Test uygulama esaslarının değiştirilmesi	Zaman değişimi, test yönergesinde değişikliğe gitmek, farklı uygulama biçimleri kullanmak vb.
	Test uygulama merkezleri veya ortamları arasındaki farklılıklar	Koşulların farklılığı, materyallerin farklılığı vb.
	Testin uygulama aşamasında meydana gelen sorunlar	Adayların yerinin değiştirilmesi, elektrik kesintisi, gürültü vb.
	Testin uygulandığı saatin farklı olması	Adaylar için performanslarının düşük olduğu saatte testin uygulanması
	Puanlama anahtarında değişiklik	Adayların değerlendirme ölçütlerinin testin uygulaması esnasında değişimi
	Teknolojik değişkenlerden kaynaklı ilgisiz varyans kaynakları	Donanımda oluşabilecek sorun, diğer teknik nedenler vb.
Bireyi etkileyen	Test maddelerini önceden alma	Test güvenliğinin ihlal edilmesi
	Adayın performansını etkileyecek test kaygısı	Başarılı olma kaygısı, gözetmenin adayın üzerinde baskı hissi oluşturması vb.
	Gözetmenin kopyaya sebebiyet vermesi	Aday istemeden gözetmenden ipucu alması
	Adayın motivasyon düşüklüğü	Adayın bilerek düşük performans göstermeyi seçmesi
	Adayın yönergeleri yanlış anlaması	Gözetmenin adayın yönergeyi anladığından emin olmaması
	Adayın testte nasıl ilerleyeceğini bilmemesi	Adayın cevabı nasıl kaydedeceğini bilememesi, test sistemini uygun kullanamaması

[*] *McCallin'dan (2006) uyarlanmıştır.*

Tabloda yer alan hata kaynakları incelendiğinde üç önemli kavram göze çarpmaktadır. Bunlar; uygulama koşullarının standartlaştırılması, sınav güvenliği ve gözetmenliktir. Bu kavramları her biri aşağıda açıklanmıştır (Downing, 2006):

- *Standartlaştırma*: Testlerin tüm bireylere mümkün olduğunca aynı koşullarda uygulanması istenmeyen değişkenlik kaynakları önlemeye, yani sonuçları hata varyansından arındırmaya yardımcı olur ve geçerliği artırır.
- *Sınav Güvenliği*: Testin uygulanmasında güvenlik her zaman en temel unsurlardandır. Özellikle yüksek riskli sınavlarda güvenlikle ilgili üst düzey önlemler alınması gerekir. Bu önlemlerden ilki test geliştirme sürecinin birçok

aşamasını kısıtlı sayıda bireyin erişimine açmaktır. Testin üretilmesi ve uygulanması sırasında sürece dışarıdan dâhil olan birey sayısı artacağından güvenlik zinciri genişler. Kâğıt kalem testlerinde test materyallerinin uygulama sahasına gönderilmesi, kuryeler, bina ve salon görevlileri tarafından teslim alınması, testlerin adaylara dağıtılması ve adaylardan toplanması, sınav evraklarının sınav merkezine gönderilmek üzere hazırlanması gibi aşamalarda sınav güvenliğini sağlamaya yönelik kurallar titizlikle uygulanmalıdır. Bilgisayar tabanlı testlerde kitapçıkların basılması, uygulanması ve değerlendirilme aşamaları olmadığından sıralanan başlıklarla ilgili güvenlik ihtiyacı kâğıt kalem testlerine kıyasla nispeten daha azdır. Ancak bilgisayar tabanlı testlerde daha başka güvenlik açıklarının olabileceği ve bunlara yönelik önlemler alınması gerektiğinin altı çizilmelidir. Örneğin, sınavların içeriği ve adayların cevapları şifrelenmeli, adayların kişisel verileri teknik yöntemler ile kontrol altına alınmalıdır (AERA, APA & NCME, 2014). Test bankasına ait dosyalar yerel sınav sunucusuna iletilmeden önce şifrelenmeli, maddelerin tamamı tek bir dosya olarak değil ihtiyaç duyuldukça kısmi olarak sunucudan çekilmeli, sınav merkezindeki sunucu bulunduğu yerde fiziksel olarak korunmalı ve tüm bu güvenlik önlemleri kayıt altına alınmalıdır (Becker & Pomplun, 2006).

- *Gözetmenlik*: Testlerin uygulanmasında sınav kurumları uygulama ve denetlemede yardımcı olması için eğitimli ve deneyimli gözetmenlerden yararlanılmalıdır. Bilgisayar tabanlı testlerde maddeler ve yanıtlar sistem üzerinden iletildiğinden bu sınavlar kâğıt kalem testlerine kıyasla daha az sayıda görevli/gözetmen ile yürütülebilmektedir.

2.6. Madde İstatistiklerinin İncelenmesi

Test geliştirmede temel alınan ölçme kuramına bağlı olarak, deneme formundaki maddelerin madde istatistikleri/özellikleri (madde güçlüğü, madde ayırt ediciliği, madde karakteristik eğrisi, değişen madde fonksiyonu), testin amacına uygun bir nihai form oluşturmak için incelenir. İlgili istatistikler hakkında kısaca bilgi sunulmuştur, daha kapsamlı bilgi edinmek için kitabın 11. bölümü incelenebilir.

2.6.1. Madde Güçlük İndeksi (p_j)

Madde güçlük indeksi, klasik test kuramı'na göre maddenin grupta doğru yanıtlama oranı olup 0 ile 1 arasında değerler alabilir. Madde güçlük indeksinin 0 olması maddeyi grupta kimsenin doğru yanıtlayamadığı anlamına gelirken 1 olması gruptaki tüm kişilerin maddeyi doğru yanıtlandığını belirtir. Anlaşılacağı üzere madde güçlük indeksi 0'a yaklaştıkça madde zorlaşırken 1'e yaklaştıkça madde kolaylaşır. Madde güçlük indeksinin 0,50 olması ise grubun yarısının maddeyi doğru cevapladığını ve maddenin orta güçlükte olduğunu yansıtır. Madde güçlüğü 0 ve 1 değerini aldığında farklı yetenek düzeyindeki bireyler birbirinden ayırt edilemez. Madde orta güçlükte olduğunda (p_j=0,50) ise farklı yetenek düzeyindeki kişiler birbirinden en iyi şekilde ayırt edilebilir ve beraberinde madde varyansı [(p_j) × (1–p_j)] maksimum değerini alır**.** Bu sebeple

testteki maddelerin çoğunun orta güçlükte olması, kolay ve zor madde sayısının daha sınırlı sayıda tutulması önerilir. Madde güçlüklerine ilişkin böyle bir dağılım, bireylerin ölçülen özellik bakımından en iyi şekilde ayırt edilmesini sağlar. Bununla beraber, testin uç gruplar için kullanımı amaçlanıyorsa daha zor veya daha kolay maddelerin ağırlıkta olduğu testler oluşturmak gerekebilir (Allen & Yen, 1979). Yani teste seçilecek maddeler için ideal güçlüğün ne olduğu testin amacına göre şekillenir. Çok sayıda başvuranın olduğu bir test ile az sayıda eleman alınması amaçlanıyor ise zor maddeler teste seçilebilir. Test sadece temel yetenek düzeyindeki bireyleri ayırt etmek amacıyla geliştirilecek ise kolay maddeler teste alınabilir.

2.6.2. Madde Ayırt Ediciliği (r_j)

Madde ayırt ediciliği, maddenin ölçülen gizil özelliğe farklı düzeylerde sahip olan kişileri birbirinden ayırt etme düzeyidir. Klasik Test Kuramı'na göre, madde ayırt ediciliği alt ve üst grup karşılaştırmalarına ve/veya madde ile toplam test puanı arasındaki korelasyona dayalı olarak hesaplanabilir. Madde ayırt edicilik düzeyi -1,00 ile +1,00 arasında değişmekte olup; maddelerin ayırt ediciliğinin 0,40 ve üzerinde olması mükemmel ayırt ediciliği gösterir. Eğer $0{,}30 \leq r_j \leq 0{,}39$ ise madde yeterince ayırt edicidir, herhangi bir düzeltme olmaksızın teste alınabilir. $0{,}20 \leq r_j \leq 0{,}29$ aralığındaki değerler ise maddenin düzeltildikten sonra teste alınabileceğini belirtir. Dolayısıyla bu maddeler uzmanlara sunulmalı, uzman görüşleri doğrultusunda gerekli revizyonlar yapılarak testte kabul edilmelidir (Crocker & Algina, 2008). Ayırt edicilik indeksi 0,20'nin altında kalan maddelerin ise testten çıkarılması gerekir. Maddede ne tür değişikliklerin yapılması gerektiği konusunda çoktan seçmeli testler için çeldirici analizleri de yol gösterici olabilmektedir.

2.6.3. Madde Karakteristik Eğrisi

Madde karakteristik eğrileri, maddenin doğru cevaplama olasılığı ile bireyin yetenek düzeyi arasındaki ilişkinin gösterildiği grafiklerdir. Bu grafikte yatay eksen yetenek düzeyini, dikey eksenin ise maddenin doğru yanıtlanma olasılığını temsil eder. Madde karakteristik eğrisine bakılarak maddeyi doğru cevaplama olasılığının yetenek düzeyine göre nasıl değiştiği görülebilir (Crocker & Algina, 1986), madde güçlük ve madde ayırt edicilik değerleri belirlenebilir. Maddenin doğru yanıtlama olasılığının 0,50 olduğu yetenek düzeyi madde güçlüğünü ifade eder. Madde karakteristik eğrisinin eğimi ise madde ayırt ediciliği hakkında bilgi verir. Daha dik olan eğriler, madde ayırt ediciliğinin yüksek olduğuna işaret ederken daha düz olan eğriler maddenin ayırt ediciliğinin düşük olduğunu yansıtır. Madde karakteristik eğrisi düz olan maddelerde, düşük ve yüksek yetenek düzeylerinde maddenin doğru yanıtlanma olasılığı, yaklaşık aynı olmaktadır (Baker, 2001). İyi bir madde için madde karakteristik eğrisi pozitif eğime ve orta düzey güçlüğe sahip olmalıdır (Allen & Yen, 1979).

2.6.4. Değişen Madde Fonksiyonu

Değişen madde fonksiyonu, maddenin aynı yetenek düzeyindeki farklı alt gruplarda (cinsiyet, yaş grubu, okul türü, etnik köken) farklı şekilde çalışıp çalışmadığını ortaya koyar (Zumbo, 1999). Örneğin, bir maddede cinsiyet değişkeni açısından değişen madde fonksiyonu gözlenmesi, aynı yetenek düzeyindeki bir kadın ile erkeğin ilgili maddeyi doğru cevaplama olasılıkların farklı olduğunu bildirir (Furr, 2022). Bir madde için değişen madde fonksiyonu saptanması o maddenin yanlı olduğu anlamına gelmez, madde üzerinde daha fazla araştırma yapmak gerektiğine ilişkin ikaz niteliği taşır (Skaggs, 2023). Çünkü bu durum maddenin karşılaştırılan gruplardan birinin aleyhine çalışmasından kaynaklanıyor olabileceği gibi gruplardan birinin maddenin ölçtüğü özellikle ilgili ön öğrenmelerinin yetersizliğinden de kaynaklanıyor olabilir (Popham, 2025). Dolayısıyla maddenin bir gruba haksız avantaj sağlayıp sağlamadığını, diğer bir deyişle yanlı olup olmadığını belirlemek için değişen madde fonksiyonu gösteren sorular uzman incelemesine sunulmalıdır. Uzmanların maddede yanlılık bulunduğuna kanaat getirmesi halinde test geliştiriciler maddeyi değiştirmeli ya da düzenlemelidirler (AERA, APA & NCME, 2014). Değişen madde fonksiyonunu incelemek için klasik test kuramına ve madde tepki kuramına dayalı analizler gerçekleştirilebilir (Camili & Shepard, 1994).

2.6.5. Madde Bankasına Maddelerin Seçimi

Madde yazımı zorlu ve yoğun emek gerektiren bir süreçtir. Maddelerin uygulamaya hazır hale gelene kadar geçtiği süreçler düşünüldüğünde maddelerin gelecek testlerde kullanılmak üzere madde bankalarında bekletilmesi test geliştirmede sıkça kullanılan bir prosedürdür. Madde bankaları maddelerin bir sistematik dâhilinde saklandığı sistemlerdir ve bu sistem belirli kurallar çerçevesinde çalışır. Madde bankası olarak günümüzde bilgisayar tabanlı sistemler kullanılmaktadır. Madde bankası oluşturmada bilgisayarların kullanımı 1976'lara uzanmaktadır (Vale, 2006). Madde yazım ilkeleri doğrultusunda hazırlanan maddeler uzman (alan, ölçme ve dil) görüşlerine sunulduktan sonra gerekli düzenlemeler yapılır ve deneme uygulamasına geçilir. Deneme uygulamadan sonra hesaplanan güçlük ve ayırt edicilik indeksleri dikkate alınarak seçilen maddeler madde bankasına alınır. Madde bankası için çeşitli sistemler kullanılabilir. Nitelikli bir madde bankası sistemi esnek ve güvenilir olmalı, maddeler madde bankasına Şekil 6'da yer verilen ilkeler doğrultusunda kaydedilmelidir (Downing, 2006).

Şekil 6

Maddelerin Madde Bankasına Kaydederken Takip Edilmesi Gereken Süreçler

1. Maddeler özgün bir numara ile saklanmalıdır.
2. Maddelerin içeriğine ve bilişsel düzeyine uygun sınıflamalar yapılmalıdır.
3. Maddelerin uygulamaya ilişkin bilgileri (uygulanma sayısı, tarih bilgisi, KTK veya MTK ya göre madde istatistikleri) sisteme girilmelidir.
4. Aynı formda kullanılmaması gereken maddeler varsa bununla ilgili bilgiler sisteme kaydedilmelidir.

Madde bankasının güvenliğinin sağlanması son derece önemlidir. Vale (2006), madde bankasında güvenlik açısından üç temel tehdit belirtmiştir: maddelerin izinsiz ele geçirilmesi, kaybolması ve açığa çıkması. Bu tehditleri önleyebilmek için madde bankasına kontrollü erişim sağlanması ve şifreleme tekniği kullanılması önerilmektedir. Kontrollü erişimde şifre ya da giriş kartı kullanılabileceği gibi ses, parmak okuma, yüz tanıma gibi sistemler devreye sokulabilir. Şifrelemede özel algoritmaların kullanıldığı şifreleme standartları tercih edilmelidir.

2.6.6. Nihai Formun Oluşturulması

Testin amacına ve özelliklerine uygun formları oluşturabilmek için madde bankasından madde istatistiklerine dayalı seçimler gerçekleştirilmelidir. Testin içerik özelliklerinin belirlenmesi aşamasında oluşturulan belirtke tablosu veya konu soru dağılım tablosundan yararlanılıp deneme uygulaması sonucunda elde edilen madde istatistikleri doğrultusunda amaca en uygun maddeler seçilir. Test için öngörülen ortalama güçlük düzeyi hesaba katılarak ilgili konu içeriği ve bilişsel düzeyde yer alan maddeler arasından uygun güçlük indeksine sahip olanlar seçilebilir. Bu aşamada madde güçlük değeri ile birlikte madde ayırt edicilik değeri de dikkate alınmalıdır. Aynı özelliği ölçen, benzer güçlük düzeyine sahip maddelerden madde ayırt edicilik değeri en yüksek olan maddelerin nihai forma seçilmesi önerilebilir.

Bazı maddelerin farklı testin farklı zamanlardaki uygulamalarda birden fazla kez kullanılması söz konusu olabilir. Madde bankasında yer alan maddelerin daha önceki uygulamalarda elde edilmiş madde istatistikleri de bu aşamada ayrıca değerlendirilmelidir. Maddelerin kullanım sıklığının artması maddelerin tanınmasına veya bireylerin maddeye maruz kalma sıklığının artmasına da yol açabileceği göz önünde bulundurulmalıdır. Sonuç olarak, madde bankasından madde istatistikleri yer alan maddelerin teste seçilmesi ile birlikte nihai formun test ve madde istatistikleri de belirli bir hata payı ile kestirilebilir.

Nihai form oluşturulurken, deneme formu hazırlanırken dikkate alınması gereken tüm hususlar göz önünde bulundurulmalıdır. Bu aşamada da maddeler uygun punto seçilerek yazılmalı, okunabilirliği açısından kontrol edilmelidir, grafikler ve şekiller sayfa düzenine uygun olarak düzenlenmelidir. Test formunda öğrencilere işlem yapabilmeleri için uygun boşluklar bırakılmalıdır. Maddelerin test formunda kolaydan zora sıralanmasına ve içerik olarak benzer olanların bir arada verilmesine dikkat edilmelidir.

Nihai formun oluşturulmasında farklı gruplar için gerekli uyarlamalar da yapılmalıdır. Örneğin görme engelli bireyler için büyük puntoyla veya Braille ile yazılmış kitapçıklar ya da ekran okuyucular (AERA, APA & NCME, 2014) kullanılabilir. Ya da okuyucu / okuyucu-işaretleyici isteyen görme engelli adaylar şekil, grafik, tablo içeren maddelerden muaf tutulabilirler.

Testlerin farklı formlarının oluşturulması söz konusu ise formlarda da maddelerin yerleşim düzeninin benzerliği sağlanmalıdır. Maddelerin sadece yerlerinin değiştirilmesi ile oluşturulan farklı kitapçıkları söz konusu ise madde yazımında geçerli olan kurallar yine göz önüne alınarak oluşturulmalıdır. Örneğin matematik testinde maddelerin seçeneklerinde yer alan sayılar büyükten küçüğe ya da küçükten büyüğe sıralanmalıdır. Testin farklı kitapçıklarında maddelerin sırasının değişmesinde, seçeneklerde madde yazım ilkeleri doğrultusunda değişiklik gerçekleştirilmemesi uygun olacaktır. Ayrıca, testteki maddelerin kolaydan zora doğru sıralanması öğrenci motivasyonu açısından önemlidir. Testlerin farklı formlarının oluşturulmasında da bu durum göz önünde bulundurulmalıdır. Aksi takdirde farklı formların alınması öğrencilerin puanlarına hata varyans kaynağı olarak karışabilir.

Nihai test formunun geliştirilmesi ile birlikte test geliştirme süreci sonlanmıştır. Test geliştirme süreci görüldüğü üzere her bir aşaması planlanan ve ilgili planın titizlikle işletildiği süreçlerden oluşmaktadır. Test geliştirme sürecinden bir takım uygulamalar söz konusudur. Bölümde bu uygulamalar da kısaca sunulmuştur.

3. TESTİN GELİŞTİRİLMESİ SONRASINDA GERÇEKLEŞTİRİLECEK UYGULAMALAR

3.1. Geçme Puanlarının / Testin Standartlarının Belirlenmesi

Tüm testler için zorunlu olmamakla beraber testlerde geçme puanlarının veya test puanlarının belirli düzeylerini belirlemeye yönelik kesme puanlarının belirlenmesi süreçleri yürütülebilir. Yüksek riskli sınavlarda savunulabilir kesme puanlarının oluşturulması geçerlik için kritik öneme sahiptir (Downing, 2006). Cizek'e (2006) göre, kesme puanlarının oluşturulma süreci standart belirleme olarak ifade edilmekte olup testin geliştirilmesi, uygulama ve raporlama süreçlerinde düşünülecek önemli bir aşamadır. Hatta test geliştirmede madde veya görevlerin belirlenme süreçlerinden itibaren dikkate alınması uygun kanıtların toplanması açısından da fayda sağlayabilir.

Dört dil becerisine ilişkin yeterlik düzeylerinin tanımlanmak istediği bir İngilizce testi için standart belirleme çalışmaları gerçekleştirilebilir. Burada testin amacına ve uygulanan gruba dayalı olarak düzeyler belirlenebilir. Uygun standart belirleme yöntemine karar verildikten sonra, testin deneme uygulamasının gerçekleştirildiği gruptan farklı ve daha geniş bir örneklemde veri toplanarak uygun analizler gerçekleştirilebilir. Standartların belirlenmesine yönelik olarak ayrıntılı bilgi için kitabın 16. bölümü incelenebilir.

3.2. Testin Uygulanması

Nihai testin uygulanmasında deneme uygulamasında var olan süreçlerin çoğunluğu geçerli olmakla birlikte birtakım farklılar söz konusudur. Deneme uygulaması ile nihai uygulamanın gerçekleştirileceği gruplar özellik bakımından eş değerdir, ancak grup büyüklüğü ile uygulanan madde sayıları farklılaşır. Ayrıca geliştirilen testin özel gereksinimli bireyleri de kapsayacak şekilde planlanması gerekiyorsa, testin bu aşamasında gerekli düzenlemelerin yapılması gerekmektedir. Buna göre nihai uygulamada gözetilmesi gereken hususlar aşağıdaki gibi sıralanabilir:

1. Asıl uygulama deneme uygulamasına kıyasla daha geniş bir grup üzerinde yürütüleceğinden görevli ve gözetmen sayıları güncellenmelidir. Deneme uygulamasında yaşanan sorunlar varsa asıl uygulama öncesinde bu sorunlarla ilgili görevlilere bilgi veya eğitim verilmelidir.
2. Deneme uygulamasında materyal veya ekipmandan kaynaklı sorun yaşandıysa nihai uygulamada bunlara ilişkin düzenlemeler gerçekleştirilmelidir.
3. Gerçek uygulamada sınavın yürütüleceği salon veya merkezlerin sayısı artacağından güvenlik zinciri genişletilip deneme uygulamasına kıyasla daha ciddi işletilmelidir.
4. Sınav sırasında uygulayıcı kurum veya gözetmen tarafından herhangi bir hata yapılması halinde, bu hataları telafi etmek için nelerin yapılacağı önceden planlanmalıdır. Örneğin sınavın adaya eşdeğer bir form ile tekrar uygulanması gibi.
5. Formlarda baskıdan kaynaklı hataların olması ihtimaline karşı sınavların gerçekleştirildiği binalarda yedek kitapçıklar ve cevap kâğıtları bulundurulmalıdır.
6. Sınav esnasında gerçekleşen herhangi bir olumsuzluk karşısında (bilgisayarda yürütülen bir sınavda elektrik kesintisi meydana gelmesi gibi) zorunlu olarak sınava ara verildiyse yaşanan süre kaybı sınav süresine eklenmelidir.

Sınava girecek adaylar ile uygulama öncesinde sınav uygulama kılavuzu/formu basılı veya elektronik ortamda paylaşılabilir. Formda sınav başvuru koşulları, başvuru süreci, sınav uygulamasında dikkat edilmesi gereken durumlar ve sınavın nasıl değerlendirileceğine ilişkin bilgilere ayrıntılı olarak yer verilebilir. Ayrıca adayların uyması gereken kurallar da kılavuzda belirtilmelidir. Sınava girmeden önce, sınav süresi ve sınavda dikkat edilmesi gereken hususlar hakkında adayların bilgilendirilmesi sınav sonuçlarına karışabilecek hata varyanslarının azaltılmasına katkı sağlar.

Sınav uygulama kılavuzunda adayların sınav binasına getirmeleri yasak olan madde/araç/cihaz (hesap makinası, kulaklık, yiyecek-içecek vb.) bilgisi yer almalıdır. Sınav binası ya da sınav salonuna hangi zaman aralığında adayların kabul edileceği bilgisi, sınav esnasında uyulması gereken kurallar (sınav esnasında konuşmama, yardımlaşmama, salon dışına çıkmama vb.) da adaylar ile paylaşılmalıdır. Sınav sonunda adayın sınav salonunu terk etmeden önce uyması gereken kurallarda yine bu kılavuzda sunulmalıdır (bkz. Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı [ÖSYM], 2024).

3.3. Testin Sonuçlarının İlanı

Testin uygulanmasından sonra testin planında belirlenmiş tarih aralığında sonuçlar birey veya kurumlara sonuç raporu ile birlikte ilan edilir. Test raporu genel olarak beş bileşene sahiptir: gönderici, mesaj, araç, amaç ve hedef kitle (Ryan, 2006). Gönderici testin uygulama ve yönetiminden sorumlu kurum ya da kuruluştur. Mesaj sınav ile iletilmek istenilen bilgilerdir. Bir test raporunda sadece sayısal veriler değil, basit ve anlaşılır olarak testin neyi kapsadığı, puanların ne anlama geldiği, kesinlik derecesi, test puanlarının yorumlanması ile ilgili yanlış anlamalar ve puanların nasıl kullanılacağının da belirtilmesi gerekmektedir (AERA, APA & NCME, 2014). Araç ise mesajın iletilme biçimidir. Mesaj basılı olarak ya da mobil uygulamada iletilebilir. Amaç ise raporun hangi gaye ile (örneğin, bireyleri seviyeleri hakkında bilgilendirmek) paylaşıldığını belirtir. Hedef kitle ise öğrenci, aile, okul yönetimi, kuruluşlar vb. olabilir.

Test sonuçlarının geçerliği hakkında yeterli bilginin sonuç raporunda yer almasına özen gösterilmelidir. Bu nedenle raporda test sonuçlarının nasıl yorumlanacağı belirtilmeli ve test puanlarına dayalı olarak hangi çıkarımlarda bulunmanın uygun olmayacağına ilişkin bilgi verilmelidir. Raporda yer alan bilgilerin ne kadar ayrıntılı olması gerektiği testin türü ile yakından ilişkilidir. Adaylar bir sertifika sınavı aldıklarında sonuçlar daha çok başarı durumunu belirlemeye yönelik olabilir. Testteki doğru ve yanlış sayıları ile sınav puanı öğrenciye raporlanabilir. Adayların bir yabancı dil sınavına katılmaları durumunda ise sonuçlar her bir beceri bağlamında raporlanabilir. Bireyler hakkında geçti-kaldı kararı verilirken tek bir kesme noktasının mı yoksa birden fazla mı kesme noktasının kullanıldığı raporda açıklanmalıdır. Ağırlıklı ortalama ya da alt ölçekler için ayrı ayrı puanlamanın gerçekleştirildiği bilgisi de rapora dâhil edilmelidir (Downing, 2006).

3.4. Testin Teknik Raporunun Hazırlanması

Testin uygulanmasından sonra gerçekleştirilecek son aşama ise test teknik raporunun hazırlanmasıdır. Test teknik raporu test geliştiriciler tarafından hazırlanan ve test geliştirme sürecinde gerçekleştirilen uygulamaların, ulaşılan sonuçların, geçerlik ve güvenirlik kanıtlarıyla beraber açık, net ve objektif olarak ele alındığı bir dokümandır. Teknik rapor uzmanlardan oluşan küçük bir grup tarafından hazırlanabilir. Raporda sınavın kaç merkezde yapıldığı, testi alan grubun büyüklüğü, katılımcıların betimsel

özellikleri, gruplar arası farklılıklara yönelik analizler, test ve madde düzeyinde istatistikler yer almalıdır. Raporda uygulanan test için öngörülen özellikler ile elde edilen bulguların ne derece örtüştüğüne dair bilgiler de sunulmalıdır. Rapor hazırlanırken anlaşılırlığı artırmak için uygun grafik ve şekillerden yararlanılabilir. Testte yer alan madde türüne göre farklılıklar gösterse de raporun aşağıdaki bilgileri içermesi beklenir (AERA, APA & NCME, 2014; Becker & Pomplun, 2006):

- ⇨ Değerlendirmede kullanılan ölçütler,
- ⇨ Norm çalışma süreçleri ve normlar,
- ⇨ Eşitleme ve ölçekleme sonuçları,
- ⇨ Boş veya hatalı girilen maddelerin nasıl değerlendirildiğine ilişkin açıklamalar,
- ⇨ Açık uçlu maddelerin puanlanmasına görev alan puanlayıcı sayısı,
- ⇨ Puanlayıcılar arası güvenirliğe yönelik analizler,
- ⇨ Puanlayıcılar arasında gözlenen farkların nasıl çözümlendiğine yönelik bilgiler,
- ⇨ Puanlama araçlarının geliştirilme süreci,
- ⇨ Ek geçerlik ve güvenirlik kanıtları,
- ⇨ Sonuçların bireylere veya kurumlara raporlanma biçimi,
- ⇨ Sınav güvenliği ile ilgili önlemler.

4. SONUÇ

Test geliştirme süreçleri literatürde farklı aşamalarda ele alınabilse de bu bölümde sekiz aşamada sunulmuştur. Her bir aşama bir sonraki aşama için ön koşul niteliğinde kabul edilebilir. Bölümde test geliştirme sürecinin ikinci aşamasını oluşturan test özelliklerinin belirlenmesi özellikle ayrıntılandırılmıştır. Bu aşamanın testin yol haritası olarak kabul edilebilecek bir plan niteliğinde sürecin ilk aşamalarında detaylı olarak verilmesi test geliştiriciler için önemli görülmektedir. Deneme uygulamasının tüm testlerde güvenlik gibi nedenler ile gerçekleştirilmesinin mümkün olamayacağına içerikte değinilmiştir. Ek olarak, madde istatistiklerinin makine öğrenmesi yöntemi ile kestirilebileceğinden veya testlerde puanlamaya dahil edilmeyen ancak istatistiklerin kestirilmesi amacıyla uygulanan maddelerin kullanılabileceği bilgisine de yer verilmiştir. Bölümde uygulama sürecinde olası hata kaynaklarından ve çözüm önerilerinden söz edilmiştir, bunun test geliştiricilere katkı sağlayacağı düşünülmektedir. Test formunun oluşturulması ile sonlandırılan test geliştirme sürecinden sonra gelen testin uygulaması aşamasında deneme uygulamasından farklı olan gerçekleştirilen işlemler de sunulmuştur. Bölümde ayrıca test geliştirici kurumun hazırlayacağı testin teknik raporuna yer verilmiş ve raporun içermesi beklenen bilgiler psikometrik özellikler ve sınav güvenliği kapsamında ele alınmıştır.

KAYNAKÇA | REFERENCES

Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory.* Brooks/Cole.

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

Baker, F. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland.

Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. ÖSYM.

Becker, D. F. & Pomplun, M.R. (2006). Technical reporting and documentation. S. M. Downing & T. M. Haladyna (Ed.), *Handbook of test development* (s. 711–723) içinde. Lawrence Erlbaum Associates.

Camili, G. & Shepard, L. A. (1994). *Methods for identifying biased test items*. Sage.

Cizek, G. J. (2006). Standart Setting. S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (s. 225–258) içinde. Lawrence Erlbaum Associates.

Cohen, R. J. & Swerdlik, M. E. (2018). *Psychological testing and assessment: An introduction to tests and measurement* (9. baskı.). McGraw-Hill Education.

College Board. (2022). *Digital SAT suite of assessments specifications overview*. https://satsuite.collegeboard.org/media/pdf/digital-sat-test-spec-overview.pdf.

Crocker, L. & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.

Çetin, B. (2019). Test geliştirme. B. Çetin (Ed.), *Eğitimde ölçme ve değerlendirme* (s. 103-126) içinde. Anı.

Doğan, N. (2019). İnsan davranışları ve ölçme araçlarını sınıflama. N. Doğan (Ed.), *Eğitimde ölçme ve değerlendirme* (s. 75-112) içinde. Pegem Akademi.

Downing, S. M. (2006). Twelve steps for effective test development. S. M. Downing & T. M. Haladyna (Ed.), *Handbook of test development* (s. 3–25) içinde. Lawrence Erlbaum Associates.

Fulari, R. & Rusert, J. (2024,). Utilizing machine learning to predict question difficulty and response time for enhanced test construction. *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications* (*BEA 2024)* (s. 528–533) içinde. Association for Computational Linguistics.

Erkuş, A. (2019). *Psikolojide ölçme ve ölçek geliştirme I: Temel kavramlar ve işlemler* (4. baskı). Pegem Akademi.

Furr, R. M. (2022). *Psychometrics: An introduction* (4. baskı). Sage.

Grounlund, N. E. (1977). *Constructing achievement test* (2. baskı). Prentice-Hall.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3. baskı). Lawrence Erlbaum Associates.

Haladyna, T. M., Downing, S. M. & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309–334. https://doi.org/10.1207/S15324818AME1503_5

Haladyna, T. M. & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.

Hambleton, R. K. (1984). Validating the test scores. R. A. Berk (Ed.), *A guide to criterion- referenced test construction* (s. 199–230) içinde. Johns Hopkins University.

Impara, J. C. & Foster, D. (2006). Item and test development strategies on minimize test fraud. S. M. Downing & T. M. Haladyna (Ed.), *Handbook of test development* (s. 91–114) içinde. Lawrence Erlbaum Associates.

Kane, M. T. (2006). Content-related validity evidence in test development. S. M. Downing & T. M. Haladyna (Ed.), *Handbook of test development* (s. 131–153) içinde. Lawrence Erlbaum Associates.

Kline, T. J. P. (2005). *Psychological testing: A practical approach to design and evaluation* (2. baskı). Sage.

LaDuca, A., Downing, S. M. &. Henzel, T. R. (1995). Test development: Systematic item writing and test construction. J. C. Impara & J. C. Fortune (Ed.), *Licensure examinations: Purposes, procedures, and practices* (s. 117-148) içinde. Buros Institute of Mental Measurements.

McCallin, R. C. (2006). Test administration. S. M. Downing & T. M. Haladyna (Ed.), *Handbook of test development* (s. 625–652) içinde. Lawrence Erlbaum Associates.

Mehrens, W. A. & Lehmann, I. J. (1991). Measurement and evaluation in education and psychology (4. baskı). Wadsworth.

Millî Eğitim Bakanlığı. Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü. (2025). *Konu soru dağılım tabloları*. https://odsgm.meb.gov.tr/www/1-donem-konu-soru-dagilim-tablolari-2025-2026/icerik/1474

Murphy, K. R. & Davidshofer, C. O. (2001). *Psychological testing: Principles and applications* (6. baskı). Pearson/Prentice Hall.

Osterlind, S. J. (2002). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats* (2. baskı). Kluwer Academic.

Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı. (2024). *2024-Yükseköğretim Kurumları Sınavı (YKS) Kılavuzu.* https://dokuman.osym.gov.tr/pdfdokuman/2024/YKS/kilavuz_d23052024.pdf

Özçelik, D. (1992). *Ölçme ve değerlendirme*. ÖSYM.

Popham, W. J. (2025). *Classroom assessment: What teachers need to know* (10. baskı). Pearson.

Thompson, S., Johnston, C. J. & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). National Center on Educational Outcomes, University of Minnesota.

Turgut, M. F. (1992). *Eğitimde ölçme ve değerlendirme*. Saydam.

Raymond, M. R. & Neustel, S. (2006). Determining the content of credentialing examination. S. M. Downing & T. M. Haladyna (Ed.), *Handbook of test development* (s. 181–223) içinde. Lawrence Erlbaum Associates.

Razavi, P. & Powers, S. J. (2025). *Estimating item difficulty using large language models and tree-based machine learning algorithms*. https://doi.org/10.48550/arXiv.2504.08804

Rodriguez, M. C. (2002). Choosing an item format. G. Tindal & T. M. Haladyna (Eds.), *Large scale assessment programs for all students* (s. 213-231) içinde. Lawrence Erlbaum Associates.

Ryan, J. M. (2006). Practices, issues, and trends in student test score reporting. S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (s. 677–710) içinde. Lawrence Erlbaum Associates.

Russell, M. K. & Airasian, P. W. (2012). *Classroom assessment: Concepts and applications* (7. baskı). McGraw-Hill.

Skaggs, G. (2023). *Test development and validation*. Sage.

Yousefpoori-Naeim, M., Zargari, S. & Hatami, Z. (2024). *Using machine learning to predict item difficulty and response time in medical tests. Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)* (s. 551–560) içinde. Association for Computational Linguistics.

URL-1. Millî Eğitim Bakanlığı (MEB). (2025). *Fen Bilimleri 8. Sınıf Konu Soru Kitapçığı*. https://cdn.eba.gov.tr/yardimcikaynaklar/2025/09/konusoru/fen8.pdf

URL-2. Smarter Content Explorer. *Test Development & Design.* https://contentexplorer.smarterbalanced.org/test-development#blueprints

Vale, C. D. (2006). Computerized item banking. S. M. Downing & T. M. Haladyna (Ed.), *Handbook of test development* (s. 261–285) içinde. Lawrence Erlbaum Associates.

Zumbo, D. B. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type item scores.* Ottowa: Directorate of Human Resources Research and Evaluation, Department of National Defense. https://faculty.educ.ubc.ca/zumbo/DIF/handbook.pdf

YAZAR HAKKINDA | ABOUT AUTHOR

PROF. DR. MELEK GÜLŞAH ŞAHİN

Gazi Üniversitesi | mgulsahsahin[at]gazi.edu.tr

ORCID: 0000-0001-5139-9777

Prof. Dr. Melek Gülşah Şahin, Gazi Üniversitesi Gazi Eğitim Fakültesi Eğitim Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme alanında öğretim üyesidir. Lisans eğitimini Hacettepe Üniversitesi, Eğitim Fakültesi, Orta Öğretim Fen ve Matematik Alanları Bölümü, Fizik Öğretmenliği Anabilim Dalı'nda 2006 yılında tamamlamıştır. Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı'nda 2008 yılında yüksek lisans; 2014 yılında ise doktora derecesini almıştır. 2019 yılında Eğitimde Ölçme ve Değerlendirme Alanında doçent ünvanını almıştır. Dr. Şahin'in lisans ve lisansüstü düzeyde yürüttüğü dersler ve tez danışmanlıklarına ek olarak, ulusal ve uluslararası düzeyde çok sayıda yayını ile projelerde görevleri bulunmaktadır. Bilgisayar ortamında bireye uyarlanmış test, çok aşamalı testler, ölçme aracı geliştirme konuları çalışmaları alanları arasında yer alır.

PROF. DR. MELEK GÜLŞAH ŞAHİN

Gazi University| mgulsahsahin[at]gazi.edu.tr

ORCID: 0000-0001-5139-9777

Prof. Dr. Melek Gülşah Şahin is a faculty member in the Division of Educational Measurement and Evaluation, Department of Educational Sciences, Gazi Faculty of Education, Gazi University. She completed, earned her bachelor's degree in Physics Education from the Department of Secondary Science and Mathematics Education, Faculty of Education, Hacettepe University, in 2006. She received her master's degree in 2008 and her Ph.D. in 2014 from the Division of Educational Measurement and Evaluation, Department of Educational Sciences, Graduate School of Educational Sciences, Hacettepe University. In 2019, she was promoted to Associate Professor of Educational Measurement and Evaluation. In addition to teaching undergraduate and graduate courses and supervising graduate theses, Dr. Şahin has numerous national and international publications and has participated in various projects. Her research interests include computerized adaptive testing, computerized multistage testing, and assessment tools development.