

Puanlayıcı Yargılarına Dayalı Ölçmelerde Geçerlik ve Güvenirlilik

Mustafa İLHAN

Neşe GÜLER

Gülşen TAŞDELEN TEKER



Mustafa İLHAN - Neşe GÜLER - Gülşen TAŞDELEN TEKER

PUANLAYICI YARGILARINA DAYALI ÖLÇMELERDE GEÇERLİK VE GÜVENİRLİK

ISBN 978-625-6287-46-4

Kitap içeriğinin tüm sorumluluğu yazarlarına aittir.

© 2024, PEGEM AKADEMİ

Bu kitabın basım, yayım ve satış hakları Pegem Akademi Yay. Eğt. Dan. Hizm. Tic. AŞ'ye aittir. Anılan kuruluşun izni alınmadan kitabın tümü ya da bölümleri, kapak tasarımı; mekanik, elektronik, fotokopi, manyetik kayıt ya da başka yöntemlerle çoğaltılamaz, basılamaz ve dağıtılamaz. Bu kitap, T.C. Kültür ve Turizm Bakanlığı bandrolü ile satılmaktadır. Okuyucularımızın bandrolü olmayan kitaplar hakkında yayınevimize bilgi vermesini ve bandrolsüz yayınları satın almamasını diliyoruz.

Pegem Akademi Yayıncılık, 1998 yılından bugüne uluslararası düzeyde düzenli faaliyet yürüten **uluslararası akademik bir yayınevidir**. Yayımladığı kitaplar; Yükseköğretim Kurulunca tanınan yükseköğretim kurumlarının kataloglarında yer almaktadır. Dünyadaki en büyük çevrimiçi kamu erişim kataloğu olan **WorldCat** ve ayrıca Türkiye'de kurulan **Turcademy.com** tarafından yayınları taranmaktadır, indekslenmektedir. Aynı alanda farklı yazarlara ait 2000'in üzerinde yayını bulunmaktadır. Pegem Akademi Yayınları ile ilgili detaylı bilgilere <http://pegem.net> adresinden ulaşılabilir.

I. Baskı: Temmuz 2024, Ankara

Yayın-Proje: Selcan Durmuş
Dizgi-Grafik Tasarım: Tuğba Kaplan
Kapak Tasarımı: Pegem Akademi

Baskı: Sonçağ Yayıncılık Matbaacılık Reklam San Tic. Ltd. Şti.
İstanbul Cad. İstanbul Çarşısı 48/48 İskitler - Ankara
Tel: (0312) 341 36 67

Yayıncı Sertifika No: 51818
Matbaa Sertifika No: 47865

İletişim

Macun Mah. 204. Cad. No: 141/A-33 Yenimahalle/ANKARA
Yayınevi: 0312 430 67 50
Dağıtım: 0312 434 54 24
Hazırlık Kursları: 0312 419 05 60
İnternet: www.pegem.net
E-ileti: pegem@pegem.net
WhatsApp Hattı: 0538 594 92 40

ÖN SÖZ

Ölçme sonuçlarına dayanarak bireyler hakkında bir karara varmadan önce ölçümlerin psikometrik özellikleri ortaya konulmalıdır. Bu amaçla sunulması gereken kanıtlar ölçme aracının puanlanışına göre farklılaşır. Sübjektif puanlanan ölçme araçlarında, ölçme sonuçlarını puanlayıcı hataları açısından da incelemek ve puanlayıcı güvenilirliğini test etmek gerekir. Eğitim bilimlerinden sağlık bilimine, spor bilimlerinden güzel sanatlara kadar birçok disiplinde puanlayıcı yargılarına dayalı ölçmelere başvurulmaktadır. Dolayısıyla farklı alanlardan araştırmacılar/eğitimciler, puanlayıcı yargılarına dayalı ölçmelerde geçerliği ve güvenilirliği konu edinen kaynaklara ihtiyaç duymaktadır. Ancak Türkçe alanyazın gözden geçirildiğinde bu ihtiyaca cevap olabilecek kaynak sayısının sınırlı olduğu, çoğunlukla objektif puanlanan ölçme araçlarından elde edilen verilerin analizine odaklanıldığı görülmektedir. Buna bağlı olarak Türkçe alanyazına puanlayıcı yargılarına dayalı ölçmelerde geçerlik ve güvenilirliği kapsamlı bir şekilde ele alan kaynaklar kazandırmanın önemli olduğu düşünülmüş, bu düşünce kitabın hazırlanmasında motive edici güç olmuştur.

Kitapta sübjektif nitelik taşıyan ölçmelerde geçerlik ile güvenirliliğin testine yönelik işlemler farklı ölçme kuramlarına göre teorik temelleri ve paket programlardaki uygulamalarıyla birlikte sunulmuştur. Kitabın ilk iki bölümünde sırasıyla puanlayıcı hataları ve puanlayıcı hatalarını azaltmaya yönelik önlemler üzerinde durulmuştur. Bunu takiben üçüncü bölümde klasik test kuramına, dördüncü bölümde genellenebilirlik kuramına ve beşinci bölümde çok yüzeyli Rasch modeline göre puanlayıcı yargılarına dayalı ölçmelerin geçerlik ve güvenirliliğini test etmeye yönelik uygulamalara yer verilmiştir. Yazarlar kitabı farklı disiplinlerden okuyuculara hitap edebilecek açıklıkta bir içerik sunma gayretiyle yapılandırmıştır. Tek bir alana yönelik örnekler ve uygulamalar sunmaktan kaçınmış, alanı ne olursa olsun kitabı eline alan her bir okuyucunun kendi çalışma alanıyla kolayca bağdaştırabileceği örnekler vermeye özen göstermişlerdir. Tabi buna rağmen kitabın gözden kaçan, geliştirilmesi gereken yönleri bulunabilir. İlerleyen basımlarda kitabın geliştirilebilmesi okuyuculardan gelen geribildirimlerle mümkün olacaktır. Bu anlamda, okuyuculardan kitap hakkındaki görüş ve önerilerini yazarlara iletmelerini rica eder, kitabın basımında emekleri olan PEGEM AKADEMİ ailesine teşekkürlerimizi sunarız. Eğitimcilere ve araştırmacılara katkı sağlaması dileğiyle...

Mustafa İLHAN
Neşe GÜLER
Gülşen TAŞDELEN TEKER

*Kitaptaki örnek uygulamalarda
kullanılan verilere aşağıdaki
karekodu okutarak ulaşabilirsiniz.*



YAZARLAR HAKKINDA

Doç. Dr. Mustafa İLHAN

Lisans eğitimini 2009 yılında Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi, İlköğretim Matematik Öğretmenliği Bölümünde tamamlamıştır. 2011 yılında Dicle Üniversitesi, Sosyal Bilimler Enstitüsü, Eğitim Bilimleri Anabilim Dalı, Eğitim Programları ve Öğretim Bilim Dalı'nda yüksek lisans derecesini; 2015 yılında Gaziantep Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitim Bilimleri Anabilim Dalı'nda doktora derecesini ve 2020 yılında Eğitimde Ölçme ve Değerlendirme alanında doçent unvanını almıştır. 2010 ile 2016 yılları arasında Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi, Matematik ve Fen Bilimleri Eğitimi Bölümü, Matematik Eğitimi Anabilim Dalı'nda araştırma görevlisi olarak çalışmıştır ve 2016 yılından bu yana aynı bölümde öğretim üyesi olarak görev yapmaktadır.

ORCID No: 0000-0003-1804-002X

Prof. Dr. Neşe GÜLER

Lisans eğitimini 1995 yılında Uludağ Üniversitesi, Necatibey Eğitim Fakültesi, Matematik Bölümünde tamamlamış, aynı yıl yüksek lisans sınavını kazanarak bölümünde yüksek lisans eğitimine ve araştırma görevlisi olarak çalışmaya başlamıştır. Matematik bölümünde yüksek lisans eğitimini bitirdikten sonra Amerika'da lisansüstü eğitim almak üzere Milli Eğitim Bakanlığı'nın yurt dışı bursunu kazanmıştır. Florida Üniversitesi'nde Eğitimde Ölçme ve Değerlendirme dalında yüksek lisansını tamamlayarak 2003 yılında Türkiye'ye dönmüş ve Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim dalında doktora eğitimine başlamıştır. 2008 yılında doktora derecesini, 2014 yılında doçent ve 2019 yılında profesör unvanını almıştır. 2003–2008 yılları arasında Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim dalında araştırma görevlisi olarak çalışmış, 2008–2018 yılları arasında Sakarya Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim dalında öğretim üyesi olarak görev yapmıştır. 2018 yılı Nisan ayından bu yana ise İzmir Demokrasi Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim dalında öğretim üyesi olarak görevini sürdürmektedir.

ORCID No: 0000-0002-2836-3132

Doç. Dr. Gülşen TAŞDELEN TEKER

Lisans eğitimini 2007 yılında Orta Doğu Teknik Üniversitesi, Eğitim Fakültesi, İlköğretim Bölümü, Fen Bilgisi Öğretmenliği Anabilim Dalı'nda tamamlamıştır. Hacettepe Üniversitesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalında 2009 yılında yüksek lisans derecesini, 2014 yılında ise doktora derecesini almıştır. 2021 yılında Eğitimde Ölçme ve Değerlendirme alanında doçent unvanını almıştır. 2010–2017 yılları arasında Sakarya Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalında öğretim görevlisi olarak çalışmıştır; 2017 yılından bu yana ise Hacettepe Üniversitesi Tıp Fakültesi, Tıp Eğitimi ve Bilişimi Anabilim Dalında öğretim üyesi olarak çalışmaktadır.

ORCID No: 0000-0003-3434-4373

İÇİNDEKİLER

Ön Söz.....	iii
Yazarlar Hakkında.....	v

PUANLAYICI YARGILARINA DAYALI ÖLÇMELERE GİRİŞ

Kaynakça	3
----------------	---

1. BÖLÜM

PUANLAYICI HATALAR

Puanlayıcı Katılığı ve Cömertliği (Rater Severity and Leniency)	6
Merkeze Yönelme (Central Tendency)	6
Halo ve Horn Etkisi	7
Ranj Sınırlaması (Range Restriction).....	8
Yanlılık (Rater Bias)	9
Karşılaştırma Etkisi (Contrast Effect).....	9
Kendine Hizmet Etme Yanlılığı (Self-Serving Bias).....	10
Bana Benzer ve Benden Farklı Etkisi (Similar/Dissimilar to me Effect).....	10
İlk İzlenim Etkisi (First Impression Error).....	11
Puanlayıcı Katılığı Kayması (Rater Severity Drift).....	11
Mantıksal Hata (Logical Error)	11
Amigo Etkisi (Cheerleader Effect)	12
Yakınlık (Proximity)	12
Prima Donna Etkisi (Prima Donna Effect).....	13
Sonralık Etkisi (Recency Effect).....	13
Öncelik Etkisi (Primacy Effect).....	13
Kaynakça.....	14

2. BÖLÜM

PUANLAYICI HATALARINI AZALTMAK İÇİN ALINABİLECEK ÖNLEMLER

Puanlamaların Birden Fazla Puanlayıcı Tarafından Yapılması	18
Puanlayıcı Eğitimleri.....	18
Puanlayıcı Hatası Eğitimi	19

Performans Boyutları Eğitimi.....	20
Davranış Gözlem Eğitimi	21
Referans Çerçevesi Eğitimi.....	21
Puanlayıcı Değişkenliği Eğitimi	22
Puanlamaların Puanlama Ölçeklerine Dayalı Olarak Yapılması	23
Kontrol Listesi.....	24
Kontrol Listesinin Üstünlük ve Sınırlılıkları.....	25
Dereceleme Ölçekleri.....	26
Dereceleme Ölçeğinin Üstünlük ve Sınırlılıkları.....	27
Rubrikler (Dereceli Puanlama Anahtarı)	27
Genel ve Göreve Özel Rubrikler.....	28
Holistik ve Analitik Rubrikler.....	29
Rubriklerin Sağladığı Faydalar	32
Rubrik Geliştirme Basamakları	33
Diğer Önlemler.....	34
Kaynakça.....	36

3. BÖLÜM

KLASİK TEST KURAMINA DAYALI GEÇERLİK VE GÜVENİRLİK ANALİZLERİ

Güvenirlik	40
Test Tekrar Test Güvenirliği.....	41
Paralel Formlar Güvenirliği	46
İç Tutarlılık Güvenirliği	47
İki Yarı Güvenirlik Katsayısı	47
Cronbach Alfa, McDonald Omega, Tabakalı Alfa ve Hiyerarşik Omega.....	50
Puanlayıcı Güvenirliği	55
Uyum Yüzdesi	55
Cohen'in Kappası.....	56
Ağırlıklandırılmış Kappa.....	60
Fleiss'in Kappası.....	63
Pearson, Spearman ve Kendall tau Korelasyonu	65
Ortalamaların Karşılaştırılması	68
Sınıf-içi Korelasyon Katsayısı.....	71

Kendall Uyuşum Katsayısı.....	78
Krippendorff Alfa.....	81
Geçerlik.....	84
Kapsama Dayalı Kanıtlar.....	86
Cevaplama Süreçlerine Dayalı Kanıtlar.....	92
Ölçme Aracının Diğer Değişkenlerle İlişisine Dayalı Kanıtlar	92
Zamandaş ve Yordayıcı Kanıtlar	92
Yakınsak ve Iraksak Kanıtlar	93
Karşıt Gruplar Arasındaki Farka Dayalı Kanıtlar	95
Ölçme Sonuçlarına Dayalı Kanıtlar	95
Ölçme Aracının İç Yapısına Dayalı Kanıtlar	96
Faktör Analizi.....	97
Açımlayıcı Faktör Analizi (AFA).....	99
Doğrulayıcı Faktör Analizi (DFA).....	109
Madde Analizi.....	129
Kaynakça.....	132

4. BÖLÜM

GENELLENEBİLİRLİK KURAMINA DAYALI GEÇERLİK VE GÜVENİRLİK ANALİZLERİ

Genellenebilirlik Kuramı.....	143
Hipotetik Ölçme Durumu.....	144
Dengelenmemiş Desenler	172
Kaynakça.....	173

5. BÖLÜM

ÇOK YÜZEYLİ RASCH MODELİNE DAYALI GEÇERLİK VE GÜVENİRLİK ANALİZLERİ

Madde Tepki Kuramı	176
Çok Yüzeyle Rasch Modeli	178
ÇYRM'nin Varsayımları	181
Tek boyutluluk	181
Yerel bağımsızlık	182

Model-veri uyumu.....	182
ÇYRM İstatistikleri	182
Uygulama 5.1	190
Programın Bilgisayara Kurulumu	190
Veri Dosyasının Yapısı	190
Analize Hazırlık	192
Analizin Uygulanması	198
Kaynakça.....	216

EK: JASP'IN KURULUMU VE PROGRAM HAKKINDA TEMEL BİLGİLER

JASP'ın Kurulumu	220
JASP'a Veri Aktarımı.....	221
JASP'a Girilen Verilerin Düzenlenmesi.....	223
JASP'ta Değişken Tanımlama ve Filtreleme.....	224
JASP'a Hesaplanmış Sütun Ekleme	227
JASP'ta Analiz Yaparken Ekrana Gelen Arayüz	229
Önerilen Okumalar	231

ŞEKİLLER LİSTESİ

Şekil 1.	Puanlayıcı yargılarına dayalı ölçmelerin genel yapısı	2
Şekil 1.1.	Puanlayıcı hatası türleri	5
Şekil 1.2.	Katılık, cömertlik, merkeze yönelme ve halo/horn etkisine ilişkin puanlama örnekleri	8
Şekil 1.3.	Karşılaştırma etkisinin şematik gösterimi.....	10
Şekil 2.1.	Puanlayıcı eğitimi modelleri	19
Şekil 2.2.	Puanlayıcı yargılarına dayalı ölçmelerde kullanılan puanlama ölçekleri.....	23
Şekil 2.3.	Kontrol listesi geliştirme sürecine ilişkin adımlar	25
Şekil 2.4.	Rubriklerin öğretmenler ile öğrencilere sağladığı faydalar.....	32
Şekil 2.5.	Rubrik geliştirme basamakları	33
Şekil 3.1.	JASP'ta test tekrar test güvenirliğinin hesaplanmasına yönelik işlem adımları-I.....	43
Şekil 3.2.	JASP'ta test tekrar test güvenirliğinin hesaplanmasına yönelik işlem adımları-II	44
Şekil 3.3.	SPSS'te test tekrar–test güvenirliğini hesaplamaya yönelik adımlar	45
Şekil 3.4.	SPSS'te iki yarı güvenirliğini hesaplama adımları-I	48
Şekil 3.5.	SPSS'te iki yarı güvenirliğini hesaplama adımları-II.....	49
Şekil 3.6.	JASP programında Cronbach alfa ve McDonald omega katsayılarının hesaplanması-I	53
Şekil 3.7.	JASP programında Cornbach alfa ve McDonald omega katsayılarının hesaplanması-II.....	54
Şekil 3.8.	KTK'da puanlayıcı güvenirliğini belirlemede kullanılabilen yöntemler	55
Şekil 3.9.	JASP programında Cohen'in kappasının hesaplanmasına yönelik adımlar-I	58
Şekil 3.10.	JASP programında Cohen'in kappasının hesaplanmasına yönelik adımlar-II.....	59
Şekil 3.11.	SPSS'te ağırlıklandırılmış kappanın hesaplanmasına yönelik adımlar-I	61
Şekil 3.12.	SPSS'te weighted kappa penceresi ve bu penceredeki criteria ekranı.....	62
Şekil 3.13.	Ağırlıklandırılmış kappa istatistiğinden elde edilen çıktılar.....	63
Şekil 3.14.	Aynı veri seti için hesaplanan Cohen'in ve Fleiss'in kappa istatistikleri	64

Şekil 3.15. SPSS'te Fleiss'in kappasını hesaplarken takip edilmesi gereken adımlar	65
Şekil 3.16. JASP'ta Pearson, Spearman ve Kendall tau korelasyonu için işlem adımları-I.....	66
Şekil 3.17. JASP'ta Pearson, Spearman ve Kendall tau korelasyonu için işlem adımları-II	67
Şekil 3.18. JASP'ta puanlayıcı ortalamalarının karşılaştırılmasına yönelik adımlar-I	69
Şekil 3.19. JASP'ta puanlayıcı ortalamasının karşılaştırılmasına yönelik adımlar-II.....	70
Şekil 3.20. Aynı veri seti için tutarlılık ve mutlak uyum anlamındaki SKK ile Pearson korelasyonu arasındaki ilişkiler	71
Şekil 3.21. SKK için model, tür ve tanımlar	73
Şekil 3.22. Çalışmanın tasarımına bağlı olarak uygun SKK formunun seçimine yönelik akış şeması.....	75
Şekil 3.23. SPSS'te SKK'nın hesaplanması	76
Şekil 3.24. SKK için SPSS'ten elde edilen çıktılar	77
Şekil 3.25. JASP'ta SKK'nın hesaplanması.....	78
Şekil 3.26. JASP'ta Kendall W katsayısının hesaplanmasına yönelik adımlar-I	79
Şekil 3.27. JASP'ta Kendall W katsayısının hesaplanmasına yönelik adımlar-II.....	80
Şekil 3.28. JASP'ta Krippendorff alfa'yı hesaplamaya yönelik adımlar.....	83
Şekil 3.29. Geçerlik kanıtı türleri.....	85
Şekil 3.30. Uzman görüş formu örneği	87
Şekil 3.31. KGO ve KGI'nin hesaplanma aşamaları	91
Şekil 3.32. KMO katsayısı yorumlanırken dikkate alınması önerilen aralıklar	100
Şekil 3.33. Faktör analizi ile PCA arasındaki kavramsal ayrım	102
Şekil 3.34. AFA'da faktörlerin döndürülmesine ilişkin karar ağacı	105
Şekil 3.35. Örnek DFA modeli	110
Şekil 3.36. İstatistiksel süreçler açısından AFA ile DFA arasındaki farklılıkların görsel temsili.....	111
Şekil 3.37. Veri dosyasının .txt uzantılı formata dönüştürülmesi	118
Şekil 3.38. "Factor Analysis for All" arayüzünde veri dosyasının açılması.....	119
Şekil 3.39. "Factor Analysis for All" arayüzünde verilerin düzenlenmesi	120
Şekil 3.40. "Factor Analysis for All" arayüzünde faktör analizi varsayımlarının test edilmesi	122

Şekil 3.41. “Factor Analysis for All” arayüzünde faktör sayısının belirlenmesi	123
Şekil 3.42. “Factor Analysis for All” arayüzünde EFA’ya ilişkin tanımlamaların yapılması.....	124
Şekil 3.43. “Factor Analysis for All” arayüzünde EFA çıktıları.....	125
Şekil 3.44. “Factor Analysis for All” arayüzünde AGA uygulaması.....	126
Şekil 3.45. “Factor Analysis for All” arayüzünde DFA komutlarının tanımlanması.....	127
Şekil 3.46. “Factor Analysis for All” arayüzünde DFA çıktıları.....	128
Şekil 3.47. Madde güçlük ve ayırt edicilik indekslerine ilişkin temel bilgiler.....	129
Şekil 4.1. EduG programında G kuramı analizlerinin yürütülmesine yönelik işlem adımları-Dosyanın açılması.....	151
Şekil 4.2. EduG programında G kuramı analizlerinin yürütmesine yönelik işlem adımları-Çapraz desende G çalışmasının yürütülmesi.....	152
Şekil 4.3. EduG programına aktarılan veri dosyasının kontrolü ve kaydedilmesi.....	153
Şekil 4.4. EduG programı uyarı mesajı.....	154
Şekil 4.5. EduG programı G çalışması çıktısı- Çaprazlanmış desen	155
Şekil 4.6. EduG programında G kuramı analizlerinin yürütülmesine yönelik işlem adımları-Çaprazlanmış desende K çalışmasının yürütülmesi	159
Şekil 4.7. K çalışması sonucu- Çaprazlanmış desen	159
Grafik 4.1. K çalışması sonuçları.....	161
Şekil 4.8. EduG programında G kuramı analizlerinin yürütülmesine yönelik işlem adımları-Yuvalanmış desende G çalışmasının yürütülmesi	163
Şekil 4.9. EduG programında G kuramı analizlerinin yürütülmesine yönelik işlem adımları-Yuvalanmış desende yanlış düzey tanımlanması sonucu alınan hata uyarısı.....	164
Şekil 4.10. EduG programı G çalışması çıktısı- Yuvalanmış desen	165
Şekil 4.11. EduG programında G kuramı analizlerinin yürütülmesine yönelik işlem adımları - Yuvalanmış desende K çalışmasının yürütülmesi	167
Şekil 4.12. K çalışması sonucu - Yuvalanmış desen	168
Şekil 4.13. G çalışması sırasında yüzeyin herhangi bir düzeyini çıkarma	169
Şekil 4.14. G-Facets analizi.....	171
Şekil 4.15. G-Facets analizi çıktıları	172
Şekil 5.1. Madde karakteristik eğrisi örnekleri.....	177

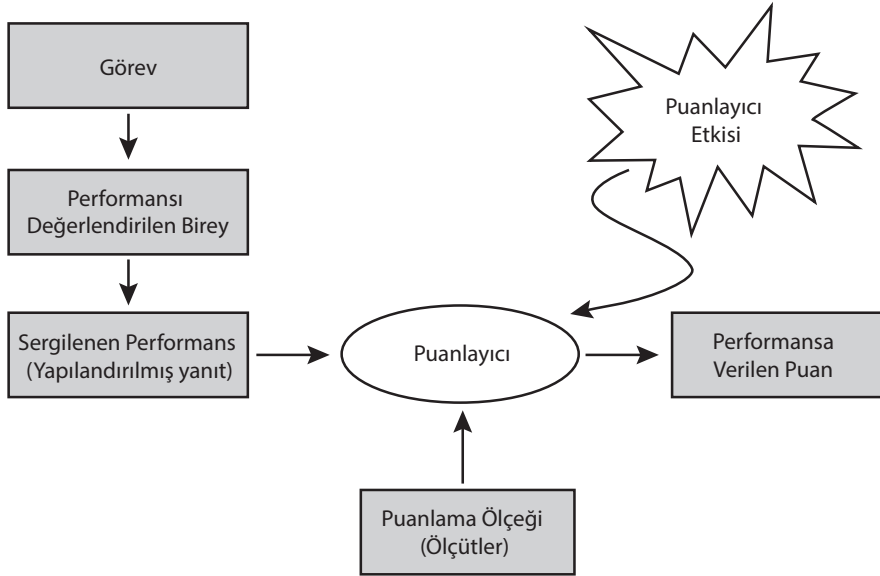
Şekil 5.1.1. İki parametrelili model.....	177
Şekil 5.1.2. Üç parametrelili model	177
Şekil 5.1.3. Dört parametrelili model.....	177
Şekil 5.2. Verilerin Excel dosyasındaki görünümü	191
Şekil 5.3. Verilerin yer aldığı Excel dosyasının metin belgesi olarak kaydedilmesi	192
Şekil 5.4. Analize hazır veri dosyası.....	197
Şekil 5.5. FACETS programında çok yüzeyli Rasch analizi uygulama adımları.....	199
Şekil 5.6. Standart artıkların yer aldığı Excel dosyasının kaydedilmesi.....	200
Şekil 5.7. Örnek uygulama için elde edilen logit harita.....	201
Şekil 5.8. FACETS’de kategori bilgi grafiği	209
Şekil 5.9. FACETS’de yanlılık grafiklerini incelemek için izlenmesi gereken adımlar ve puanlayıcı×görev etkileşimi ait t-değerlerini içeren grafik.....	212
Şekil 5.10. Yanlılık düzeltme uygulamasına ilişkin syntax.....	215
Şekil 6.1. JASP karşılama ekranı	220
Şekil 6.2. JASP’a veri aktarımında izlenen adımlar	221
Şekil 6.3. JASP’ta veri dosyasının görünümü.....	222
Şekil 6.4. JASP’ta değişkenin ölçek düzeyini değiştirme	223
Şekil 6.5. JASP’ta veri düzenleme	223
Şekil 6.6. JASP’ta değişken etiketleme ve tıklama yöntemiyle veri filtreleme-I.....	224
Şekil 6.7. JASP’ta değişken etiketleme ve tıklama yöntemiyle veri filtreleme-II	225
Şekil 6.8. JASP’ta sürük bırak yöntemiyle veri filtreleme-I.....	225
Şekil 6.9. JASP’ta sürük bırak yöntemiyle veri filtreleme-II	226
Şekil 6.10. Hesaplanmış Sütun Oluşturma-I.....	227
Şekil 6.11. Hesaplanmış Sütun Oluşturma-II	228
Şekil 6.12. JASP’ın aerodinamik arayüzü	229

TABLÖLAR LİSTESİ

Tablo 2.1. Konuşma becerisinin değerlendirilmesine yönelik kontrol listesi.....	24
Tablo 2.2. Çocuklarda davranış problemlerini değerlendirmeye yönelik dereceleme ölçeği.....	26
Tablo 2.3. Holistik rubrik örneği	29
Tablo 2.4. Analitik rubrik örneği	30
Tablo 3.1. KTK'da kullanılan güvenilirlik hesaplama yöntemleri.....	40
Tablo 3.2. Kappa istatistiği yorumlanırken başvurulabilecek eşik değerler	57
Tablo 3.3. Puanlayıcı güvenilirliğini incelerken ortalamaları karşılaştırmak için uygulanabilecek testler	68
Tablo 3.4. KGO için kritik değerler	89
Tablo 3.5. KGO için örnek uygulama.....	90
Tablo 3.6. Kültürel ve sosyal zeka verileri için çoklu özellik çoklu yöntem korelasyon katsayıları.....	94
Tablo 3.7. Örtük değişken analizlerinin sınıflandırılması.....	97
Tablo 3.8. Tanımsız, neredeyse tanımlı ve tanımlı model örneği.....	112
Tablo 3.9. Uyum indekslerine ilişkin eşik değerler	115
Tablo 4.1. Örnek veri seti	144
Tablo 4.2. G çalışması sonucu kestirilen varyans değerleri - Çaprazlanmış desen.....	156
Tablo 4.3. ömmp örneğine ilişkin K çalışması sonucu.....	160
Tablo 4.4. Örnek veri seti (yuvalanmış desen)	162
Tablo 4.5. G çalışması sonucu kestirilen varyans değerleri - Yuvalanmış desen.....	166
Tablo 5.1. ÇYRM'nin çeşitli puanlayıcı hataları için sunduğu istatistiksel göstergeler.....	188
Tablo 5.2. ÇYRM analizleri için syntax'da yer alan temel komutlar ve anlamları	193
Tablo 5.3. Birey yüzeyine ait ölçüm raporu.....	203
Tablo 5.4. Puanlayıcı yüzeyine ait ölçüm raporu.....	205
Tablo 5.5. Görev yüzeyine ait ölçüm raporu	207
Tablo 5.6. Kategori istatistikleri	208
Tablo 5.7. Beklenmedik yanıtlar	210
Tablo 5.8. Yanlılık/Etkileşim raporu.....	211
Tablo Ek 1. Analiz komutlarının uygulandığı paneldeki ikonlar ve anlamları	230

PUANLAYICI YARGILARINA DAYALI ÖLÇMELERE GİRİŞ

Puanlayıcının bir puanlama aracı yardımıyla bireylerin bir veya daha fazla alandaki performansını değerlendirmesi puanlayıcı yargılarına dayalı (rater-mediated) ölçme olarak tanımlanır. Değerlendirmede otomatik puanlama sistemleri kullanılsa bile bireylerin performansını puanlayan algoritmaların geliştirilmesine gerçek puanlayıcılar rehberlik eder (Engelhard ve Wind, 2019). Dolayısıyla yine yapılan değerlendirme puanlayıcı yargılarından tamamen arınık olmaz. von Eye ve Mun (2005) gerek akademik gerekse akademik olmayan birçok alanda puanlayıcı yargılarına dayalı ölçmelerin karşımıza çıktığını ifade eder ve puanlayıcılar arası uyumu bu ölçmelerdeki kilit faktör olarak nitelendirir. Bu durumu somullaştırmak için ise şu örneklerden bahseder. Olimpiyat oyunlarında jimnastik, artistik paten, senkronize yüzme gibi birçok alanda sıralamalar ve madalya almaya hak kazanan kişiler puanlayıcı yargılarına dayalı olarak belirlenir. Bu süreçte uç noktalarda kalan puanlayıcılar puanlama havuzundan çıkarıldıktan sonra sıralamalara karar verilir. Tıp alanında hastaya en uygun tedavinin önerildiğinden emin olmak için konulan tanıyı netleştirmeden önce birden fazla doktorun görüşüne başvurulur. Ceza davalarında en doğru kararı verebilmek için birden fazla hâkimin olayı değerlendirmesi istenir ve ancak hâkimler sanığın suçluluğuna yönelik ortak bir görüş beyan ettikleri takdirde sanık cezaya çarptırılır. Gözleme ve içerik analizine dayalı çalışmalarda araştırmacılar tutarsız kodlamalar üzerinde tartışarak güvenilirliği arttırmaya çalışır. von Eye ve Mun'un (2005) sıraladığı bu örnekler çoğaltılabilir. Söz gelimi, yemek yarışmalarında hangi tabağın görsel ve lezzet anlamında daha başarılı olduğuna karar verirken birden fazla şefin değerlendirmesine başvurulur. Benzer şekilde, resim öğretmenliği programına seçilecek öğrencileri belirlemek için uygulanan özel yetenek sınavlarında değerlendirmeler en az iki üyeden oluşan bir kurul tarafından yapılır. Yine açık uçlu soruların kullanıldığı değerlendirmelerde öğrencinin yetenek düzeyine dair daha doğru çıkarımlar elde etmek amacıyla öğrencilerin cevapları birden fazla puanlayıcı tarafından puanlanır ve puanlayıcılar arasında bir uzlaş yoksa farklı puanlayıcılardan görüş alınır. Aynı işlem bilimsel makalelere ilişkin hakem değerlendirme süreçlerinde de takip edilir. Tüm bu örneklerden anlaşılacağı gibi puanlayıcı yargılarına dayalı ölçmeler doğası itibarıyla sübjektiftir (bkz. Şekil 1).



Şekil 1. Puanlayıcı yargılarına dayalı ölçmelerin genel yapısı (Eckes, 2019)

Puanlayıcı yargıları üzerine kurulu ölçmelerde, puanlayıcı kaynaklı hatalar objektif testlerde (çoktan seçmeli sınavlar, doğru-yanlış testleri vb.) söz konusu olmayan birçok sorunu beraberinde getirir. Buna bağlı olarak ölçümlerin psikometrik niteliği hakkında cevap aranması gereken sorular artar. Puanlanışı yönüyle subjektif nitelik taşıyan bir ölçme işlemi yürütülüyorsa aşağıdaki soruların yanıtlanması gerekir (Engelhard ve Wind, 2018):

- Puanlayıcı hatalarını azaltmak için ne tür tedbirler alabiliriz?
- Yapılan puanlamaların kalitesini ve doğruluğu nasıl denetleyebiliriz?
- Puanlama hatalarını nasıl tespit edebilir ve düzeltebiliriz?
- Eğitim Araştırmaları Birliği (American Educational Research Association), Amerikan Psikologlar Birliği (American Psychological Association) ve Eğitimde Ölçme Ulusal Konseyi'nin (National Council on Measurement in Education) *Eğitsel ve Psikolojik Test Standartlarında* tanımladığı geçerlik, güvenilirlik ve adillik gibi temel konuları puanlayıcı yargılarına dayalı ölçmelere nasıl uygulayabiliriz?

Bu sorulardan anlaşılacağı gibi puanlayıcı yargılarının devreye girdiği ölçmelerde geçerli ve güvenilir sonuçlara ulaşabilmek için puanlayıcı kaynaklı hataların bilinmesi, bu hataları minimuma indirecek önlemlerin alınması, puanlayıcılar arası uyumun sağlanıp sağlanmadığının ortaya konulması ve varsa uç noktalarda