
GAZİ ÜNİVERSİTESİ
Gazi
1926

PISA 2018 TESTİNDE BİREY-UYUM İNDEKSLERİNE GÖRE ANORMAL YANIT DESENLERİNİN BELİRLENMESİ

Esra Kamacı

YÜKSEK LİSANS TEZİ
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI

GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

ŞUBAT 2023

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren 3(üç) yıl sonra tezden fotokopi çekilebilir.

YAZARIN

Adı : Esra

Soyadı : Kamacı

Bölümü : Eğitimde Ölçme ve Değerlendirme Anabilim Dalı

İmza :

Teslim Tarihi :

TEZİN

Türkçe Adı: PISA 2018 Testinde Birey-Uyum İndekslerine Göre Anormal Yanıt Desenlerinin Belirlenmesi

İngilizce Adı: Determining Aberrant Response Patterns according to Person-Fit Indexes in the PISA 2018 test

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduğumu, yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu beyan ederim.

Yazarın Adı Soyadı: Esra Kamacı

İmza:

JÜRİ ONAY SAYFASI

Esra KAMACI tarafından hazırlanan "PISA 2018 Testinde Birey-Uyum İndekslerine Göre Anormal Yanıt Desenlerinin Belirlenmesi" adlı tez çalışması aşağıdaki jüri tarafından oy birliği/oy çokluğu ile Gazi Üniversitesi Eğitim Bilimleri Anabilim Dalı' nda Yüksek Lisans olarak kabul edilmiştir.

Danışman: Prof. Dr. Dilara BAKAN KALAYCOĞLU
(Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi)
Başkan: Prof. Dr. Burcu ATAR
(Eğitimde Ölçme ve Değerlendirme, Hacettepe Üniversitesi)
Üye: Doç. Dr. Melek Gülşah ŞAHİN
(Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi)

Tez Savunma Tarihi:

Bu tezin Eğitim Bilimleri Anabilim Dalı'nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Şaban Çetin
Eğitim Bilimleri Enstitüsü Müdürü …….

Ablam ve Annem'e

TEŞEKKÜR

Yüksek lisans eğitim sürecimin her aşamasında yoluma ışık olan, kıymetli vaktini ayırarak bilimsel her konuda engin bilgi ve tecrübesi ile bana destek olan çok değerli danışman hocam Prof. Dr. Dilara BAKAN KALAYCIOĞLU'na sonsuz sevgi saygı ve şükranlarımı sunarım.

Tez savunma jürimde yer alarak araştırmama görüş ve önerileriyle katkılarını sunan değerli hocalarım Prof. Dr. Burcu ATAR ve Doç. Dr. Melek Gülşah ŞAHİN'e çok teşekkür ederim.

Hayatım boyunca girdiğim her yolda, yaptığım her işte; bana güvenen, koşulsuz sevgisi ve duası ile hep benimle olan annem Sevim KAMACI ve babam Ahmet KAMACI'ya, en büyük şansım, güç kaynağım, destekçim olan ablam Özlem KAMACI TORUN ve abim Yusuf KAMACI'ya teşekkür ederim.

Eğitim hayatımın başından bu güne kadar beni daima yüreklendiren ve yanımda olan yol arkadaşım Ertan ÜRKMEZ'e teşekkür ederim.

Tezimin son aşamasında karşılaştığım zorluklarda sevgiyle yanımda olan bana huzurlu bir çalışma ortamı sağlayan Songül ve İbrahim AVCI'ya, akademik yaşamıma katkı sağlayan tüm değerli hocalarıma ve başta Songül Aybüke TOKGÖZ olmak üzere tüm arkadaşlarıma teşekkür ederim.

Esra KAMACI

PISA 2018 TESTİNDE BİREY-UYUM İNDEKSLERİNE GÖRE ANORMAL YANIT DESENLERİNİN BELİRLENMESİ

(Yüksek Lisans Tezi)

Esra Kamacı

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Şubat 2023

ÖZ

Eğitimsel ve psikolojik testlerde amaç gerçek puanlara yakın toplam puanlar elde etmektir. Testten alınan toplam puanlar, testi alan bireyin yanıt verme süreçlerinin gerçek yönlerini doğru bir şekilde yansıtamazsa geçersiz puan çıkarımları yapılabilir. Toplam puanlarla ilgili geçersiz çıkarımlara yol açabilecek durumlardan bir tanesi testi alan bireyin zor maddeleri doğru yanıtlarken kolay maddeleri yanlış yanıtlamasıdır. Test maddelerini bu şekilde yanıtlayan bireyleri tespit etmek için birey uyum indeksleri hesaplanmalıdır. Bu araştırmaya konu olan birey uyum indeksleri bireyin yanıt deseninin örneklemdeki diğer yanıt desenleri ile uyumunu araştırmak için kullanılan ve daha önceki çalışmalarda performansı sınanmış parametrik olmayan H^t, U3 ve C* indeksleridir. Çalışma için gerekli veri seti OECD tarafından 15 yaş grubundaki öğrencilerin belirli alanlarda kazandıkları bilgi ve becerileri değerlendirmek için yapılan ve uluslararası bir araştırma olan PISA'ya katılan ülkeler arasından Amerika Birleşik Devletleri, Çin, Güney Kore, Japonya, Rusya ve Türkiye'ye ait ikili puanlama yapılmış maddelerden elde edilmiştir. Veri seti R studio programına yüklenerek programda bulunan PerFit kütüphanesi yardımıyla indeksler hesaplanmış, her örneklem için kesme puanları belirlenmiş ve bu puanlar yardımıyla tüm ülkelerde anormal yanıt desenine sahip bireyler tespit edilmiştir. Öğrencilerin indeks puan ortalamalarının ülkeler ve formlar bazında karşılaştırmaları için ANOVA testi yapılmış ve ortalamalar arasındaki farklılaşmaların anlamlı olduğu tespit edilmiştir. Ortalamalar arasında meydana gelen farklılaşmaların hangi ülkeler ve formlardan kaynakladığını tespit etmek için Post-Hoc testi yapılmıştır. Anormal yanıt desenine sahip öğrencilerin madde yanıt grafikleri

çizdirilmiştir ve öğrenci sayısının ülkelere ve formlara göre karşılaştırmak için $\chi 2$ ilişki testi yapılmış ve öğrenci sayıları arasında anlamlı bir fark bulunmamıştır.

Anahtar Kelimeler : Birey uyum analizi, H^t, U3, C*, PISA
Sayfa Adedi : ix + 95
Danışman : Prof. Dr. Dilara BAKAN KALAYCOĞLU

DETERMINING ABERRANT RESPONSE PATTERNS ACCORDING TO PERSON-FIT INDEXES IN THE PISA 2018 TEST

(M.S. Thesis)

Esra Kamacı

GAZI UNIVERSTY

INSTITUTE OF EDUCATIONAL SCIENCES

February 2023

ABSTRACT

The aim of educational and psychological tests is to obtain total scores that are close to the real scores. If the total scores obtained from the test do not accurately reflect the true nature of the test taker's response process, invalid score inferences may be made. One of the situations that can lead to invalid inferences about total scores is when the test taker answers difficult items correctly but answers easy items incorrectly. To identify test takers who respond to the test items in this way, individual fit indices should be calculated. The non-parametric H^t, U3, and C* indices, which have been studied in previous research and are used to examine the conformity of the test taker's response pattern with other response patterns in the sample, are the subject of this research. The data set for this study was obtained from the binary scoring items of the United States, China, South Korea, Japan, Russia, and Turkey, which participated in the PISA, an international research conducted by the OECD to assess the knowledge and skills acquired by 15-year-old students. The data set was uploaded to the R studio program and the indexes were calculated with the help of the PerFit library in the program, cutting scores were determined for each sample and with the help of these scores, individuals with aberrant response patterns were identified in all countries. ANOVA test was performed for the comparison of the index score averages of the students on the basis of countries and forms and it was determined that the differences between the averages were significant. Item response graphs of students with aberrant response patterns were drawn, and $\chi 2$ relationship test was performed to compare the number

of students by country and form, and no significant difference was found between the number of students.

Key Words : Person-Fit Analysis, H^t, U3, C*, PISA
Page Number : ix + 95
Supervisor : Prof. Dr. Dilara BAKAN KALAYCOĞLU

İÇİNDEKİLER

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU ..... **i**

ETİK İLKELERE UYGUNLUK BEYANI ..... **ii**

JÜRİ ONAY SAYFASI ..... **iii**

ÖZ ..... **vi**

ABSTRACT ..... **viii**

İÇİNDEKİLER ..... **x**

TABLOLAR LİSTESİ ..... **xiv**

ŞEKİLLER LİSTESİ ..... **xvi**

KISALTMALAR LİSTESİ ..... **xvii**

BÖLÜM I **1**

GİRİŞ **1**

PISA **2**

PISA 2018 ..... **4**

Birey Uyum Analizi **5**

Birey Uyum İndeksleri **6**

Parametrik Olmayan Birey Uyum İndeksleri **7**

Guttman Modeline Dayalı İndeksler ***8***

G İndeksi *8*

G İndeksi* *9*

Uyuşmaya Dayalı İndeksler ***11***

A, D, E indeksleri *11*

Kovaryansa Dayalı İndeksler ***12***

rpbis, rbis indeksleri *12*

H^t İndeksi *13*

U3 İndeksi *13*

C İndeksi *14*

C İndeksi* *14*

Hangi Birey Uyum İndeksi Kullanılmalıdır? **17**

Birey Uyum Analizi Süreci **17**

Anormal Yanıt Deseni **18**

Anormal Yanıt Desenlerinin Oluşma Sebepleri **19**

Birey Uyum İstatistiklerinin Eğitim Bağlamında Kullanımı **22**

Madde Yanıt Grafikleri (Person Response Function Plot [PRF]) **23**

Kesme Puanı **25**

Araştırmanın Amacı ve Önemi **27**

Araştırma Problemi **28**

Sınırlılıklar **29**

Tanımlar **29**

BÖLÜM II **30**

İLGİLİ ARAŞTIRMALAR **30**

İlgili Araştırmalar Özet **30**

BÖLÜM III 35

YÖNTEM 35

Yöntem 35

Evren 36

Örneklem 36

Veri Analizi 37

Bulgular 55

PISA 2018 Araştırmasına Amerika Birleşik Devletleri, Çin, Güney Kore, Japonya, Rusya Ve Türkiye'den Katılan Ve İlk Beş Formu Alan Öğrencilerin H^t Birey Uyum İndeks Puan Ortalamaları Farklılık Göstermekte Midir? 56

Ülkelerin ve Formların H^t İndeks Puan Ortalamaları Üzerindeki Etkileşim Etkisi Nedir? Öğrencilerin H^t İndeks Puan Ortalamaları Ülkelere ve Formlara Göre Farklılık Göstermekte Midir? 56

H^t İndeksi Tarafından Yanıt Deseni Anormal Olarak Belirlenen Öğrenci Sayıları Ülkelere Göre Farklılık Göstermekte Midir? 59

H^t İndeksi Tarafından Yanıt Deseni Anormal Olarak Belirlenen Öğrenci Sayıları Formlara Göre Farklılık Göstermekte Midir? 60

PISA 2018 Araştırmasına Amerika Birleşik Devletleri, Çin, Güney Kore, Japonya, Rusya ve Türkiye'den Katılan Ve İlk Beş Formu Alan Öğrencilerin U3 Birey Uyum İndeks Puan Ortalamaları Farklılık Göstermekte Midir? ... 61

Ülkelerin ve Formların U3 İndeks Puan Ortalamaları Üzerindeki Etkileşim Etkisi Nedir? Öğrencilerin U3 İndeks Puan Ortalamaları Ülkelere ve Formlara Göre Farklılık Göstermekte Midir? 61

U3 İndeksi Tarafından Yanıt Deseni Anormal Olarak Belirlenen Öğrenci Sayıları Ülkelere Göre Farklılık Göstermekte Midir? 65

U3 İndeksi Tarafından Yanıt Deseni Anormal Olarak Belirlenen Öğrenci Sayıları Formlara Göre Farklılık Göstermekte Midir? *65*

PISA 2018 araştırmasına Amerika Birleşik Devletleri, Çin, Güney Kore, Japonya, Rusya ve Türkiye'den Katılan Ve İlk Beş Formu Alan Öğrencilerin C* Birey Uyum İndeks Puan Ortalamaları Farklılık Göstermekte Midir? ... **66**

Ülkelerin ve Formların C İndeks Puan Ortalamaları Üzerindeki Etkileşim Etkisi Nedir? Öğrencilerin C* İndeks Puan Ortalamaları Ülkelere ve Formlara Göre Farklılık Göstermekte Midir?* *67*

C İndeksi Tarafından Yanıt Deseni Anormal Olarak Belirlenen Öğrenci Sayıları Ülkelere Göre Farklılık Göstermekte Midir?* *70*

C İndeksi Tarafından Yanıt Deseni Anormal Olarak Belirlenen Öğrenci Sayıları Formlara Göre Farklılık Göstermekte Midir?* *71*

BÖLÜM IV 69

SONUÇ, TARTIŞMA VE ÖNERİLER 69

Sonuç ve Tartışma 69

Öneriler 73

KAYNAKLAR 73

EKLER 83

Ek-1 84

Ek-2 86

Ek-3 90

Ek-4 92

Ek-5 94

TABLOLAR LİSTESİ

Tablo 1. *Birey Uyum İndeksleri* 7

Tablo 2. *Guttman Modeline İlişkin Örnek Bir Uygulama* 10

Tablo 3. H^t*, U3 ve C* İndekslerinin Hesaplandığı Örnek Bir Uygulama* 15

Tablo 4. *R Studio PerFit Kütüphanesinde Bulunan İstatistikler* 16

Tablo 5. *12 Maddelik Bir Testte Anormal Yanıt Desenleri* 21

Tablo 6. *Alt formları Alan Öğrenci Sayılarının Ülkelere Göre Dağılımı* 37

Tablo 7. *Formların Betimsel İstatistik Tabloları* 38

Tablo 8. *Alt formların Cronbach Alfa Güvenirlik Katsayıları* 39

Tablo 9. *Alt Formların Ortalama Güçlüğü* 40

Tablo 10. *Form-1 Madde Güçlük İndeksleri* 41

Tablo 11. *Form-1 İndeks Puanlarının Ortalama ve Standart Sapma Değerleri* 44

Tablo 12. H^t *İndeksi İçin Kesme Puanları* 45

Tablo 13. *U3 İndeksi İçin Kesme Puanları* 46

Tablo 14. *C* İndeksi İçin Kesme Puanları* 46

Tablo 15. H^t *indeksine göre Anormal Yanıt Desenine Sahip Birey Sayısı ve Oranı* 47

Tablo 16. *U3 indeksine göre Anormal Yanıt Desenine Sahip Birey Sayısı ve Oranı* 48

Tablo 17. *C* indeksine göre Anormal Yanıt Desenine Sahip Birey Sayısı ve Oranı* 49

Tablo 18. *Form-1 Anormal Yanıt Veren Birey Sayısının C*, H^t ve U3 İndekslerine Göre Dağılımı* 49

Tablo 19. *Form-1 Anormal Yanıt Veren Birey Sayısının H^t, U3 ve C* İndekslerine Göre Dağılımı* 51

Tablo 20. *Form-1 de H^t indeksine göre anormal yanıt deseni gösteren bireyler* 52

Tablo 21. *Form-1 ABD Anormal Yanıt Desenine Sahip Bireylerin PRF Grafikleri* 53

Tablo 22. *Hesaplanan İndekslere Ait Korelasyon Katsayıları* 54

Tablo 23. *H^t İndeks Ortalamalarına Ait İki-Yönlü ANOVA Tablosu* 57

Tablo 24. *H^t İndeks Puanları Ülkelere Ait Post-Hoc Sonuçları* 58

Tablo 25. *H^t İndeks Puanları Formlara Ait Post-Hoc Sonuçları* 59

Tablo 26. *H^t İndeksi ülkeler arasındaki χ^2 İlişki Testi* 60

Tablo 27. *H^t İndeksi formlar arasındaki χ^2 İlişki Testi* 60

Tablo 28. *U3 İndeks Ortalamalarına Ait İki-Yönlü ANOVA Tablosu* 62

Tablo 29. *U3 İndeks Puanları Ülkelere Ait Post-Hoc Sonuçları* 64

Tablo 30. *U3 İndeks Puanları Formlara Ait Post-Hoc Sonuçları* 64

Tablo 31. *U3 İndeksi ülkeler arasındaki χ^2 İlişki Testi* 65

Tablo 32. *U3 İndeksi formlar arasındaki χ^2 İlişki Testi* 66

Tablo 33. *C* İndeks Ortalamalarına Ait İki-Yönlü ANOVA Tablosu* 67

Tablo 34. *C* İndeks Puanları Ülkelere Ait Post-Hoc Sonuçları* 69

Tablo 35. *C* İndeks Puanları Formlara Ait Post-Hoc Sonuçları* 69

Tablo 36. *C* İndeksi ülkeler arasındaki χ^2 İlişki Testi* 70

Tablo 37. *C* İndeksi formlar arasındaki χ^2 İlişki Testi* 71

ŞEKİLLER LİSTESİ

Şekil 1. Beklenen yanıt deseni 24

Şekil 2. Anormal yanıt deseni 24

Şekil 3. H^t İndeksi kesme puanı (R studio) 26

Şekil 4. U3 İndeksi kesme puanı (R studio) 26

Şekil 5. C* İndeksi kesme puanı (R studio) 27

Şekil 6. H^t İndeksinden elde edilen puanlara ait histogram 42

Şekil 7. U3 İndeksinden elde edilen puanlara ait histogram 43

Şekil 8. C* İndeksinden elde edilen puanlara ait histogram 43

Şekil 9. H^t İndeks Puan Ortalamaları 58

Şekil 10. U3 İndeksine Ait Puan Ortalamaları 63

Şekil 11. C* İndeksine Ait Puan Ortalamaları 68

KISALTMALAR LİSTESİ

DMF	Değişen Madde Fonksiyonu
KTK	Klasik Test Kuramı
MEB	Milli Eğitim Bakanlığı
MTK	Madde Tepki Kuramı
OECD	Ekonomik İşbirliği ve Kalkınma Örgütü
PISA	Uluslararası Öğrenci Değerlendirme Programı
PRF	Madde Yanıt Grafiği

BÖLÜM I

GİRİŞ

Test puanlarının geçerliği psikometride önemli bir konudur. Test puanlarından, sınava girenlerin davranışları hakkında test oturumunun ötesinde çıkarımlar yapmak için kullanılır (Crocker & Algina, 2008). Geçerlik, testlerin kullanım amacı için test puanlarını destekleyen teori ve kanıtların düzeyidir (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014).

Geçerlik ile ilgili kanıt toplama süreci 2014 yılında yayınlanan standartlarda (APA, AERA & NCME, 2014) testin içeriği (kapsamı), tepki (cevaplama, yanıtlama) süreçleri, diğer değişkenlerle ilişkiler, içsel yapı ve testlerin sonuçları şeklinde sınıflandırılmıştır. Her bir kanıt toplama süreci için farklı analizlerin yapılması önerilmektedir, örneğin; içsel yapıya ilişkin kanıt toplama sürecinde faktör analizi yapılırken, test içeriğine ilişkin kanıt toplama sürecinde uzman görüşü alınabilir veya tepki sürecine ilişkin kanıt toplamak amacıyla birey uyum analizi yapılabilmektedir.

Test katılımcısı yanıtlama sürecinde rastgele, yaratıcı, dikkatsiz, uykulu, kopya çekerek vb. sebeplerle maddelere tutarsız yanıtlar verebilir, bu durum test puanlarının geçerliğini tehdit eder ve bu test puanların kullanılmasını sorunlu hale getirir (Meijer, 1996). Bu şekilde yanıtlanan maddelerden alınan puanlar test katılımcısının toplam puanının bir göstergesi olamaz (Avşar & Emons, 2021). Bu nedenle eğitimsel ve psikolojik testlerle ilgili çeşitli kılavuzlarda ve yönergelerde, test katılımcısından elde edilen cevap verilerinin kalitesinin kontrol edilmesi önerilmektedir (Olson & Fremer, 2013). Literatür, bu bağlamda birey uyum

analizinin, test puanlarının geçerliğini kontrol etmek için kullanılabileceğini göstermiştir (Walker, Jemings & Engelhard, 2018).

Birey uyum analizi, bir test katılımcısının verdiği yanıtların oluşturduğu madde yanıt deseninin varsayımsal bir modele uyup uymadığını tespit etmek amacıyla yapılmaktadır. Birey uyum analizi için birey uyum indeksleri hesaplanır ve elde edilen sonuçlara göre modele uygun olmayan yanıt deseni tespit edilir ve anormal olarak tanımlanır (Walker, 2017). Bireylerin verdiği yanıtlarda anormal bir duruma rastlandığında araştırmacı toplam puanlar hakkında yorum yaparken dikkatli olmalıdır. Teorik (Karabatsos, 2003; Meijer, 1996; Rudner, 1983; Rupp, 2013) ve deneysel (Emmen, 2011) araştırmalarla desteklenen birey uyum analizleri, puan yorumlama ve kullanımı ile ilgili geçerlik kanıtları sağlamasına rağmen henüz büyük ölçekli K-12 eğitim testi uygulamalarına uyarlanmamıştır (Cui Roberts, 2018; De La Torre & Deng, 2008).

Bu çalışmada öğrencilerin öğrenme çıktılarının uluslararası değerlendirmesine imkan sağlayan PISA verileri üzerinde birey uyum analizi yapılmıştır.

Seçilen ülkelerdeki öğrencilerin yanıt desenlerinin incelenmesi amacıyla, PISA araştırmasından elde edilen veri setinde çeşitli birey uyum indeksleri hesaplanmış ve daha sonra hesaplanan indeksler yardımıyla, bireyin testi yanıtlama sürecine yönelik yorumlar yapılmıştır.

PISA

PISA (Uluslararası Öğrenci Değerlendirme Programı), Ekonomik İşbirliği ve Kalkınma Teşkilatı (Organisation for Economic Cooperation and Development-OECD) tarafından 15 yaş grubundaki öğrencilerin belirli alanlarda kazandıkları bilgi ve becerileri değerlendiren, üçer yıllık döngülerle yapılan uluslararası bir araştırmadır (MEB, 2019). Öğrencilerin öğrenme çıktılarında, uluslararası değerlendirmeyi sağlar (OECD, 2019). PISA araştırması sonrasında yayınlanan verilere dayalı raporlar, eğitimcilere ve politika yapıcılara ülkelerinin eğitim sisteminin güçlü ve zayıf yönlerini, zaman içinde kaydedilen ilerleme ve iyileştirme

fırsatları hakkında faydalı bilgiler sağlamaktadır (OECD, 2019). Ülkelerin eğitim performanslarını etkileyen faktörleri ölçerek modern toplumların beklentilerine ve çabalarına katkıda bulunmaktadır (Uçar & Karsak, 2021).

PISA çalışmaları, OECD'ye üye ülkelerin eğitim sistemlerinin performanslarını izlemek ve değerlendirmek amacıyla uygulanmaya başlanmıştır. PISA, güvenilir ve sürdürülebilir verilere ulaşmak amacıyla 1997 yılında resmi olarak yürürlüğe girmiştir. İlk değerlendirmeye ise 2000 yılında başlanmış ve sonrasında üç yıllık döngülerle uygulanmaya devam edilmiştir.

OECD tarafından, uygulanan PISA, matematik, fen ve okuma alanlarında öğrencilerin okuryazarlık becerilerinin değerlendirildiği, en yüksek katılımlı uluslararası izleme çalışmasıdır (OECD, 2019). Her bir PISA uygulamasında; okuma, matematik, fen okuryazarlığı alanlarından bir tanesi temel alan olarak değerlendirilmektedir. PISA araştırması, zorunlu eğitimin sonunda örgün eğitime devam eden, 15 yaş grubu öğrencilerin, öğretim programında yer alan kazanımlara ulaşmış olma yeterliliklerini değil; günümüz bilgi toplumunda karşılaştıkları soru ve sorunlarda, sahip oldukları bilgiyi kullanma becerilerini ölçmeyi amaçlamaktadır. Eğitim göstergelerinin çeşitliliği ve içeriği açısından, literatürdeki en kapsamlı araştırmalar arasındadır. Eğitim alanında çalışan akademisyenlere zengin bir veri tabanı sunmakta, ülke içindeki çalışmalara örnek teşkil etmekte ve özellikle nitelikli ölçme araçlarının geliştirilmesine fayda sağlamaktadır.

PISA araştırmasına katılacak okul ve öğrencilerin seçimi, OECD tarafından aşamalı tesadüfi (seçkisiz) yöntemle belirlenmektedir (MEB, 2018).

PISA 2000: İlk kez uygulanan PISA araştırmasına 30'u OECD üyesi olmak üzere toplam 32 ülke katılmıştır. PISA 2000 araştırması, okuma ağırlıklıdır (OECD, 2003). Türkiye bu uygulamaya katılmamıştır.

PISA 2003: OECD ülkelerinden 33, toplam 41 ülke bu araştırmaya katılmıştır (OECD, 2020). PISA 2003 araştırması matematik, ağırlıklıdır. Türkiye ilk kez 2003 yılında PISA'ya katılmıştır (MEB, 2018).

PISA 2006: Bu araştırmaya OECD ülkelerinden 39, toplam 57 ülkeden katılım olmuştur (OECD, 2020). PISA 2006 araştırması, fen ağırlıklıdır.

PISA 2009: PISA 2009'a, 39'u OECD üyesi olmak üzere toplam 65 ülke katılmıştır (OECD, 2020). PISA 2009 araştırması, 2000 yılındaki gibi okuma ağırlıklıdır, ancak bu uygulamadan farkı, okuma çerçevesinin güncellenmiş olması ve elektronik ortamda sunulan metinlerin değerlendirmesini içermesidir (OECD, 2009).

PISA 2012: PISA'nın beşinci döngüsü olan PISA 2012 araştırmasına 39'u OECD üyesi olmak üzere 64 ülke katılmıştır (OECD, 2020). PISA 2012 araştırması, matematik ağırlıklıdır.

PISA 2015: PISA'nın altıncı döngüsü olan PISA 2015 araştırması, 39'u OECD üyesi olmak üzere 72 ülkenin katılımıyla uygulanmıştır (OECD, 2020). PISA 2015 araştırmasının ağırlıklı alanı fen okuryazarlığıdır.

PISA 2018

Bu araştırmaya konu olan PISA 2018 uygulamasına, 39'u OECD üyesi olmak üzere toplam 79 ülkeden, 32 milyon 15 yaşındaki çocuğu temsil eden yaklaşık 600.000 öğrenci katılmıştır (OECD, 2019). Bilgisayar tabanlı testler kullanılarak yapılan sınavda, çoktan seçmeli ve açık uçlu maddeler sorulmuştur. Her bir değerlendirme iki saat sürmüştür. PISA'nın yedinci döngüsü olan bu araştırmada ağırlıklı alan, okuma alanı olmuştur. Bu döngüye diğer döngülerden farklı olarak; finansal okuryazarlık, küresel yeterlik ve refah çerçeveleri de eklenmiştir. Dolayısıyla bu döngüde öğrencilerin bilgi, beceri ve refahını değerlendirmek amacını taşıyan ve oldukça kapsamlı hazırlanan uluslararası testler ve anketler uygulanmıştır (Hacettepe Üniversitesi, 2020).

PISA 2018 matematik uygulaması, 24 ayrı form üzerinden gerçekleştirilmiştir. Bu formlarda ortak sorular olduğu gibi her bir form için ortak olmayan sorular da mevcuttur. Her formu, uygulamaya katılan öğrencilerin bir kısmı almaktadır. Çalışma kapsamında, 24 paralel formun ilk 5 tanesini alan öğrenciler üzerinde analizler gerçekleştirilecektir. OECD,

kişilerin kimlik bilgilerini paylaşmadığından, birey bazında yapılan analizler herhangi kişisel bir bilgi içermemektedir.

Birey Uyum Analizi

Testler test katılımcıları için farklı geçerlik derecelerine sahiptir (Ghiselli, 1972). Testten elde edilen ölçümler örneklemin tamamı için geçerli olsa bile, bazı bireyler anormal yanıt desenine sahip olduğu için o bireylere ait puanların geçerliliği tartışmalı hale gelmiş olabilir (Drasgow & Suertler, 1987). Bireylerin test puanları benzer olsa da yanıt desenlerinin farklı bir hikayesi olabilir. Genellikle test uygulandıktan sonra bireylerin test puanları yorumlanır fakat bireylerin yanıt desenlerini değerlendirmek için bir uygulama yapılmaz (Rudner, 1983; Tellegen, 1998). Bireylerin yanıt desenleri incelendiğinde bazı bireylere ait yanıt desenlerinin mantıklı olmadığı tespit edilebilir. Bu gibi durumlarda yanıt desenlerini incelemek için birey uyum analizi yapılmalıdır (Drasgow, Levine & McLaughlin, 1987; Parsons, 1983).

Eğitimsel ve psikolojik test literatüründe, sınav katılımcısının testteki (ölçme aracındaki) maddelere verdiği yanıtların değerlendirilmesine; birey güvenirliği-person reliability- (Lumsden, 1997), uygunluk ölçümü-appropriateness measurement (Levine & Rubin, 1979), birey uyumu-person fit (Smith, 1986), birey uyumsuzluğu-person aberrancy (Meijer, 1996) ve birey ölçeklenebilirliği-person scalability (Reise & Flannery, 1996) adı verilmiştir. Birey uyum analizinin pratik uygulamada yapılması test katılımcısının toplam puanından yola çıkarak verilecek yanlış eğitim kararlarını önleyebilir (Walker, 2017). Örneğin; başarısız bir birey anormal yanıtlama davranışı sergileyerek, hak etmediği bir eğitim programına dâhil olabilir veya başarılı bir birey anormal yanıtlama davranışı sonucunda, alınması gereken bir eğitim programına alınmayabilir. Bunun yanı sıra anormal yanıtlama davranışı sergileyen bireylerin, öğrenme ilerlemelerinin değerlendirilmesi ve bu ilerlemeye bağlı olarak müfredatın planlanması süreci zorlaşabilir. Örnekte açıklandığı üzere; yanlış ölçümler, sınav katılımcıları için olumsuz sonuçlara neden olabilir.

Birey uyumu analizleri, test katılımcılarının gerçek puanından düşük veya yüksek test puanları almalarıyla sonuçlanan anormal madde yanıt desenlerini tanımlamak için kullanılabilir (Karabatsos, 2003).

Birey Uyum İndeksleri

Bireylerin testten aldığı toplam puanlar; genel yetenek veya başarı düzeylerini belirlemek için kullanılır. Bir test katılımcısının toplam puanı hakkında daha ayrıntılı bilgi elde etmek için bireysel olarak madde yanıt desenleri incelenebilir. Bu madde yanıt desenleri çeşitli tespitler için de kullanılabilir (Karabatsos, 2003).

İndeksler, bireylerin gözlemlenen yanıt deseninin; KTK (Klasik Test Kuramı), MTK (Madde Tepki Kuramı) ve Rasch gibi modellerde beklenen yanıt deseniyle veya örneklemdeki diğer bireylerin yanıt desenleriyle ne kadar iyi eşleştiğini belirler (Reise, 2000). Birey uyum indeksleri anormal yanıt desenlerini tespit eder fakat bireyin yanıt deseninindeki anormal davranışın nedenini açıklamaz (Meijer & Sijtsma, 1995). Anormal yanıt desenine sahip bireylerin, test tarafından ölçülemeyen faktörlerden etkilenmiş olabileceğine dair bir gösterge niteliğindedir (Emons, Meijer & Sijtsma, 2002).

Bir dizi ölçüm çerçevesinde uygulanmak üzere çeşitli birey uyum indeksleri geliştirilmiştir. Meijer ve Sijtsma (2001) birey uyum indekslerini incelerken KTK'ye ya da gruba dayalı, MTK'ya dayalı indeksler olarak ayırmıştır. Tablo 1'de literatür tarafından kabul görmüş bu sınıflandırma göz önünde bulundurularak sıklıkla kullanılan birey uyum indeksleri özetlenmiştir.

Tablo 1

Birey Uyum İndeksleri

KTK'ye veya Gruba Dayalı (Parametrik Olmayan) Birey Uyum İstatistikleri
G (Guttman, 1944, 1950)
*G** (Van Der Flier, 1997)
rpbis, *rbis* (Donlon & Fischer, 1968)
C (Sato, 1975)
U (Van Der Flier, 1980; Meijer, 1994)
A_i, D_i, E_i (Kane & Brennan, 1980)
*C** (Harnisch & Linn, 1981)
ZU_3 (Van Der Flier, 1982)
NCI, *ICI* (Tatsuoka & Tatsuoka, 1983)
H^t (Sijtsma, 1986; Sijtsma & Meijer, 1992)
MTK'ya Dayalı (Parametrik) Uyum İndeksleri
l_θ (Levine & Rubin, 1979)
D(θ)(Weiss, 1973; Trabin & Weiss, 1983)
ECI (Tatsuoka, 1984)
lz (Drasgow, Levine, & Williams, 1985)
JK, *O/E* (Drasgow, Levine & McLaughlin, 1987)
lzm (Drasgow, Levine & McLaughlin, 1991)
c (Levine & Drasgow, 1988)

Birey uyum indeksleri farklı şekillerde sınıflandırılmaktadır. Bu sınıflandırmalardan en yaygın olanı, parametrik ve parametrik olmayan birey uyum indeksleri şeklindedir. Araştırmaya konu olan parametrik olmayan birey uyum indeksleri, sınava giren kişilerin test maddelerine verdikleri yanıtların oluşturduğu veri kümesinden hareketle hesaplanmaktadır.

Parametrik Olmayan Birey Uyum İndeksleri

Parametrik olmayan birey uyum indeksleri MTK, KTK gibi modellerin varsayımlarına dayanmamaktadır, bireyin yanıt deseninin örneklemdeki diğer yanıt desenleriyle uyumunu değerlendirmek için kullanılan bir dizi ölçümü içermektedir (Karabatsos, 2003). Tüm birey uyum indekslerinin nihai hedefi bireyin yanıt deseninin makullüğünü denetlemektir. Bireyin yanıt deseninin beklenen yanıt desenlerinden ne kadar saptığını belirleyen indeksleri birbirinden ayıran husus yukarıda sözü edilen beklentinin nasıl oluştuğudur. Beklenen yanıt

deseni modele göre veya örnekleme göre oluşmaktadır, bu desenden sapan bireylerin yanıt deseni anormal olarak sınıflandırılmaktadır.

Genel olarak, parametrik olmayan birey uyum indekslerinde beklenen; bireyin doğru yanıtlama oranı (güçlüğü) 1'e yakın maddeleri doğru yanıtlama eğiliminde, güçlüğü 0'a yakın olan maddeleri ise yanlış yanıtlama eğiliminde olmasıdır. Parametrik olmayan birey uyum indekslerinde beklenti, örneklemden tahmin edilen madde özelliklerine bağlı olarak belirlenir (Karabatsos, 2003).

Parametrik olmayan birey uyum indekslerinin her birinin nasıl hesaplandığı konusunda değişkenlikler aşağıda sunulmuştur.

Guttman Modeline Dayalı İndeksler

Klasik Test Kuramı bir yanıt deseninden hangi özelliklerin beklenilebilir olduğunu belirlemek için basit ve sezgisel bir tekniğe dayanmaktadır. Bu modele göre bir maddenin güçlüğü (p) bireyin yetenek düzeyinden büyük veya eşit ise doğru yanıtlanma olasılığı 1'dir, bireyin yetenek düzeyinden daha küçük ise doğru yanıtlanma olasılığı 0'dır (Guttman, 1944; Meijer & Sijtsma, 2001).

Gruba dayalı birey uyum indekslerinin birçoğu Guttman modeline (1944;1950) dayanmaktadır.

G İndeksi

Parametrik olmayan ilk birey uyum indeksi Guttman tarafından türetilen G indeksidir. Meijer ve Sijtsma (2001) tarafından açıklandığı gibi en kolay maddeler doğru yanıtlandığında geriye kalanlar ise yanlış yanıtlandığında Guttman desenine uygun bir yanıt deseni ortaya çıkarken, en zor maddeler doğru en kolay maddeler ise yanlış yanıtlandığında anormal yanıt deseni oluşur ve bu desen ters Guttman olarak da bilinir.

Maddeler doğru yanıtlandığında 1, yanlış yanıtlandığında 0 ile kodlanır ve madde sırası 1 ve 2 şeklinde ifade edildiğinde 2'nci sıradaki madde 1'inci maddeden daha zor bir madde olacak şekilde sıralandığında 0-0, 1-0, 0-1 ve 1-1 yanıt desenleri ortaya çıkmaktadır.

0-0 yanıt desenine sahip birey hem kolay hem de zor maddeyi yanlış yanıtlamıştır. 1-1 yanıt desenine sahip birey her iki maddeyi de doğru yanıtlamıştır. Bu iki yanıt deseninde de bir hata söz konusu değildir. 1-0 yanıt desenine sahip olan birey kolay olan ilk maddeyi doğru yanıtlarken zor olan maddeyi yanlış yanıtlamıştır bu yanıt deseni Guttman tarafından Mükemmel Desen olarak adlandırılmaktadır. Fakat 0-1 yanıt desenine sahip bir birey kolay maddeyi yanlış yanıtlarken zor maddeyi doğru yanıtlamıştır ve bu desen Guttman hatası olarak tanımlanmıştır (Meijer & Sitjma, 2001; Yormaz, 2020). Bireyin yanıt desenindeki bu hataların sayısı G indeksini vermektedir. G temel indeksi olup Guttman'a dayalı indekslerin tümü bu hatalı yanıt desenlerinin ağırlıklandırılmış sayılarına bağlıdır. Türetilen indekslerin sonuçları farklı ağırlıklandırmalar nedeniyle farklılaşmaktadır. G indeksi yukarıda örneklendirilen durumdaki (0-1) yanıt deseninin sayısıdır.

G İndeksi*

Test uzunluğu arttıkça G değeri de artma eğiliminde olduğundan G* indeksi (van der Flier, 1977) türetilmiştir. Bir birey N sayıda maddeye doğru yanıt verdiğinde G indeksi maksimum N.(N-1) değerini almaktadır. G* indeksi, G indeksini, alabileceği maksimum değere bölerek [0,1] arasında sınırlandırır (Karabatsos, 2003). Tablo 2'de kolaydan zora doğru sıralanmış 4 test maddesi vardır. Bu maddeler karşısında farklı yanıt desenleri sergileyen 9 adet bireye ait yanıtlar sunulmuş ve bireylerin yanıt deseninden ortaya çıkan tepki çiftleri verilmiştir. Tablo incelendiğinde A bireyi tüm maddeleri doğru, E bireyi ise tüm maddeleri yanlış yanıtladığı için Guttman hatası yapmamıştır. B, C ve D bireyleri ise kolay maddeleri doğru yanıtlarken maddeler zorlaştıkça yanlış yanıtlar verdikleri için Guttman hatası yapmamış olup yanıt desenleri normaldir.

Tablo 2

Guttman Modeline İlişkin Örnek Bir Uygulama

	Maddeler						
Birey	1 (kolay)	2	3	4 (zor)	Tepki Çiftleri	G	G*
A	1	1	1	1	(1,1)(1,1)(1,1)(1,1)(1,1)(1,1)	0	0.00
B	1	1	1	0	(1,1)(1,1)(1,1)(1,0)(1,0)(1,0)	0	0.00
C	1	1	0	0	(1,1)(1,0)(1,0)(1,0)(1,0)(0,0)	0	0.00
D	1	0	0	0	(1,0)(1,0)(1,0)(0,0)(0,0)(0,0)	0	0.00
E	0	0	0	0	(0,0)(0,0)(0,0)(0,0)(0,0)(0,0)	0	0.00
F*	0	1	0	0	(0,1)(0,0)(0,0)(1,0)(1,0)(0,0)	1	0.33
G*	1	0	1	0	(1,0)(1,1)(1,0)(0,1)(0,0)(1,0)	1	0.33
H*	1	0	0	1	(1,0)(1,0)(1,1)(0,0)(0,1)(0,1)	2	0.50
I*	0	1	1	1	(0,1)(0,1)(0,1)(1,1)(1,1)(1,1)	3	1.00
Güçlük	0.67	0.55	0.44	0.33	Madde Çiftleri	$G = \sum n_{uh}\,(1 - n_{ue})$	$G^* = \frac{G}{X_i\,(n - X_i)}$
					(12) (13) (14) (23) (24) (34)		

(*)Anormal yanıt desenini ifade etmektedir.

Yormaz, S. (2019). *Test güvenliği açısından bireyler arasındaki olası iş birliğinin incelenmesi* (doktora tezi). Mersin Üniversitesi Eğitim Bilimleri Enstitüsü, Mersin.

G bireyi en kolay maddeyi doğru yanıtladıktan sonra daha zor olan maddeye yanlış yanıt vermiştir. G bireyinin yanıt deseninin devamında beklenen daha zor madeleri de yanlış yanıtlamasıdır fakat ikinci maddeden daha zor olan üçüncü maddeye doğru yanıt vermiştir. Tepki çiftleri incelendiğinde bir tane (0,1) tepki çifti bulunmaktadır bir tane Guttman hatası yapmıştır dolayısıyla sapma G=1'dir ve anormal yanıt deseni sergilemiştir. H bireyi de G'ye benzer şekilde anormal yanıt desenine sahiptir. Tepki çiftleri incelendiğinde iki tane (0,1) tepki çifti bulunmaktadır yani iki tane Guttman hatası yapmış olup iki birim sapma vardır G=2'dir. I bireyi sadece en kolay maddeyi doğru yanıtlamıştır. Oysa beklenen zor maddeler

doğru yanıtlandığında kolay maddenin de doğru yanıtlanmasıdır. Bu durumda bireyin yanıt deseninde 3 birim sapma vardır ve G=3'tür.

Uyuşmaya Dayalı İndeksler

Kane ve Brennan'ın (1980) geliştirmiş olduğu indekslerdir.

A, D, E indeksleri

Uyuşmaya dayalı ilk indeks A indeksidir, bireyin maddeye verdiği tepki ile maddenin güçlüğünün eşleşmesine dayalıdır. Maddeler kolaydan zora doğru sıralanadığında A indeksi aşağıdaki formül yardımıyla hesaplanmaktadır.

A.KB =$\sum Xi\rho i$

Xi= n kişisinin i maddesindeki 1-0 puanı

ρi= i maddesinin doğru yanıtlanma oranı

A.KB indeksinin küçük değerleri yanıt deseninin uyumsuz olduğunun habercisidir.

D indeksi ise A indeksine bağlı olarak; anormal yanıt desenlerinin miktarını hesaplamaya çalışır ve A.KB(max) n kişisinin total puanı ile elde edilebilecek en büyük değer olmak üzere aşağıdaki formül yardımıyla hesaplanır.

D.KB = A.KB(max)−A.KB

D.KB büyük ise yanıt desenin uyumsuz olduğundan söz edilebilir.

D indeksinin [0,1] aralığında normalleştirilmesiyle elde edilen, E indeksi ise A indeksine bağlı olarak; E.KB = A.KB/A.KB(max) ile hesaplanır. E.KB indeksinin küçük değerleri yanıt deseninin uyumsuz olduğunun habercisidir.

Kovaryansa Dayalı İndeksler

rpbis, rbis indeksleri

Donlon ve Fischer (1968) bireyin puan vektörü ile örneklemin madde doğru yanıtlama oranı arasındaki korelasyonun birey uyum analizi için kullanılmasını önermiştir. Bu indekste düşük değerler uyumsuz yanıt vektörlerinin göstergesidir. Çift serili korelasyon aşağıdaki formül yardımıyla hesaplanmaktadır.

$$r_{bis} = \frac{\bar{Y}_i - \bar{Y}_r}{S_{Y_r}} . \frac{p_r q_r}{u}$$

- $\bar{Y}_i$ = Doğru yanıtlanan maddelerin ortalaması
- $\bar{Y}_r$ = r maddesini doğru yanıtlayanların ortalaması
- $S_{Y\,r}$= r maddesini doğru yanıtlayanların standart sapması
- p_r = r maddesinin güçlüğü
- q_r = 1- p_r
- y = Birim normal dağılım eğrisi altında kalan alanda p_r ve 1- p_r'yi ayıran ordinat yüksekliği

Nokta çift serili korelasyon ise aşağıdaki formül yardımıyla hesaplanmaktadır.

$$r_{pbis} = \frac{\bar{Y}_i - \bar{Y}_r}{S_{Y_r}} . \sqrt{p_r q_r}$$

- $\bar{Y}_i$ = Doğru yanıtlanan maddelerin ortalaması
- $\bar{Y}_r$ = r maddesini doğru yanıtlayanların ortalaması
- S_{Yr}= r maddesini doğru yanıtlayanların standart sapması
- p_r = r maddesinin güçlüğü
- q_r = 1- p_r

H^t İndeksi

Sijtsma (1986), bir maddenin Guttman (1944, 1950) modeline göre ölçeklenebilirliği hakkında bilgi sağlayan indeksi, birey düzeyine adapte ederek H^t indeksini geliştirmiştir.

H^t bir bireyin yanıt deseni ile kalan (n-1) yanıt deseni arasındaki uyumu ölçer (Mousavi, Cui & Rogers, 2019). Bu indekste veri matrisinin satırları toplam puana S_n(n=1,…,N) göre artan bir şekilde sıralandığında, birey uyumu aşağıdaki formülle hesaplanır.

$$H_n^t = \frac{\sum_{n \neq m}(t_{nm} - t_n t_m)}{\sum_{m<n}(t_m - t_n t_m) + \sum_{n<m}(t_n - t_n t_m)}$$

- S_n yanıt veren bireyin toplam puanı
- I madde sayısı
- $t_n = S_n / I_n$ n bireyinin doğru yanıtladığı maddelerin oranı
- $t_m = S_m / I_m$ m bireyinin doğru yanıtladığı maddelerin oranı
- t_{mn} hem m hem n bireyinin doğru yanıtladığı maddelerin oranı

H^t indeksi, çeşitli simülasyon çalışmalarında en iyi performans gösteren birey uyum indeksi olarak belirlenmiştir (Karabatsos, 2003; Sijtsma, 1986; Sijtsma & Meijer, 1992). H^t indeksi bireyin yanıt deseni uyumsuz ise 0'a, uyumlu ise 1'e yaklaşır.

U3 İndeksi

U3 indeksi (Emons, Meijer & Sijtsma, 2002; Van der Flier, 1980, 1982), maddeler azalan bir doğru yanıtlama oranına göre sıralanmış olduğunda, mükemmel Guttman vektörü için (sadece en kolay maddelere doğru yanıt verilen) 0, ters Guttman vektörü için 1 değerini alır. U3 indeksi 1'e doğru gittikçe yanıt desenlerinin uyumsuzluğu artmaktadır. U3 puanları doğru yanıt sayısına bağımlıdır.

$$U3_n = \frac{f(\mathrm{x}_\mathrm{n}^*) - f(\mathrm{x}_\mathrm{n})}{f(\mathrm{x}_\mathrm{n}^*) - f(\mathrm{x}_\mathrm{n}')}$$

U3 indeksi bir öğrencinin madde yanıt deseni ile Guttman modeli arasındaki uyumu ölçer (Mousavi vd., 2019).

C İndeksi

C indeksinde (Sato, 1975), bir test katılımcısının madde yanıt deseni mükemmel Guttman vektörü ise indeks değeri 0'a eşittir. Madde puan vektörü ile madde toplam vektörü arasındaki kovaryans sıfıra eşitse C indeksi 1'e eşit olur. Madde puan vektörü ile madde toplam vektörü arasındaki kovaryans negatif ise C indeksi 1'den büyüktür bu nedenle, birey için hesaplanan C indeks değeri büyüdükçe gruptaki diğer bireylerin yanıt davranışına kıyasla davranışının anormal olduğu ifade edilir. C indeksinin sabit bir üst sınırı olmadığından yorumlanması sorunlu olabilir. Bu nedenle Harnisch ve Linn (1981) tarafından C* indeksi önerilmiştir.

$$C_n = \frac{Cov(x_n, p)}{Cov(x_n^*, p)}$$

C İndeksi*

Harnisch ve Linn (1981) tarafından önerilen bu indeks Guttman vektörünün tersidir. Test katılımcısı ters Guttman vektörüne sahip olduğunda yani en zor maddelere doğru, en kolay maddelere ise yanlış yanıt verdiğinde C* indeksi 1 değerini alırken, birey mükemmel Guttman vektörüne sahip olduğunda ise 0 değerini alır. Bu sebeple 1'e yaklaşması anormal yanıt deseninin göstergesidir. Bu indeks bütün maddeleri doğru ya da yanlış yanıtlayan bireyler için hesaplanamaz.

$$C_n^* = \frac{Cov(x_n^*, p) - Cov(x_n, p)}{Cov(x_n^*, p) - Cov(x_n', p)}$$

- X_i ikili madde puanlarından oluşan rastgele değişken (i=1,2,…,I)
- X_{ni} madde i'deki n bireyinin gözlenen puanı
- p madde güçlük indeksi

- p_i (i maddesini doğru yanıtlayan bireylerin oranı)
- X_n* yalnızca en kolay öğeler (örneğin, en büyük p değerlerine sahip) için doğru yanıtlar içeren Guttman vektörü

Anlatılan birey uyum indekslerini açıklamak amacıyla Tablo 3'te 4 madde ve bu maddelere verilen yanıt desenleri verilmiştir. Her birey için H^t, U3 ve C* indeksleri hesaplanmıştır.

Tablo 3

H^t, U3 ve C İndekslerinin Hesaplandığı Örnek Bir Uygulama*

	Maddeler				Birey Uyum İndeksleri		
Birey	1 (kolay)	2	3	4 (zor)	C*	H^t	U3
A	1	1	1	1	---	---	---
B	1	1	1	0	0,00	0,27	0,00
C	1	1	0	0	0,00	0,25	0,00
D	1	0	0	0	0,00	0,27	0,00
E	0	0	0	0	---	---	---
F*	0	1	0	0	0,33	-0,09	0,33
G*	1	0	1	0	0,25	0,00	0,25
H*	1	0	0	1	0,50	-0,25	0,50
I*	0	1	1	1	1,00	-0,81	1,00
Güçlük	0.67	0.55	0.44	0.33	(*)Anormal yanıt desenini ifade etmektedir.		

Tablo 4 incelendiğinde maddelerin tamamı doğru veya yanlış yanıtlandığında indekslerin hesaplanamadığı görülmektedir. Bireyin yanıt desenindeki Guttman hataları arttıkça U3 ve C* artma eğilimi gösterirken; H^t indeksi negatif değerler alarak azalma eğilimi göstermiştir.

R Studio PerFit Kütüphanesi (Tendeiro, 2015) içerisinde bulunan parametrik olmayan indeksler tablo 4'te sunulmuştur. PerFit kütüphanesi yardımıyla programa yüklenen veri setinin özelliklerine uygun olan seçilerek birey uyum indeksi hesaplanabilmektedir.

Tablo 4

R Studio PerFit Kütüphanesinde Bulunan İstatistikler

Birey Uyum İstatistiği R fonksiyonu	Referans	Madde Tipi	Birey Uyum istatistiği Türü
r.pbis	(Donlon & Fischer, 1968)	İkili Puanlama	Parametrik Olmayan
C.Sato	(Sato, 1975)	İkili Puanlama	Parametrik Olmayan
G, Gnormed	(Van Der Flier, 1977; Meijer, 1994)	İkili Puanlama	Parametrik Olmayan
A.KB, D.KB, E.KB	(Kane & Brennan, 1980)	İkili Puanlama	Parametrik Olmayan
U3, ZU3	(Van Der Flier, 1980, 1982)	İkili Puanlama	Parametrik Olmayan
C*	(Harnisch & Linn, 1981)	İkili Puanlama	Parametrik Olmayan
NCI	(Tatsuoka & Tatsuoka, 1982,1983)	İkili Puanlama	Parametrik Olmayan
Iz	(Drasgow, Levine, & Williams, 1985)	İkili Puanlama	Parametrik
Izpoly	(Drasgow, Levine, & Williams, 1985)	Çoklu Puanlama	Parametrik
H^t	(Sijtsma, 1986)	İkili Puanlama	Parametrik Olmayan
Gpoly	(Molenaar, 1991)	Çoklu Puanlama	Parametrik Olmayan
Gnormed.poly	(Molenaar, 1991; Emons, 2008)	Çoklu Puanlama	Parametrik Olmayan
Izstar	(Snijders, 2001)	İkili Puanlama	Parametrik
U3poly	(Emons, 2008)	Çoklu Puanlama	Parametrik Olmayan

Hangi Birey Uyum İndeksi Kullanılmalıdır?

Yukarıda açıklandığı üzere çok sayıda birey uyum indeksi vardır. İndekslerin kullanışlılığı, anormal yanıt desenini tespit etme gücü ve tespitin doğruluğuna bağlıdır (Emons, 2008). Bu indekslerin gücünün simüle edilmiş veriler kullanarak karşılaştıran çalışmalar bulunmaktadır. Örneğin Karabatsos (2003) çalışmasında parametrik olmayan H^t indeksini, anormal yanıt gösteren bireyleri tespit etmede en iyi performans gösteren, C* ve U3 indesklerini de H^t indeksinden sonra en iyi performasa sahip indeksler olarak belirlemiştir. Tendeiro ve Meijer (2014) çalışmalarında ikili puanlama yapılan maddeler için farklı grup tabanlı indeksleri karşılaştırmış ve en yüksek güce sahip olan indeksin parametrik olmayan H^t ardından U3 ve C indeksleri olduğu sonucuna varmışlardır. İndekslerin gücünü simüle eden veya gerçek deneysel veride karşılaştıran birçok çalışma yapılmıştır (Birenbaum, 1985, 1986; Harnisch Linn, 1981; Kogut, 1987; Meijer, 1994; Nering & Meijer, 1998; Rudner, 1983) fakat anormal yanıt desenini tespit etmede hangi indeksin en iyi performans gösterdiği konusunda evrensel bir anlaşma yoktur.

İndekslerin gücünün yanında, birey uyum indeksi seçmek için başka bir kriter de teorik bir örneklem dağılımının var olup olmadığıdır, mevcut dağılıma uygun bir indeks seçmek esastır. U3 ve H^t'nin bilinen bir teorik dağılımı yoktur ve önceki araştırmalarda en çok atıf alan ve kullanılan iki indekstir (Emons vd., 2002; Karabatsos, 2003).

Birey Uyum Analizi Süreci

Meijer ve Sitjtsma (2001) ve Rupp (2013) tarafından birey uyum analizinin nasıl gerçekleştirileceği formüle edilmiştir.

1. İstatistiksel tespit adımı: Bu adım Mousavi ve Tendeiro (2016)'un önerdiği birey uyum analizi hesaplama prosedürüyle gerçekleştirilmiştir. Bu prosedüre göre;

a) Birey uyum indeksleri seçilir.

b) İndekslere bağlı olarak kesme değerleri seçilir.

c) Veriler için indeksler hesaplanır.

d) Kesme değerine bağlı olarak uyumsuz yanıt desenleri belirlenir.

2. Sayısal tablo adımı: Hesaplanan indeks sonuçlarının bir tablosu yapılarak bütünsel bakış sağlanır.

3. Grafiksel keşif adımı: Anormal yanıt deseni tespit edilen bireylerin ve normal yanıt deseni gösteren bireylerin madde yanıt grafikleri çizilerek yanıt davranışları görselleştirilir.

4. Nicel keşif adımı: Anormal yanıt deseni gösteren bireyler ile görüşmeler yapılarak sesli düşünme protokolleri uygulanır.

5. Nitel açıklama adımı: Anormal yanıt desenlerinin olası nedenleri açıklanır.

Anormal Yanıt Deseni

Ölçme sürecinde bireyin maddelere verdiği yanıtlar bilgi beceri ve yetenek düzeyiyle uyumlu olmalıdır. Belirli bir yetenek seviyesindeki bireyin maddelere verdiği yanıtlar sonucunda oluşan yanıt deseninin uygunluğu örneklemdeki benzer yetenek düzeyinde olan bireylerin yanıt desenlerine dayanarak belirlenmektedir. Bireyin sahip olduğu yanıt deseni ve toplam puanı, beklenen yanıt deseni ve puanına uygun değilse anormal olarak sınıflandırılmaktadır. Örneğin; düşük yetenek düzeyine sahip bir test katılımcısının, birkaç zor maddeyi doğru cevaplamış olması, beklenmedik bir durumu temsil etmektedir. Buradaki ilk zorluk, yanıt desenleri incelenirken, beklenti düzeyinden sapma derecesine yönelik bir karara varabilmektir. Bu karara varabilmek içinse öncelikle beklenti düzeyinin tanımlanması gerekir. Bu beklenti düzeyini tanımlayabilmek ve nihayetinde anormal yanıt desenlerini belirlemek amacıyla, çok sayıda birey uyum indeksi geliştirilmiştir (Der Flier, 1997; Guttman, 1944,1950; Harnisch & Linn, 1981; Kane & Brennan, 1980; Karabatsos, 2003; Meijer, 1994; Sato, 1975; Sijtsma, 1986; Sijtsma & Meijer, 1992; Tatsuoka & Tatsuoka, 1983; Van Donlon & Fischer, 1968). Geliştirilen birey uyum indeksleri anormal yanıt desenlerini tespit etmeye yardımcı olmaktadır fakat anormal yanıt davranışının nedeni

hakkında kesin bilgiler vermemektedir. Örneğin kişisel, bilişsel ve psikolojik davranışlar bireyin yanıtlama sürecini etkilemiş olabilir ve farklı davranış türleri benzer türde madde yanıt desenleri üretebilir (Meijer, Muitjens & van der Vlueten, 1996). Maddeleri kolaydan zora doğru dizilmiş bir testte birey ilk sıralardaki kolay maddeleri yanlış cevaplandırırken takip eden daha zor maddeleri doğru cevaplandırdığında bireyin yanıt deseni anormal olarak nitelendirilir. Bu davranışının sebebi bireyin hile yapması veya kopya çekmesi olabileceği gibi, birey sınavın ilk anlarından gergin veya heyacanlı olduğundan başlangıç sorularını yanlış cevaplamış da olabilir. Anormal yanıt deseninin bir çok oluşma sebebi bulunmaktadır bu sebeplerden bazıları aşağıda sunulmuştur.

Anormal Yanıt Desenlerinin Oluşma Sebepleri

Meijer (1996), anormal yanıt desenlerine neden olabilecek 7 farklı davranış biçiminden ve bu davranışların, yanıt desenlerinde kendini nasıl gösterebileceğinden söz etmiştir.

1. Uykusuzluk
2. Şanslı tahmin
3. Hile / Kopya çekmek
4. Yetiştirememe
5. Hizalama hataları
6. Yaratıcı düşünerek yanıt verme
7. Hazırbulunuşluk eksikleri

Uykusuzluk; birey uykulu olduğunda soruları dikkatsiz okuyacağından, bazı kolay maddelerde beklenenden fazla sayıda hata yapar ve bu durum anormal yanıt deseninin oluşmasına sebep olur.

Şanslı Tahmin; birey orta zorluktaki maddeler için tahminler yapar. Bu tahminler sonucu, zor maddeleri doğru yanıtlarken, nispeten daha kolay olan maddeleri yanlış yanıtlayabilir.

Bu durum anormal yanıt desenini meydana getirir. Tahmin davranışı skolastik doğası nedeniyle test puanlarını fazlaca etkilememektedir. Test katılımcısı, bazı maddeleri doğru tahmin ederken, bazılarını yanlış tahmin etmektedir. Bu nedenle test puanları için önemli ölçüde farklılaşmaya sebep olmayabilir (Meijer, 1996).

Kopya Çekmek; birey sınava giren bir başkasının cevaplarının bir kısmını kopyaladığında, bir kısmını ise kendi cevaplandırdığında yetenek düzeyiyle uyumsuz, anormal yanıt deseni oluşturur (Belov & Amstrong, 2010).

Yetiştirememe; birey doğru yanıtladığından emin olana kadar bir sonraki maddeyi cevaplamayı reddeder. Bu durum ona vakit kaybettirdiğinden, sınavdaki tüm maddeleri cevaplamaya vakti yetmez. Tüm maddeleri cevaplandırmaması ise anormal yanıt deseni meydana getirir.

Hizalama Hataları; sınavın formatı gereği bireyin cevaplarını kaydetmesi için ayrı bir cevap kâğıdı verildiğinde, birey sınav kâğıdında boş bıraktığı maddenin cevabını cevap kâğıdında boş bırakmayı unutursa sonraki maddelerin cevaplarını kaydırdığı için bu durum anormal yanıt desenine yol açar.

Yaratıcı Düşünerek Yanıt Verme; birey testteki kolay maddeler üzerinde gereğinden fazla düşünerek bulduğu sonucun cevap olamayacak kadar basit olduğunu düşünüp yanlış cevap verir. Yüksek yetenek düzeyine sahip birey kolay maddeleri yanlış cevaplandırdığında anormal yanıt deseni oluşur.

Hazırbulunuşluk Eksikleri; madde zorluk seviyeleri belirli bir içerikle örtüştüğünde bu içeriğin alt yetenek düzeylerinde eksiklik olanlar (yani konunun belli bir kısmı hakkında daha az bilgiye sahip olanlar) alt yetenek düzeyindeki kolay maddeleri doğru yanıtlayamazken üst yetenek düzeyindeki zor maddeleri doğru cevaplayarak anormal yanıt desenleri meydana getirir. Bu davranışların oluşturması beklenen anormal yanıt desenleri Tablo 5'te özetlenmiştir.

Tablo 5

12 Maddelik Bir Testte Anormal Yanıt Desenleri

	Maddeler											
Davranış	1	2	3	4	5	6	7	8	9	10	11	12
Uykusuzluk	0	0	0	1	1	1	1	1	1	1	0	1
Şanslı Tahmin	1	1	1	1	0	0	1	0	0	0	0	1
Kopya Çekmek	1	1	0	1	0	1	0	0	0	1	1	1
Yetiştirememek	1	1	1	1	1	1	0	0	0	0	0	0
Hizalama Hataları	1	1	1	1	1	0	1	1	0	0	0	0
Yaratıcı Düşünerek Yanıt Verme	0	0	0	0	1	1	1	1	0	1	1	1
Hazırbulunuşluk Eksikleri	0	0	1	0	1	1	1	0	1	1	1	0
p	,90	,85	,83	,82	,57	,55	,50	,49	,30	,25	,21	,15

Meijer, R. R. (1996). Person-fit research: An introduction. *Applied Measurement in Education*, *9*(1), 3-8.

Meijer (1996) tarafından sağlanan örnek yanıt desenlerinin dikkate alınması önemlidir. Sınav maddelerinin artan zorluk sırasına göre sınava girenlere sunulduğu varsayıldığında baştaki maddeler en kolay ve kolay sonlara doğru olan maddeler ise zor ve en zor olarak sıralanmış olsun. Gerçek hayatta sınava giren bireylerin azımsanamayacak kadarı verilen sürede tüm maddelere cevap vermeyi yetiştirememektedir. Bunun yanı sıra maddelerin kolaydan zora doğru sıralanan bir sınav uygulamasının günlük yaşamda gerçekleşmesi zordur ve bu durum anormal yanıt desenleri hakkında yapılan yorumları sınırlandırır.

Meijer (1996) açıkladığı 7 farklı davranış biçimine ek olarak anormal yanıt desenlerine neden olabilecek davranışları araştıran farklı çalışmalar derlenmiştir.

Levin ve Rubin (1979), çalışmasında anormal yanıt deseninin nedeninin; madde köklerini yanlış okumak ve cevapları optiğe kodlarken hata yapmak -örneğin testteki madde 10'u

cevaplar için verilmiş olan optikte madde 11'e kodlamak ve devamındaki maddeleri de bu sebepten dolayı kaydırmak olduğu sonucuna varmıştır.

Okuma becerilerinin eksikliği ve yorumlama sorunları; Meijer, Egberink, Emons ve Sijtsma (2008) ve Meijer & Tendeiro (2014) çalışmalarında farklı etnik grupların katıldığı bir sınavda; sınavı ana dilinde almayan bazı öğrencilerin okuma veya yorumlama eksikliği nedeniyle anormal yanıt davranışı ortaya çıkarabileceğini ifade etmiştir.

Meijer, vd., (2008), Harter'ın (1985) Çocuklar için Benlik Algılama Profili'nden elde edilen verileri, 8 ile12 yaş arası bir örneklemde analiz ettiğinde bazı çocuklar için madde puanlarının, daha az gelişmiş bir benlik kavramı ve sorulan maddelerin anlamını anlamamaları nedeniyle özellik düzeylerini yeterince yansıtmadığını görmüştür. Ayrıca, Meijer ve Tendeiro (2014), yüksek riskli bir eğitim testinde yetersiz İngilizce bilgisinin sonucunda farklı etnik gruplar için anormal yanıt desenleri bulmuşlardır.

Test Alma Motivasyonu; Ferrando (2012), çalışmasında Nevrotiklik ve Dışadönüklük Ölçeği'ni dolduran bazı bireylerin yanıt deseninin rastgele olduğunu bulmuştur ve bu durumun testi alma motivasyonunun düşük olmasından kaynakladığı sonucuna ulaşmıştır. Niessen, Meijer ve Tendeiro (2014), araştırmaları için veri toplarken, bazı öğrencilerin yanıt desenlerinin, test alma motivasyonlarının düşük olmasının bir sonucu olarak rastgele olduğunu ve bunun da anormal yanıt deseniyle sonuçlandığını bulmuşlardır.

Birey Uyum İstatistiklerinin Eğitim Bağlamında Kullanımı

Eğitim ortamlarında sınavlar yoluyla öğrencilerin başarı düzeyleri ölçülerek, elde edilen ölçümler sonucunda kararlar alınır. Öğrenciler hakkında doğru eğitim kararlarının verilebilmesi yapılan doğru ölçümlerle mümkündür. Bu ölçümler, bir eğitim derecesi alacak nitelikli öğrencilerin belirlenmesine, öğrenme ilerlemelerinin izlenmesine, eğitimcilerin müfredatlarını verimli bir şekilde uyarlamalarına, politika yapıcıların önemli eğitim sorunlarına karar vermelerine yardımcı olabilir. Öğrencilerin sınav sonucunda aldığı toplam puanlar, öğrencilerin yeterlilik düzeyi hakkında uygulayıcıya bilgi sağlaması amacıyla rapor

edilirken bazı durumlarda toplam puanlar sınav katılımcısının yeterlilik düzeyi hakkında yanlış bir izlenim verebilir ve olumsuz sonuçlara yol açabilir (Karabotsos, 2003).

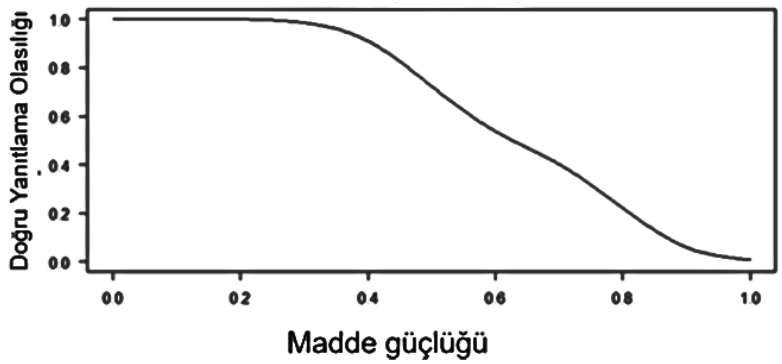
Sınav katılımcıları bazı test sorularını yanlış yorumlayarak, ön bilgileri yetersiz olmasına rağmen şanslı tahminler yaparak veya kopya çekerek anormal yanıtlama davranışı sergilerler. Buna bağlı olarak da geçersiz puanlar alırlar. Bu durumun anlaşılabilmesi için testlerin bir anlamda kalite kontrol araçları yardımıyla geçerlikleri rutin olarak kontrol edilmelidir. Bu kontrol araçlarından biri olan birey uyum indekslerinin hesaplanması ve puanlarının raporlanması tavsiye edilmektedir (Tendeiro & Meijer, 2012).

Eğitimsel ve psikolojik ölçümde, test geliştirme ve analizinde kullanılan birçok birey uyum indeksi önerilmiştir. Bu indeksler yardımıyla gözlemlenen yanıt deseni, ölçme sürecinde kullanılan ölçüm modeliyle veya örneklemdeki diğer öğrencilerin yanıt desenleriyle uyumsuz olan öğrencileri belirlenir (Cui & Li, 2015). Gözlemlenen yanıt deseni öğrencinin yetenek seviyesiyle tutarsız olduğunda, bireyin anormal yanıtlama davranışında bulunduğu ifade edilebilir. Fakat bir sınav oturumunda yüksek yetenek düzeyine sahip bir öğrencinin yanına oturan düşük yetenekli öğrenci tüm cevapları arkadaşından kopyaladığında, yanıt deseni gerçek yetenek seviyesiyle uyumsuz olmasına rağmen gözlemlenen yetenek seviyesiyle uyumludur. Bu birey uyum analizinin bir sınırlılığıdır. İndeksler yanıt deseni, gözlemlenen yetenek seviyesiyle tutarsız olduğu sürece doğru tespitler yapar (St-Onge, Valois, Abdous & Germain, 2011).

Madde Yanıt Grafikleri (Person Response Function Plot [PRF])

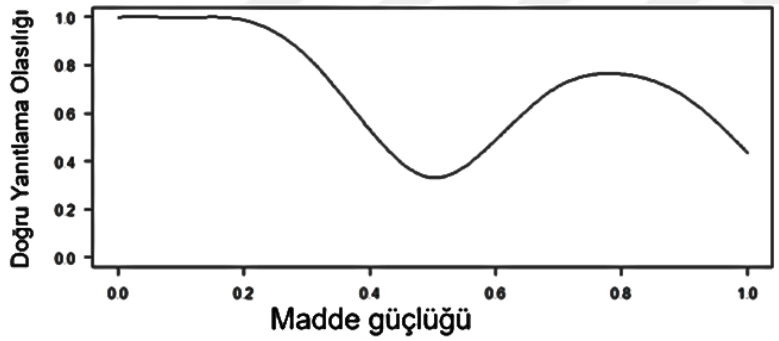
Trabin ve Weiss (1983), anormal madde yanıt desenlerini belirleyebilmek için bireyin yanıt deseninden oluşan bir tepki fonksiyonu yazmayı önermiştir. Bu fonksiyon, maddelerin güçlüklerine göre kolaydan zora doğru sıralanarak bireylerin sergiledikleri yanıt desenlerini bir fonksiyon olarak görselleştirilip incelemek ve yorumlamak için geliştirilmiş sezgisel bir yöntemdir. Bireylerin test puanlarından hareketle yapılacak çıkarımlara katkı sağlamaktadır.

Şekil 1 beklenen yanıt desenine bir örnek olarak verilebilirken, Şekil 2 anormal yanıt desenine sahip bireye ait bir yanıt fonksiyonudur.



Şekil 1. Beklenen yanıt deseni

Şekil 1'deki yanıt desenine sahip bireyin kolay maddelere doğru cevap verme olasılığı 1 iken sorular zorlaştıkça doğru yanıtlama olasılığı azalmış ve 0'a yaklaşmıştır.



Şekil 2. Anormal yanıt deseni

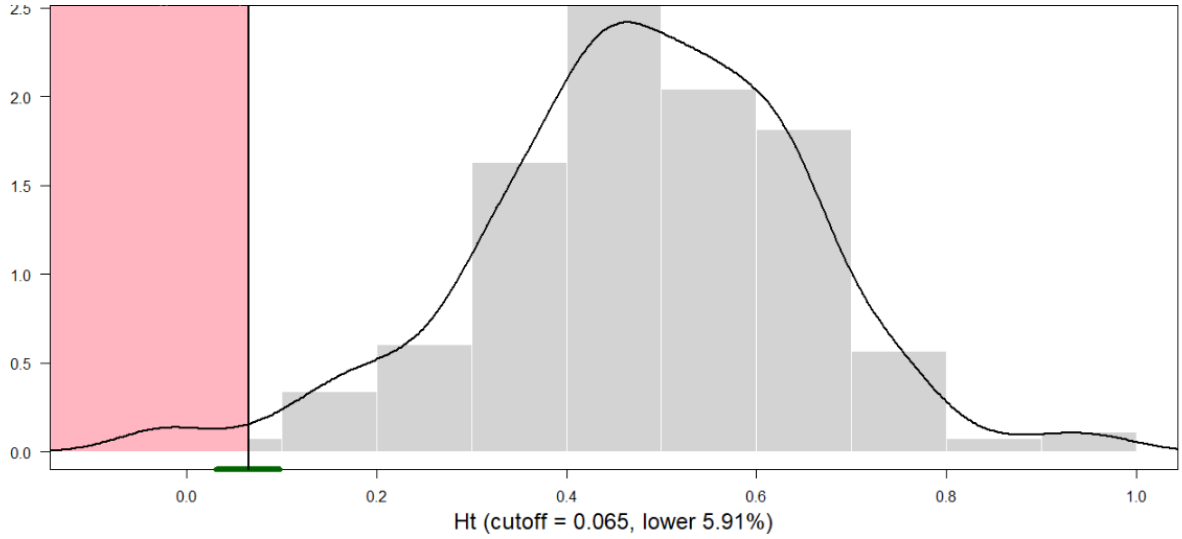
Şekil 2'deki yanıt desenine sahip olan bireyin kolay sorulara doğru cevap verme olasılığı 1 iken sorular zorlaştıkça doğru yanıt verme olasılığı azalmış fakat madde güçlüğü 0,50 civarında olan maddeleri doğru yanıtlama olasılığı tekrar artmaya başlamış ve 0,80 civarında yeniden azalışa geçmiştir. Doğru cevaplama olasılığındaki bu dalgalanma bireyin yanıt deseninin anormal olarak sınıflandırılmasına sebep olmuştur.

Kesme Puanı

Birey uyum analizi yapılırken, bireylerin indeks puanı hesaplandıktan sonra belli bir kesme puanının altında veya üstünde puana sahip bireylerin yanıtları anormal olarak sınıflandırılmaktadır. H^t indeksi için kesme değerinden daha küçük, U3 ve C* indeksleri için ise kesme değerinden daha büyük puana sahip bireyler tespit edilmektedir.

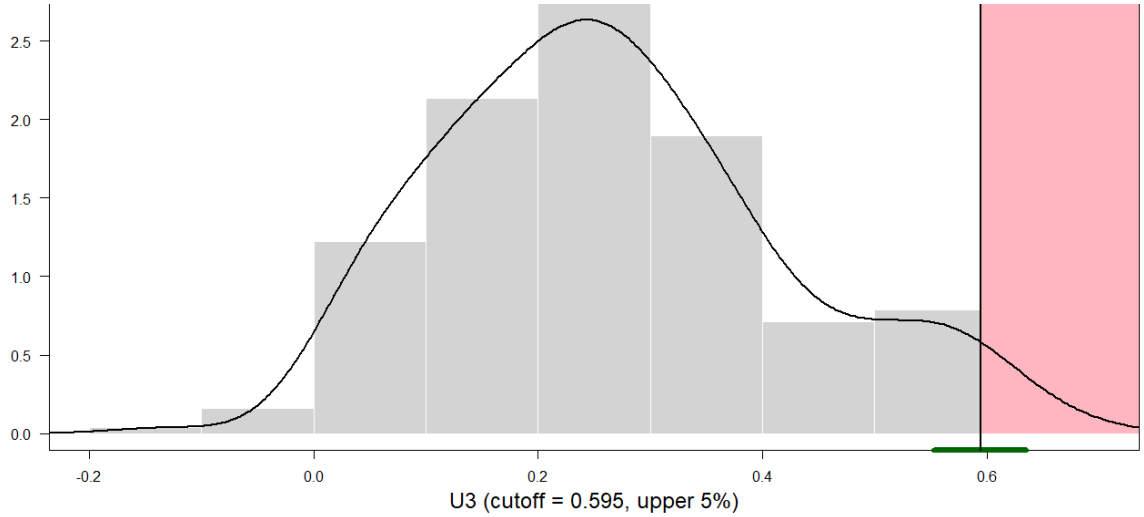
Literatürde indekslerin kesme puanları için kesin bir görüş yoktur (Uçar, 2021). Her bir indeks için farklı kesme puan değerleri önerilmiştir. Örneğin C için 0,5 önerilmiş olup bu değerden büyük olan indeks puanları anormal olarak sınıflandırılmıştır (Miller, 1996; Sato, 1975). C* indeksi için ise 0,3 kesme puanı önerilmiş ve puanları bu değerden büyük olan bireyler anormal olarak sınıflandırılmıştır (Harnisch & Linn, 1981). U3 indeksi için kesme puanı 0,5 olarak önerilmiş bu değerden yüksek puanlar anormal olarak sınıflandırılmıştır. H^t indeksi için ise, Sijtsma ve Meijer (1992) 0,3 kesme puanını önerirken Karabatsos (2003) daha iyi sonuçlar için 0,22 kesme puanının kullanılması gerektiğini ifade etmiştir. Fakat güncel çalışmalar her örneklem için farklı kesme puanlarının belirlenmesini önermektedir (Mousavi, Cui & Rogers, 2019). Felt, Castaneda, Tiemensma ve Depaoli (2017), çalışmalarında sabit bir kesme puanı belirlemek yerine indeks puanlarına ait histogram çizilerek, çizilen histogramın üzerinde veri setine uygun bir kesme puanının belirlenmesinin daha doğru bir yöntem olacağını belirtmiştir.

R studio programında kesme puanları örnekleme göre 0,05 anlamlılık düzeyinde 'cutoff' fonksiyonu yardımıyla tahmin edilmektedir. Kesme puanları örnekleme göre tahmin edildiğinden Rstudio programına yüklenen veri seti için, o örnekleme uyan bir kesme puanı belirlenmektedir. Aşağıda programın H^t, C* ve U3 indeksler için tahmin ettiği kesme puanları sırasıyla Şekil 3, 4 ve 5'te ifade edilmiştir.



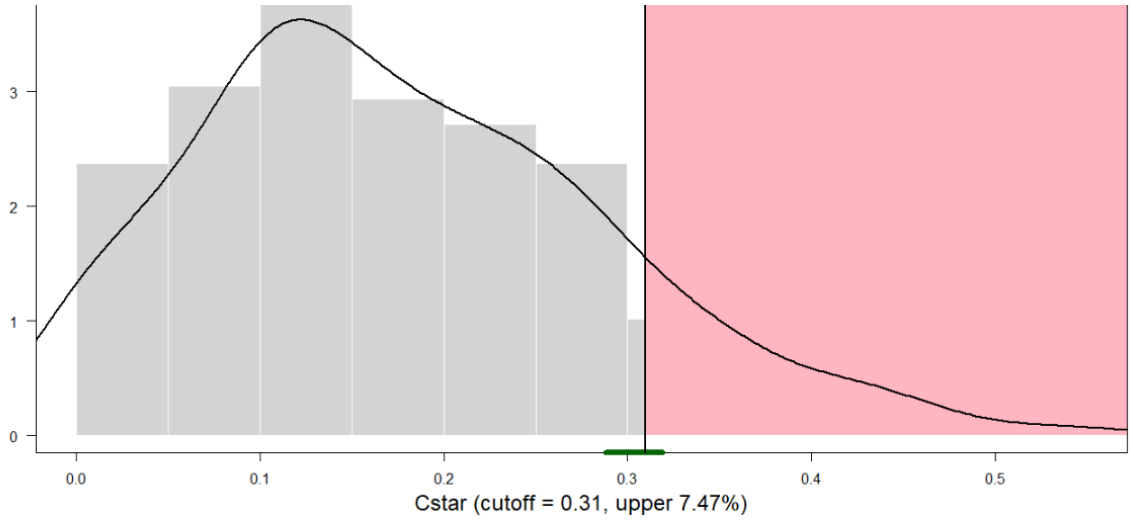
Şekil 3. H^t İndeksi kesme puanı (R studio)

Şekil 3, hesaplanan H^t puanlarının dağılımını göstermektedir. Kesme puanı dikey çizgi ile ifade edilmiştir. Dağılımın sol tarafındaki taranmış alan, anormal yanıt desenlerini belirleyen değer aralığını göstermektedir. Kesme puanını güven aralığı dikey çizginin etrafındaki yeşil işaretleyiciyle belirtilmiştir. Kesme puanından küçük olan H^t puanları anormal yanıt deseni olarak işaretlenmektedir.



Şekil 4. U3 İndeksi kesme puanı (R studio)

Şekil 4, hesaplanan U3 puanlarının dağılımını göstermektedir. Kesme puanından büyük olan U3 puanları anormal yanıt deseni olarak işaretlenmektedir.



Şekil 5. C* İndeksi kesme puanı (R studio)

Şekil 5, hesaplanan C* puanlarının dağılımını göstermektedir. Kesme puanı dikey çizgi ile ifade edilmiştir ve literatürle uyumludur (Harnisch & Linn, 1981). Kesme puanından büyük olan C* puanları, anormal yanıt deseni olarak işaretlenmektedir.

Araştırmanın Amacı ve Önemi

Uluslararası çalışmalarda daha çok birey uyum indekslerinin istatistiksel gücüne yoğunlaşılmış, pratik uygulamalarda kullanılabilirliği görece geri planda kalmış (Emmen, 2011) ve yapılan birçok çalışma simülatif veriler üzerinden yürütülmüştür (Amstrong & Shi, 2009; Karabatsos, 2003; Meijer, 1994; Runder, 1983). Çalışmanın öncelikli amacı daha önce simülatif veri çalışmalarında performansı sınanmış (Avşar, 2019; Karabatsos, 2003; Rudner, 1983; Tendeiro & Meijer, 2014) parametrik olmayan H^t, U3 ve C* indekslerinin, gerçek veri üzerinde uygulanması ve bu veride anormal yanıt deseni gösteren bireylerin tespit edilmesidir.

Araştırmaya katılan öğrencilerin yanıt desenlerini, kamuoyu ile paylaşan geniş ölçekli testlerden PISA 2018 matematik uygulamasının veriler kullanılarak birey uyum analizi gerçekleştirilmiştir. PISA 2018 uygulamasına Amerika Birleşik Devletleri, Çin, Güney Kore, Japonya, Rusya ve Türkiye'den katılan ve uygulamadaki 24 paralel form arasından ilk

5 formu alan öğrencilerin yanıt desenleri üzerinde parametrik olmayan H^t, U3 ve C* indeksleri hesaplanmıştır. Hesaplanan indeks puanlarının ülkelere ve formlara göre farklılaşmasını incelenmiştir. İndeks puanları kesme puanına göre anormal olan öğrenci sayıları tespit edilmiş ülke ve formlara göre farklılaşmasını incelenmiştir. Çalışma, Rstudio Programı PerFit Kütüphanesi'nin birey uyum analizi sürecindeki kullanımı için pratik bir rehber sağlamıştır.

Araştırma Problemi

1) PISA 2018 araştırmasına Amerika Birleşik Devletleri, Çin, Güney Kore, Japonya, Rusya ve Türkiye'den katılan ve ilk beş formu alan öğrencilerin H^t birey uyum indeks puan ortalamaları farklılık göstermekte midir?
   a) Ülkelerin ve formların H^t indeks puan ortalamaları üzerindeki etkileşim etkisi nedir? Öğrencilerin H^t indeks puan ortalamaları ülkelere ve formlara göre farklılık göstermekte midir?
   b) H^t indeksi tarafından yanıt deseni anormal olarak belirlenen öğrenci sayıları ülkelere göre farklılık göstermekte midir?
   c) H^t indeksi tarafından yanıt deseni anormal olarak belirlenen öğrenci sayıları formlara göre farklılık göstermekte midir?
2) PISA 2018 araştırmasına Amerika Birleşik Devletleri, Çin, Güney Kore, Japonya, Rusya ve Türkiye'den katılan ve ilk beş formu alan öğrencilerin U3 birey uyum indeks puan ortalamaları farklılık göstermekte midir?
   a) Ülkelerin ve formların U3 indeks puan ortalamaları üzerindeki etkileşim etkisi nedir? Öğrencilerin U3 indeks puan ortalamaları ülkelere ve formlara göre farklılık göstermekte midir?
   b) U3 indeksi tarafından yanıt deseni anormal olarak belirlenen öğrenci sayıları ülkelere göre farklılık göstermekte midir?

c) U3 indeksi tarafından yanıt deseni anormal olarak belirlenen öğrenci sayıları formlara göre farklılık göstermekte midir?

3) PISA 2018 araştırmasına Amerika Birleşik Devletleri, Çin, Güney Kore, Japonya, Rusya ve Türkiye'den katılan ve ilk beş formu alan öğrencilerin C* birey uyum indeks puan ortalamaları farklılık göstermekte midir?

   a) Ülkelerin ve formların C* indeks puan ortalamaları üzerindeki etkileşim etkisi nedir? Öğrencilerin C* indeks puan ortalamaları ülkelere ve formlara göre farklılık göstermekte midir?

   b) C* indeksi tarafından yanıt deseni anormal olarak belirlenen öğrenci sayıları ülkelere göre farklılık göstermekte midir?

   c) C* indeksi tarafından yanıt deseni anormal olarak belirlenen öğrenci sayıları formlara göre farklılık göstermekte midir?

Sınırlılıklar

PISA 2018 matematik uygulaması 24 ayrı form üzerinden gerçekleştirilmiştir. PISA 2018 araştırmasına katılan 79 ülke arasından Amerika Birleşik Devletleri, Çin, Güney Kore, Japonya, Rusya ve Türkiye şeklinde seçilen ülkelerin ilk 5 formu alan öğrencileri analize dâhil edilmiştir.

Tanımlar

KTK (Klasik Test Kuramı): KTK, gerçek puanı, ölçme sonuçlarından elde edilen puanla tahmin etmeye çalışan bir kuramdır.

Gözlenen puan (X) = Gerçek puan (T) + Hata puanı (E) üzerine kurgulanmıştır (Crocker & Algina, 1986).

BÖLÜM II

İLGİLİ ARAŞTIRMALAR

Bu bölümde birey uyum indeksleri ile ilgili alan yazında daha önce gerçekleştirilmiş çalışmalara yer verilmiştir.

İlgili Araştırmalar Özet

Birey uyum indeksleri 20.yy başlarından (Cronbach, 1946; Fowler, 1954; Glaser, 1949, 1950, 1951, 1952; Guttman 1944, 1950; Mosier, 1940; Sherif & Cantril, 1945, 1946; Spearman, 1910; Thurstone, 1927) beri çalışılan bir konudur. Literatür incelediğinde indekslerin gücünün sınandığı araştırmaların yoğun olduğu görülmektedir (Drasgow, Levine & Willams, 1997; Harnisch & Linn, 1981; Meijer, 1994; Meijer vd., 1996; Reise & Due, 1991; Rudner, 1983).

Rudner (1983) simülatif verilerle gerçekleştirdiği çalışmasında parametrik U, W ve l_o indeksleri ile parametrik olmayan r_{pbis}, r_{bis}, NCI ve C* indekslerini karşılaştırmıştır. Parametrik olmayan indeksler arasında yüksek korelasyonlar bulmuştur. Manipüle edilen faktörler test uzunluğu (I=45, 80) ve anormal yanıt davranışının türü (anormal olarak düşük veya yüksek) şeklindedir. C* indeksinin parametrik indekslerden daha iyi performans gösterdiği sonucuna varmıştır.

Harnisch ve Linn (1981), okuma ve matematik testlerinden elde edilen deneysel verileri kullanarak C, C*, rpbis, Ai, Di, Ei ve NCli indekslerini hesaplamış ve bu birey uyum indeksleri arasındaki korelasyonları incelemiştir. Harnisch ve Linn her iki test için

hesaplanan indeksler arasındaki korelasyonların Ai dışındaki indeksler için 0,65 ve 0,90 arasında değişmekte olduğunu bulmuşlardır. Ai indeksi ise diğerlerinin her biriyle yaklaşık olarak 0,40 korelasyon göstermiştir. Çalışmaları sonucunda ortalama C* puanları arasında farklı okul türlerinde, öğrenciler arası farklılaşmalar bulmuşlar ve bu farklılaşmanın okulların eğitim sürecinde kullandıkları müfredattan kaynaklanabileceğini belirtmişlerdir.

Meijer (1994) çalışmasında U3, U1 ve U1'in normsuz versiyonunun performansını karşılaştırmıştır ve çalışmanın sonunda, madde ayırıcılığı, anormal yanıt davranışının türü (şanslı tahmin, hile, rasgele yanıtlama) ve test uzunluğu gibi faktörlerin indekslerin tespit oranlarında değişikliğe sebep olduğunu bulmuştur. Karşılaştırdığı U3, U1 ve U1'in normsuz versiyonunun performasının birbirine benzer olduğunu tespit etmiştir.

Meijer, Muijtjens, ve Van der Vlueten (1996), çalışmalarında parametrik olmayan C, C* ve ZU3 indekslerinin anormal yanıt desenini algılama performanslarını karşılaştırmış ve C* ve ZU3 indekslerinin neredeyse aynı güçte olduğu sonucuna ulaşırken C'nin daha az algılama gücünün olduğunu ifade etmişlerdir. Ayrıca deneysel örneklerde tanı aracı olarak birey uyum indekslerinin kullanılabileceğini vurgulamışlardır. Araştırmacılar, birey uyum indekslerinin test sonuçları altında gizlenen bir gizil özellikten fazlası olduğunu ve kişileri izlemek için başvurulabilecek bir yol olduğunu belirtmişlerdir.

Karabatsos (2003), 25'i parametrik 9'u parametrik olmayan 36 farklı birey uyum indeksinin simülasyon verileri üzerinde anormal cevap desenlerini tespit etme performanslarını farklı test uzunluklarında ve beş farklı anormal yanıtlama durumunda; kopya, yaratıcı yanıt verme, tahmin etme, dikkatsiz ve rastgele yanıt verme (cheaters, creative responding, guessing, careless ve random respndents) karşılaştırmıştr. H^t indeksini kopya çekme, yaratıcı yanıtlar verme, şanslı tahminde bulunma davranışı gösteren bireyleri tespit etmede en iyi performans gösteren indeks olarak, MCI ve U3 indesklerini de rastgele cevaplayanları en etkili tespit eden indeksler olarak belirlemiştir.

Emmen (2011), çalışmasında 2000 Hollandalı yetişkine uygulanan kişilik testi verilerine birey uyum analizi yapmıştır. Katılımcılar kişilik anketinde; dürüstlük, uyumluluk,

vicdanlılık, duygusallık, aşırılık ve değişime açık olma gibi altı alt faktörden oluşan Likert tipi anket maddelerine yanıt vermektedir. Anket soruları Likert tipi puanlandığından U3 indeksinin, genelleştirilmiş U3poly versiyonu kullanılmıştır. Analiz sonucunda tespit edilen anormal yanıt desenine sahip katılımcıların yanıt desenleri ayrıntılı incelenerek hangi alt faktörün anormal yanıt verme davranışı ile ilgili olduğunu bulmak için regresyon analizi yapılmış ve korelasyonlar hesaplanmıştır.

Tendeiro ve Meijer (2014) çalışmalarında, beklenen desene göre anormal yanıt desenlerini tespit etmek için kullanılabilecek farklı yaklaşımlardan söz etmiştir. Parametrik olmayan indekslerin avantajı, verilerle ilgili katı model varsayımlarına dayanmamaları ve hesaplarının görece kolay olmasıdır. Fakat daha önceki araştırmalardan, değişen test koşullarında indekslerin gücünü belirlemeye yönelik simülasyon çalışmaları yapılmıştır. İndeksler, simüle edilmiş verileri kullanarak anormal yanıt desenlerini algılama gücüne göre karşılaştırılmıştır. Uygulayıcıların farklı indeksler arasında seçim yapmalarına yardımcı olabilecek bazı yönergeler sağlanmıştır. Genel olarak, U3, C* ve U1 tarafından takip edilen H^t indeksinin, anormal yanıt desenlerini tespit etmede en yüksek performansı gösterdiği bulunmuştur. Test uzunluğuyla madde ayırt ediciliğinin indekslerin yanıt desenlerinde anormalliği algılama gücü üzerinde orta ila büyük etkilere sahip olduğundan söz edilmiştir.

Cui ve Li (2015), araştırmalarında bilişsel ve tanısal değerlendirme için birey uyum analizinin önemli bir araştırma alanı olduğunu çünkü anormal yanıt davranışlarının tespit edilmemesinin öğrenci niteliklerinin yanlış yorumlanmasına neden olacağını ifade etmiştir. Çalışmalarında ilk olarak madde yanıt teorisi çerçevesinde geliştirilmiş olan I_z indeksi yardımıyla hem gerçek hem de simülatif verilerle birey uyum analizi gerçekleştirilmiştir. Birey uyum analizinin, bilişsel tanılar yapmak için de kullanılabileceğini ifade etmişlerdir. Son olarak da yeni bir birey uyum indeksi olan RCI'yi tanıtmışlardır.

Mousavi, Tendeiro ve Younesi (2016), çalışmalarında birey uyum indekslerinin kullanımını daha yaygın ve erişilebilir duruma getirmek amacıyla yazılmış olan R Studio programında yüklü Perfit kütüphanesi kullanılarak birey uyum analiz sürecinin nasıl gerçekleşeceğini ifade etmiş, ardından İran'da bulunan lise öğrencileri üzerinde yapılan psikolojik bir

çalışmada 430 kişiye uygulanmış olan kişilik ve zeka testi verileri üzerinden H^t birey uyum indekslerini hesaplamıştır. Düşük düzeyde nevrotik olan kişilerin ve nispeten daha yüksek IQ puanına sahip kişilerin anormal tepki verme davranışının diğer bireylere göre daha düşük olduğunu bulmuşlardır.

Tendeiro, Meijer ve Niessen (2016) çalışmalarında birey uyum analizleri için PerFit R kütüphanesinin kullanımından söz etmiş çeşitli analizler için kütüphanenin kullanımı konusunda pratik bir rehber sağlamıştır.

Avşar (2019), çalışmasında parametrik ve parametrik olmayan birey uyum indekslerinin gücünü karşılaştırmak için simüle edilmiş veri setlerini kullanmıştır. Tip-I hata oranı arttıkça indekslerin tespit oranları artma eğilimindedir ve bu duruma ek olarak test uzunluğu arttıkça indekslerin tespit oranları da artmıştır. Simüle edilmiş koşulların genelinde parametrik olmayan indekslerin gücü, parametrik indeksleri geçmiştir.

Avsar ve Emons (2021), PISA 2015 anket verisi üzerinde öncelikle SIBTEST ile DIF hesaplamaları yapmış ardından parametrik olmayan birey uyum indekslerinden G_N^p ve $U3^p$ indeksleriyle uyum göstermeyen öğrencileri belirlemiştir. Belirlenen öğrencilerin DMF üzerinde etkisi olduğu sonucuna ulaşmışlardır.

Yapılan literatür taramasında birey uyum indekslerinin özellikle, simülatif olarak üretilen veriler üzerinde çalışıldığı, PISA verileri üzerinde yapılmış birey uyum indeks çalışmalarının sınırlı olduğu gözlenmiş ve özellikle büyük ölçekli gerçek veriler üzerinde daha fazla araştırılma yapılması gerektiği görülmüştür. Birey uyum indekslerinin simülatif veriler üzerinden gücünü sınayan birçok çalışmada araştırma sonuçlarının gerçek veri setinde uygulanabilirliğinin önemi vurgulanmıştır. Örneğin, birey uyum indekslerinin anlamlılığını test etmek üzere önerilen teorik dağılımın gerçek veri setinde gözlenip gözlenmediği araştırılmıştır (Meijer, 2003; Nering, 1995; Rupp, 2013).

Aydoğan (2021), çalışmasında birey uyum indekslerinin testin psikometrik özelliklerine etkisini incelemiştir. İlk olarak uluslarası bir uygulama araştırması olan TIMSS başarı testinde kullanılan ikili puanlama yapılmış maddelerden oluşmuş veri setinde birey uyum

indekslerini hesaplamış ve öğrencilerin yaklaşık %6'sının anormal yanıt desenine sahip olduğu tespit edilmiştir. Anormal yanıt desenine sahip öğrenciler analizden çıkarıldığında, testin tek boyutluluğa daha iyi uyum sağladığı tespit edilmiştir. Daha sonra birey uyum indesklerinin testin psikometrik özelliklerine etkisini incelemek amacıyla birçok alt koşul altında (1PLM ve 2PLM ile uyumlu veri; "b" parametresi (-1,5/+1,5) ve (-3/+3) ranjı; "a" parametresi ortalaması 1; 1,5 ve 2 uyumsuzluk yüzdesi %5, %10 ve %20; 6 farklı uyumsuz yanıt türü) simülatif veriler üretilmiştir. Anormal yanıt desenlerinin "b" parametre ranjı genişledikçe daha rahat ikinci bir boyut, birinci boyutta eksi faktör yükleri ve fazlalığının bağımlı madde çifti oluşturduğu tespit edilmiştir.

Avcu (2022), çalışmasında birey uyum analizi ve uç yanıt stili arasındaki ilişkiyi incelemiştir. Bu doğrultuda 727 üniversite öğrencisine Temel Kişilik Özellikleri Envanteri uygulamış ve elde edilen veri seti üzerinde parametrik birey uyum indekslerinden l_z indeksini hesaplamış ve uyumsuz bireylerin oranını %10 olarak tespit etmiştir. Anormal yanıt desenine sahip öğrencilerin, yanıt desenleri ayrıntılı olarak incelendiğinde uç yanıtlama stilini yanıt deseni normal olan öğrencilere göre daha yoğun şekilde kullanma eğiliminde oldukları bulunmuştur.

BÖLÜM III

YÖNTEM

Yöntem

Öncelikle analize dâhil edilen ülkelere ait betimsel istatistikler sunulmuştur. Ardından veri setini tanımak amacıyla formların ülkeler bazında aritmetik ortalama ve standart sapma değerleri hesaplanmıştır. Formaların iç tutarlılığını analiz etmek için Cronbach Alfa güvenirlik katsayıları hesaplanmıştır. Formların ülkelere göre ortalama güçlükleri ve formlara ait soruların madde güçlük parametreleri hesaplanmıştır.

Birey uyum analizi için Guttman ölçeğine dayalı parametrik olmayan indekslerden ikili puanlama yapılmış maddelerde kullanılan ve çeşitli simülasyon çalışmalarında iyi performans gösteren (Karabatsos, 2003; Sijtsma, 1986; Sijtsma & Meijer, 1992) H^t, U3 ve C* indeksleri ve elde edilen indeks puanları arasındaki korelasyonlar hesaplanmıştır. İndeks puanlarına ait aritmetik ortalama ve standart sapma puanları tablolaştırılmıştır. Bu üç indeksin sonuçlarına göre anormal yanıt desenine sahip öğrenciler belirlenmiş, ülkeler ve formlar bazında sayıları sunulmuştur. Anormal yanıt desenine sahip öğrencilerin H^t, U3 ve C* indekslere göre dağılımı verilmiştir. İndekslerin tespit ettiği öğrencilerin yanıt desenleri tablolaştırılmış ve PRF grafikleri yardımıyla görselleştirilmiştir.

Farklı ülkelerden sınava giren ve farklı formları alan öğrenciler için hesaplanan birey uyum indekslerinin ortalamaları üzerindeki ülke ve form etkileşimini belirlemek amacıyla iki yönlü ANOVA yapılmıştır. Birey uyum indeks puanlarının ülkelere ve formlara göre de farklılaşıp faklılaşmadığı da yani ana etkiler de değerlendirilmiş ve etki büyüklükleri

raporlanmıştır. Ayrıca, birey uyum indesklerinde farklılaşma varsa farkın hangi ülkeler veya formlar arasından olduğunu tespit etmek amacıyla Post-Hoc testleri de yapılmıştır. Son olarak, her bir indekse göre anormal yanıt deseni gösterdiği tespit edilen birey sayıları ile ülkeler arasında istatistiksel açıdan anlamlı bir ilişki olup olmadığını belirlemek amacıyla χ^2 ilişki testi yapılmış ve sonuçlar raporlanmıştır.

Evren

Araştırmanın evrenini; PISA'nın hedef kitlesi olan, ülkeler içinde yer alan eğitim kurumlarına devam eden ve yedinci sınıf ve üzeri olan, 15 yaş üç ay ve 16 yaş iki ay arasındaki öğrencilerdir. PISA 2018 uygulamasına katılan 37'si OECD üyesi olan 79 ülkeden yaklaşık 600.000 öğrencidir.

PISA 2018 araştırmasında matematik testi 24 ayrı alt forma ayrılmıştır. Bu alt formların her birini alan öğrenciler yaklaşık 20 soru cevaplamıştır. Bu sorular ölçtükleri beceriler bakımından hiyerarşik bir yapıda hazırlanmıştır. İkili puanlama yapılmış maddeler 0-1 olarak kodlanmıştır.

Örneklem

Araştırmanın örneklemi amaçlı örnekleme yöntemi ile seçilmiştir. Seçilen ülkelerin PISA 2018 uygulamasında matematik alanındaki puan sıralamasında üst ve orta sıralarda olmasına dikkat edilmiştir (OECD, 2020). Araştırma kapsamında belirlenen altı ülke: Amerika Birleşik Devletleri, Çin, Güney Kore, Japonya, Rusya ve Türkiye'dir. Bu ülkelerin PISA matematik testi başarı sıraları incelendiğinde; araştırmaya dört eyaletiyle (Pekin, Şangay, Jiangsu ve Zhejiang) katılan Çin 1. olmuştur. Japonya 6. ve Güney Kore ise 7. sıradadır. Son olarak Rusya 31., ABD 38. ve Türkiye 43. sırada bulunmaktadır.

Araştırmada 24 form arasından ilk 5 formdaki ikili puanlama yapılmış (0-1 ile kodlanmış) maddeler seçilerek analize dâhil edilmiştir.

Veri Analizi

PISA 2018 uygulamasına ait veriler OECD'nin sitesinde yayınlanmış ve http://www.oecd.org/pisa/data/2018database/ internet adresinden SPSS dosyası formatında indirilmiştir. Veri setinde ülke kodu yardımıyla veri seçme özelliği kullanılarak her bir ülke için farklı veri seti dosyaları oluşturmuştur. PISA 2018 uygulaması matematik okuryazarlığı testi ilk beş formu; form-1 ve form-4 17 madde, form-2 ve form-3 18 madde, form-5 ise 16 adet madde içermektedir. Teste giren öğrencilerden test maddelerinden en az birini, sınav süresi yetmediğinden veya herhangi başka bir nedenden dolayı göremediği için cevaplayamayanlar ile maddeyi boş bırakanlar, maddenin doğru cevabını bilmedikleri için doğru cevaplamadıkları varsayılarak "0" olarak kodlanmıştır. Ülkelerin formu alan öğrenci sayıları ve toplam puanları hakkında betimsel istatistikler IBM SPSS Statistics 23 programı yardımıyla hesaplanmıştır.

PISA 2018 matematik uygulaması ilk beş formuna ait betimsel istatistikler aşağıda sunulmuştur. Tablo 6'da her bir ülkeden alt formları alan ve toplamda analize dâhil edilen öğrenci sayıları verilmiştir.

Tablo 6

Alt Formları Alan Öğrenci Sayılarının Ülkelere Göre Dağılımı

Ülkeler	Form-1	Form-2	Form-3	Form-4	Form-5	Toplam
ABD	180	177	184	180	187	908
Çin	469	471	457	462	460	2319
Japonya	225	235	242	236	238	1176
Kore	186	185	180	180	179	910
Rusya	204	205	213	202	204	1028
Türkiye	255	254	280	272	264	1325
Toplam	1519	1527	1556	1532	1532	7666

Tablo 6'da matematik alt formlarından ilk beş formu alan öğrenci sayıları gösterilmektedir. En çok katılım Çin (2319) ve en az katılım ABD'den (908) olmuştur. Çin, PISA araştırmasına Pekin, Şangay, Jiangsu ve Zhejiang olarak dört bölgesiyle katıldığından öğrenci sayısı diğer ülkelerden görece fazladır. Her bir formu cevaplyan öğrenci sayıları 1519 ile 1556 arasında değişmekte olup toplamda 7666 öğrencinin yanıt deseni incelenmiştir.

Tablo 7'de her bir form için ülkelerin toplam puan ortalamaları ve standart sapmaları verilmiştir.

Tablo 7

Formların Betimsel İstatistik Tabloları

Ülkeler	Formlar	Formu Alan Öğrenci Sayısı	Toplam Puan Ortalamaları	Toplam Puan Standart Sapmaları
ABD	Form-1	180	7,77	3,86
	Form-2	177	7,83	3,88
	Form-3	184	8,81	3,81
	Form-4	180	7,49	3,51
	Form-5	187	6,48	3,21
Çin	Form-1	469	12,67	3,61
	Form-2	471	13,26	3,44
	Form-3	457	13,41	3,42
	Form-4	462	11,73	3,19
	Form-5	460	10,81	3,26
Japonya	Form-1	225	10,84	4,10
	Form-2	235	9,90	4,06
	Form-3	242	10,07	4,04
	Form-4	236	9,94	3,49
	Form-5	238	8,35	3,22
Kore	Form-1	186	9,81	4,14
	Form-2	185	10,01	3,47
	Form-3	180	10,97	3,80
	Form-4	180	9,37	3,81
	Form-5	179	8,12	3,62

Rusya	Form-1	204	8,10	3,70
	Form-2	205	8,50	3,73
	Form-3	213	9,14	3,76
	Form-4	202	7,65	3,30
	Form-5	204	7,12	3,44
Türkiye	Form-1	255	7,23	3,73
	Form-2	254	8,55	4,03
	Form-3	280	8,53	4,11
	Form-4	272	6,16	3,25
	Form-5	264	5,37	2,75

Toplam puana ilişkin betimsel istatistik tablosu ülke bazında incelendiğinde tüm formlar için en yüksek ortalamaya sahip ülkenin Çin, en düşük ortalamaya sahip ülkenin ise Türkiye olduğu görülmüştür. Bu durum ülkelerin PISA 2018 matematik testi başarı sıralamalarıyla uyumludur.Tablo formlar bazında incelendiğinde ise tüm ülkeler için form-5 puan ortalamasının diğer formlara göre daha düşük olduğu görülmektedir.

Tablo 8'de alt formların iç tutarlığını analiz etmek için Cronbach alfa değerleri raporlanmıştır.

Tablo 8

Alt formların Cronbach Alfa Güvenirlik Katsayıları

Ülkeler	Form-1	Form-2	Form-3	Form-4	Form-5
ABD	0,79	0,79	0,77	0,79	0,78
Çin	0,75	0,77	0,77	0,76	0,79
Japonya	0,77	0,78	0,79	0,79	0,79
Kore	0,74	0,74	0,75	0,80	0,82
Rusya	0,75	0,74	0,76	0,75	0,81
Türkiye	0,74	0,81	0,83	0,77	0,71

Cronbach Alfa katsayıları indelendiğinde tüm formlar için 0,7'den büyük olduğu görülmektedir bu değer genel kabule uygundur (Kılıç, 2016). Tablo 9'da uygulanan formların ortalama güçlüklere verilmiştir.

Tablo 9

Alt Formların Ortalama Güçlüğü

Ülkeler	Form-1	Form-2	Form-3	Form-4	Form-5
ABD	0,45	0,43	0,49	0,44	0,40
Çin	0,70	0,74	0,75	0,69	0,67
Japonya	0,60	0,55	0,56	0,59	0,52
Kore	0,54	0,56	0,61	0,55	0,50
Rusya	0,45	0,47	0,50	0,45	0,44
Türkiye	0,40	0,47	0,47	0,36	0,33

Tablo 9'dan görüldüğü gibi Form-5'in tüm ülkeler için diğer formlardan daha zor olduğu görülmüştür ve ülkeler için formların ortalama güçlükleri incelendiğinde ise ülkelerin genel başarı sıralamaları ile uyumlu olarak Çin'in en yüksek ve Türkiye'nin en düşük ortalamaya sahip olduğu görülmektedir.

Tablo 10 da Form-1 için maddelerin güçlük indeksleri verilmiştir.

Tablo 10

Form-1 Madde Güçlük İndeksleri

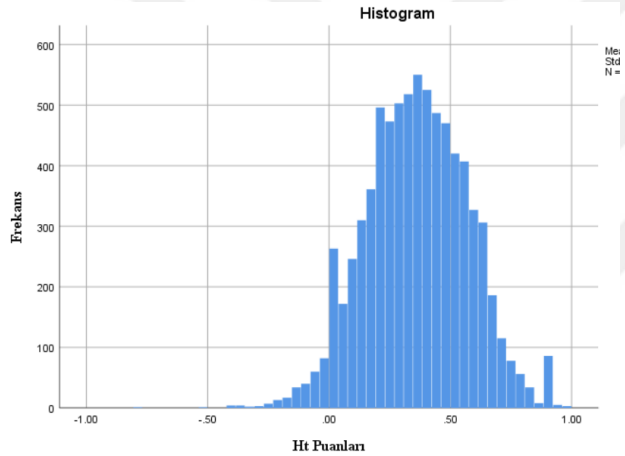
Maddeler	ABD	Çin	Japonya	Kore	Rusya	Türkiye
1	0,75	0,79	0,82	0,74	0,75	0,59
2	0,27	0,51	0,54	0,39	0,37	0,24
3	0,66	0,81	0,64	0,79	0,64	0,61
4	0,46	0,71	0,57	0,52	0,56	0,45
5	0,29	0,75	0,58	0,47	0,34	0,38
6	0,41	0,55	0,5	0,47	0,44	0,38
7	0,44	0,71	0,58	0,57	0,4	0,33
8	0,4	0,75	0,58	0,54	0,36	0,29
9	0,67	0,97	0,84	0,73	0,69	0,81
10	0,28	0,5	0,52	0,42	0,25	0,22
11	0,57	0,68	0,88	0,78	0,63	0,52
12	0,46	0,72	0,56	0,49	0,42	0,33
13	0,61	0,84	0,69	0,69	0,63	0,52
14	0,36	0,77	0,54	0,66	0,48	0,4
15	0,37	0,8	0,55	0,5	0,42	0,44
16	0,35	0,63	0,48	0,46	0,36	0,33
17	0,31	0,59	0,43	0,27	0,21	0,27

Tablo 10 incelendiğinde 5. maddenin Çin'de zor bir madde iken ABD de sınavı alan öğrencilere kolay geldiği; 14. ve 15. ABD'de zor bir maddeyken Çin'deki öğrencilere kolay geldiği söylenebilir. Madde güçlük indeksleri diğer formlar için de hesaplanarak Ek-2 kısmında raporlanmıştır.

Birey uyum analizlerini yapmak üzere her bir form ve her ülkeye ait veri dosyası `'data()'` fonksiyonu yardımıyla R studio programına aktarılmıştır. R studio programında mevcut olan PerFit paketi (Tendeiro, 2015) `'library(PerFit)'` komutu yardımıyla programa yüklenmiştir.

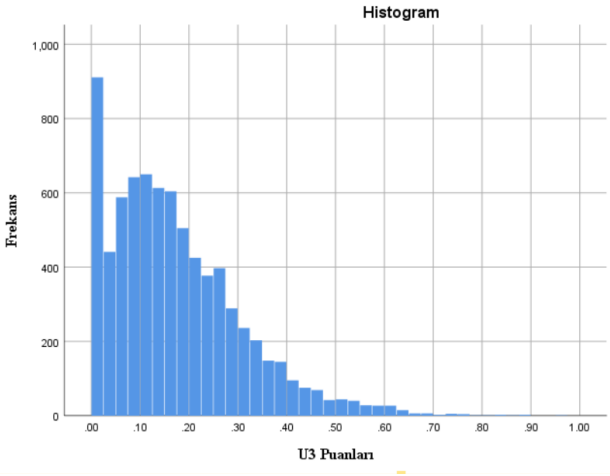
Katılımcıların indeks puanları araştırmaya konu olan ve veri seti ile uyumlu olan H^t, U3 ve C* indeksleri `'Ht.out<-Ht(Data), Cstar.out<-Cstar(Data), U3.out<-U3(Data)'` komutlarıyla hesaplanmıştır.

Birey uyum analizi sonucunda elde edilen H^t, U3 ve C* indeks puanlarına ait histogramlar sırasıyla Şekil 6, 7 ve 8'de sunulmuştur.



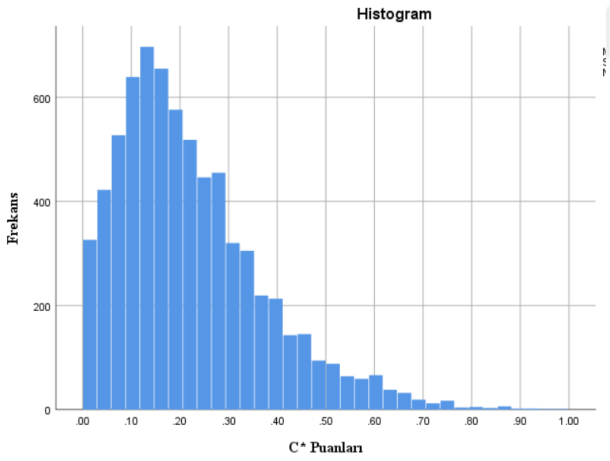
Şekil 6. H^t İndeksinden elde edilen puanlara ait histogram

Histogram incelendiğinde H^t puanları -1 ve 1 arasında değer alabilirken en yüksek frekansı 0,5 ve 0 arasında gösterdiği görülmektedir.



Şekil 7. U3 İndeksinden elde edilen puanlara ait histogram

Histogram incelendiğinde U3 puanları 0 ve 1 arasında değer alabilirken, en yüksek frekansı 0'da gösterdiği görülmektedir.



Şekil 8. C* İndeksinden elde edilen puanlara ait histogram

Histogram incelendiğinde C* puanları 0 ve 1 arasında değer alabilirken, en yüksek frekansı 0,1 ve 0,2 arasında gösterdiği görülmektedir.

Tablo 11'de birey uyum analizi sonucunda elde edilen puanların aritmetik ortalama ve standart sapma değerleri verilmiştir.

Tablo 11

Form-1 İndeks Puanlarının Ortalama ve Standart Sapma Değerleri

Form-1	H^t		U3		C*	
	$\bar{x}$	SS	$\bar{x}$	SS	$\bar{x}$	SS
ABD	0,16	0,13	0,29	0,16	0,29	0,16
Çin	0,15	0,11	0,22	0,13	0,28	0,16
Japonya	0,12	0,12	0,23	0,16	0,25	0,15
Kore	0,18	0,13	0,26	0,15	0,26	0,15
Rusya	0,20	0,13	0,26	0,14	0,26	0,14
Türkiye	0,19	0,13	0,26	0,15	0,27	0,16

H^t indeksi için ortalamanın en yüksek olduğu ülke Rusya olurken en düşük olduğu ülke ise Japonya'dır. U3 indeksine ait ortalamanın en yüksek olduğu ülke ABD olurken en düşük olduğu ülke ise Çin'dir. C* indeksine ait ortalamanın en yüksek olduğu ülke ABD olurken en düşük olduğu ülke ise Japonya'dır. Diğer formlar için hesaplanan ortalama ve standart sapma tabloları Ek-3 kısmında raporlanmıştır.

H^t indeks puanı hangi değere sahip olan katılımcının anormal yanıt deseni gösterdiğine karar vermek amacıyla kesme puanı `'Ht.cut<- cutoff(Ht.out)'` komutuyla belirlenmiştir. Benzer şekilde diğer indeksler için de kesme puanları belirlenmiştir.

Tüm öğrenciler için indeks puanları hesaplandıktan sonra; Rstudio Programı, öğrencinin indeks puanının bulunduğu örnekleme göre normal olup olmadığını kestirmek için 0,05 anlamlılık düzeydinde kesme puanları belirlemiştir. H^t indeksi için belirlenen kesme puanları Tablo 12'de sunulmuştur.

Tablo 12

H^t İndeksi İçin Kesme Puanları

Ülkeler	Form-1	Form-2	Form-3	Form-4	Form-5
ABD	-0,08	-0,04	0,20	0,31	0,28
Çin	-0,02	0,09	0,07	0,26	0,31
Japonya	-0,07	0,08	0,15	0,21	0,35
Kore	-0,02	0,10	0,16	0,21	0,27
Rusya	-0,02	0,07	0,17	0,22	0,32
Türkiye	-0,01	0,09	0,14	0,09	0,24

Tablo 12 incelendiğinde kesme puanlarının her ülke ve form için birbirinden farklı olduğu gözlemlenmektedir. H^t indeksi için belirlenen kesme puandan düşük puana sahip öğrencilerin yanıt desenlerinin anormal olarak tespit edilmiştir. Örneğin ABD'de Form-1'i alan öğrencilerden indeks puanı -0,08'den düşük olan 9 öğrencinin yanıt deseninin anormal olduğu aşağıda raporlanmıştır. U3 indeksi için belirlenen kesme puanları Tablo 13'te sunulmuştur tablo incelendiğinde kesme puanlarının her ülke ve form için birbirinden farklı olduğu gözlemlenmektedir. U3 indeksi için belirlenen kesme puanından yüksek puana sahip öğrencilerin yanıt desenleri anormal olarak tespit edilmiştir.

Tablo 13

U3 İndeksi İçin Kesme Puanları

Ülkeler	Form-1	Form-2	Form-3	Form-4	Form-5
ABD	0,61	0,53	0,34	0,37	0,24
Çin	0,40	0,37	0,43	0,37	0,31
Japonya	0,53	0,39	0,39	0,35	0,26
Kore	0,52	0,39	0,36	0,35	0,30
Rusya	0,51	0,44	0,35	0,27	0,28
Türkiye	0,56	0,40	0,37	0,30	0,21

C* indeksi için belirlenen kesme puanları Tablo 14'te sunulmuştur.

Tablo 14

C İndeksi İçin Kesme Puanları*

Ülkeler	Form-1	Form-2	Form-3	Form-4	Form-5
ABD	0,60	0,55	0,37	0,31	0,31
Çin	0,41	0,49	0,50	0,39	0,33
Japonya	0,55	0,42	0,39	0,38	0,26
Kore	0,48	0,42	0,39	0,34	0,30
Rusya	0,51	0,45	0,40	0,35	0,30
Türkiye	0,57	0,44	0,38	0,40	0,34

Tablo 14 incelendiğinde kesme puanlarının her ülke ve form için birbirinden farklı olduğu gözlemlenmektedir. C* indeksi için belirlenen kesme puandan yüksek puana sahip öğrencilerin yanıt desenlerinin anormal olarak tespit edilmiştir.

Belirlenen kesme puanına dayalı olarak anormal yanıt deseni oluşturan öğrencilerin numaraları `'flagged.resp'` komutuyla belirlenmiştir. Her bir form için ülkelere göre anormal yanıt deseni sergileyen öğrenci sayıları tablolaştırılmıştır.

Tablo 15'te ise kesme puanları yardımıyla tespit edilen, H^t indeksine göre anormal yanıt gösteren öğrencilerin sayısı ve oranı gösterilmiştir.

Tablo 15

H^t indeksine göre Anormal Yanıt Desenine Sahip Öğrenci Sayısı ve Oranı

	H^t	Form-1	Form-2	Form-3	Form-4	Form-5
ABD	N	9	12	18	17	16
	%	5	6,8	9,8	9,4	8,5
Çin	N	37	24	23	42	43
	%	7,8	5,1	5	9,1	9,3
Japonya	N	15	13	16	16	22
	%	5,8	5,5	6,6	6,7	9,2
Kore	N	13	7	12	14	19
	%	6,9	3,9	6,6	7,7	10,6
Rusya	N	10	14	15	20	14
	%	4,9	6,8	7	9,9	6,8
Türkiye	N	17	20	18	30	24
	%	6,7	7,8	6,4	11	9,1

H^t indeksine göre anormal yanıt gösteren öğrencilerin yüzdeleri incelendiğinde en yüksek yüzdenin Kore'ye ait form-5 testinde elde edildiği görülmektedir. En düşük yüzdenin ise yine Kore'de form-2 testinde tespit edildiği görülmüştür.

Tablo 16

*U3 indeksine göre Anormal Yanıt Desenine Sahip Ö*ğrenci *Sayısı ve Oranı*

U3		Form-1	Form-2	Form-3	Form-4	Form-5
ABD	N	11	9	2	5	16
	%	6,1	5,1	1,9	2,7	8,5
Çin	N	26	28	26	30	21
	%	5,5	5,9	5,7	6,5	4,5
Japonya	N	9	17	18	18	12
	%	4	7,2	7,4	7,6	5
Kore	N	11	8	18	12	16
	%	5,9	4,3	10	6,6	8,9
Rusya	N	11	15	15	16	11
	%	5,3	7,3	7	7,9	4,1
Türkiye	N	21	12	16	26	23
	%	8,2	4,7	5,7	9,5	8,7

U3 indeksine göre anormal yanıt gösteren öğrencilerin yüzdeleri incelendiğinde en yüksek yüzdenin Türkiye'ye ait form-4 testinde elde edildiği görülmektedir. En düşük yüzdenin ise yine ABD'de form-3 testinde tespit edildiği görülmüştür.

Tablo 17

C indeksine göre Anormal Yanıt Desenine Sahip Öğrenci Sayısı ve Oranı*

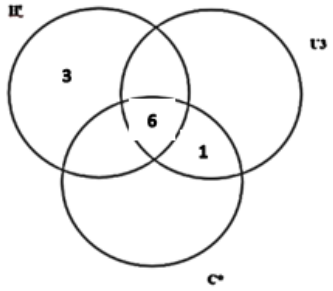
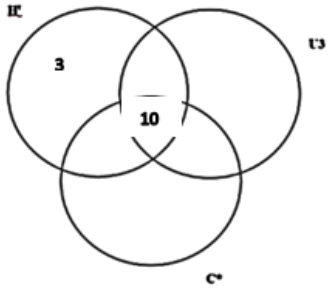
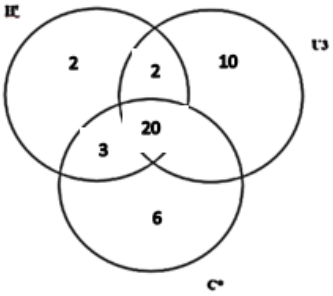
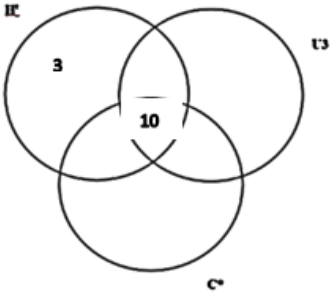
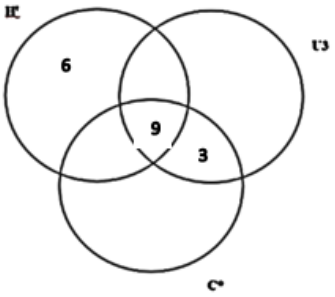
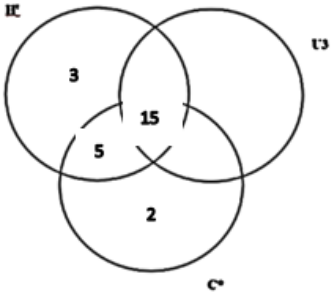
C*		Form-1	Form-2	Form-3	Form-4	Form-5
ABD	N	7	11	11	13	14
	%	3,9	6,2	5,9	7,2	7,5
Çin	N	25	33	21	40	30
	%	5,3	7	4,6	8,6	6,5
Japonya	N	12	13	16	17	14
	%	5,3	5,5	6,6	7,2	5,8
Kore	N	10	8	10	9	13
	%	5,3	4,3	5,5	5	7,2
Rusya	N	10	13	12	20	11
	%	4,9	6,3	5,6	9,9	4,1
Türkiye	N	22	14	16	19	18
	%	8,6	5,5	5,7	7	6,8

C* indeksine göre anormal yanıt gösteren öğrencilerin yüzdeleri incelendiğinde en yüksek yüzdenin Rusya'ya ait form-4 testinde elde edildiği görülmektedir. En düşük yüzdenin ise yine ABD'ye ait form-1 testinde tespit edildiği görülmüştür. Bulunan öğrenci sayılarındaki farklılaşmanın ülkere ve formlara göre anlamlı olup olmadığını tespit etmek için χ^2 ilişki testi yapılmış ve araştırma sorularına ait bulgular kısmında raporlanmıştır.

Tablo 18 Form-1 Anormal yanıt desenine sahip öğrenci sayısının C*, H^t ve U3 indekslerine göre dağılımı verilmiştir. Venn şemasının kesişim bölgesindeki birey sayısı her üç istatistiğe göre de yanıt deseni anormal olarak belirlenen öğrenci sayıları verilmiştir.

Tablo 18

Form-1 Anormal Yanıt Veren Öğrenci Sayısının C, H^t ve U3 İndekslerine Göre Dağılımı*

ABD			Kore
Çin			Rusya
Japonya			Türkiye

Form-1 için venn şeması ile ifade edilen öğrenci sayıları Form-1 için Tablo 19'da özetlenmiş; diğer formlar için ise Ek-4 kısmında tablolaştırılmıştır.

Tablo 19

Form-1 Anormal Yanıt Veren Öğrenci Sayısının H^t, U3 ve C İndekslerine Göre Dağılımı*

Form-1	H^t	U3	C*	$H^t \cap U3 \cap C^*$
ABD	9	7	7	6
Çin	27	32	25	20
Japonya	15	12	12	9
Kore	13	10	10	10
Rusya	10	11	10	10
Türkiye	23	15	22	15

Tablo 19 incelendiğinde ABD'de form-1'de H^t indeksi tarafından 9 öğrenci, U3 indeksi tarafından 7 öğrenci, C* indeksi tarafından 7 öğrenci tespit edilmiştir. Öğrencilerden 6'sı her üç indeks tarafından da tespit edilmiştir. Diğer ülkeler için de durum benzer şekilde olup ortak tespit edilen öğrencilerin sayısı tabloda sunulmuştur. Form-1 de H^t indeksine göre anormal yanıt deseni gösterdiği tespit edilen 9 öğrencinin yanıt desenleri ve H^t indeks puanları Tablo 20'de verilmiştir.

Tablo 20

Form-1 de H^t İndeksine Göre Anormal Yanıt Desenine Sahip Öğrenciler

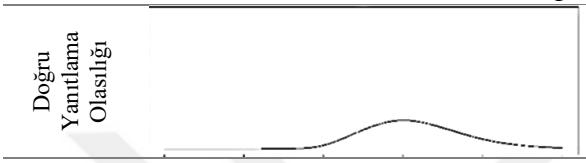
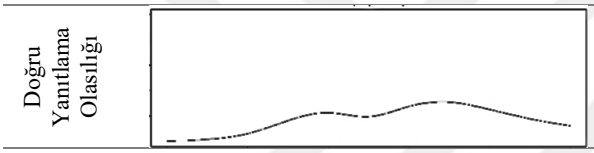
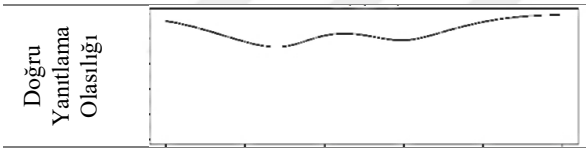
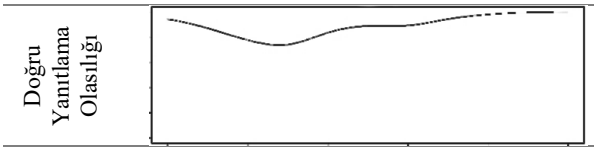
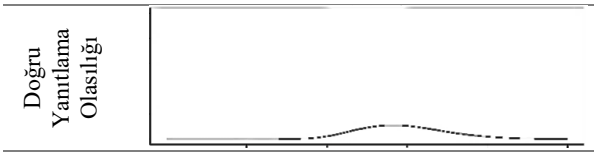
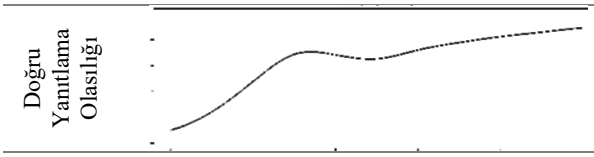
	R Studio Programının Maddeleri Analize Dâhil Ettiği Sıra																	
Öğrenci Sıra No	1	9	3	13	11	4	12	7	6	8	15	14	16	17	5	10	2	H^t
	,75	,67	,66	,61	,57	,46	,46	,44	,41	,40	,37	,36	,35	,31	,29	,28	,27	
14	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	-0,13
21	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0	0	0	-0,08
26	1	0	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	-0,91
29	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	-0,21
63	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	-0,91
73	0	1	1	1	0	1	1	0	0	1	1	1	1	0	1	1	1	-0,11
89	0	0	0	1	1	0	0	0	0	0	1	1	0	0	1	0	1	-0,10
155	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	-0,13
165	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	-0,11

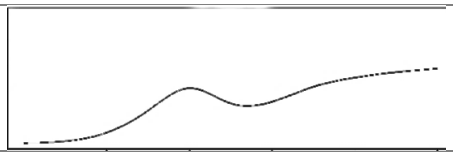
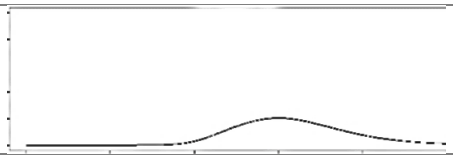
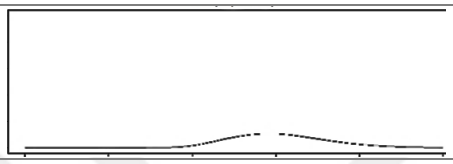
Tablo 20'de ABD'den form-1'i alan ve H^t indeksi tarafından anormal yanıt desenine sahip olduğu tespit edilen 9 öğrencinin yanıt desenleri sunulmuştur. Rstudio programı yardımıyla maddeler kolaydan zora doğru sıralanmıştır. Örneğin 2. sırada 9. madde bulunurken, 2. madde 17. yani son sırada analize dahil edilmiştir. Sıra numarası 14 olan birey daha kolay olan ilk 9 maddeyi doğru cevaplayamazken, 10 ve 11. sıradaki daha zor maddelere doğru cevap vermiş ve testin kalanındaki maddeleri doğru cevaplayamamıştır. H^t indeks puanı -0,13 olan bu bireyin yanıt deseninin anormal olduğu tespit edilmiştir. C* ve U3 indeksine göre tespit edilen öğrencilerin yanıt desenlerine yönelik tablolar Ek-5 kısmında raporlanmıştır. Tabloda anormal yanıt deseni gösteren öğrenciler genellikle kolay olan ilk maddeleri yanlış yanıt verirken maddeler zorlaştıkça doğru cevapları artma eğilimi göstermiştir.

Analizler sonucunda madde yanıt desenlerini görselleştirmek ve tanımlamak amacıyla `'PRFplot (Data, respID=)'` komutu kullanılarak anormal yanıt deseni gösteren bireylerin madde yanıt grafikleri çizdirilmiştir. Rstudio Programında çalışılan kodlar Ek-1 kısmında raporlanmıştır. Sayısal verileri görselleştirmek amacıyla tabloda yanıt deseni verilmiş olan öğrencilerin madde yanıt grafikleri aşağıdaki Tablo 21'de sunulmuştur.

Tablo 21

Form-1 ABD Anormal Yanıt Desenine Sahip Öğrencilerin PRF Grafikleri

	14 numaralı öğrenci kolay maddelere yanlış yanıt verirken maddeler zorlaştıkça doğru yanıtlama olasılığı önce artma sonra azalma eğilimi göstermiştir. H^t puanı=-0,13
	21 numaralı öğrenci kolay maddelere yanlış yanıt verirken maddeler zorlaştıkça doğru yanıtlama olasılığında dalgalanmalar meydana gelmiştir. H^t puanı=-0,08
	26 numaralı öğrenci kolay maddelere doğru yanıt verirken maddeler zorlaştıkça doğru yanıtlama olasılığı önce azalmış daha sonra artma eğilimi göstermiştir. H^t puanı=-0,91
	29 numaralı öğrenci kolay maddelere doğru yanıt verirken maddeler zorlaştıkça doğru yanıtlama olasılığı önce azalmış sonra artış göstermiştir. H^t puanı=-0,21
	63 numaralı öğrenci kolay maddelere yanlış yanıt verirken maddeler zorlaştıkça doğru yanıtlama olasılığı önce artmış sonra azalarak 0'a yaklaşmıştır. H^t puanı=-0,91
	73 numaralı öğrenci kolay maddelere yanlış yanıtlar verirken maddeler zorlaştıkça doğru yanıtlama olasılığı artarak 1'e yaklaşmıştır. H^t puanı=-0,11

Doğru Yanıtlama Olasılığı 		89 numaralı öğrenci kolay maddelere yanlış yanıtlar verirken maddeler zorlaştıkça doğru yanıtlama olasılığı önce artma sonra azalma ardından tekrar artma eğilimi göstermiştir.
	Madde Güçlüğü	H^t puanı=-0,10
Doğru Yanıtlama Olasılığı 		155 numaralı öğrenci 14 numaralı bireye benzer bir yanıt deseni oluşturmuştur.
	Madde Güçlüğü	H^t puanı=-0,13
Doğru Yanıtlama Olasılığı 		165 numaralı öğrenci kolay maddelere yanlış yanıt verirken maddeler zorlaştıkça doğru yanıtlama olasılığı önce küçük bir artış gösterip sonra 0'a yakın şekilde devam etmiştir.
	Madde Güçlüğü	H^t puanı=-0,11

Tablo 21'de yanıt desenleri verilen 9 öğrenciye ait yanıt grafikleri çizdirilmiş ve yorumlanmıştır.

İndeks puanları SPSS 23 programına aktarılarak toplam puanlara ait veri dosyasıyla birleştirilmiştir. İndeks puanlarının normal dağılım gösterip göstermediğini test etmek için Shapiro Wilk testi kullanılmıştır. Shapiro Wilk testine göre tüm formlar ve ülkelerin indeks puanları normal dağılama sahiptir. Hesaplanan birey uyum indeksleri arasındaki ilişkileri incelemek amacıyla Spearman korelasyon katsayısı hesaplanmıştır. Tablo 22'de analiz sürecinde kullanılan H^t, U3 ve C* indeksleri arasındaki korelasyonlar incelenmiştir.

Tablo 22

Hesaplanan İndekslere Ait Korelasyon Katsayıları

	H^t	U3	C*
H^t	1	-,90	-,90
U3	-,90	1	,99
C*	-,90	,99	1

Tablo 22'de hesaplanan birey uyum indeksleri arasındaki ilişkiler incelenmiş ve C* ve H^t indeksi arasında negatif korelasyon gözlenmiştir (Spearman's ρ = -,90, p = 0,00, ρ^2 =0,81). İki değişken arasında yüksek düzeyde bir ilişki vardır ve öğrencilerin H^t puanları yükseldikçe C* puanları düşme eğilimindedir. C* ve U3 indeksi arasında pozitif bir korelasyon gözlenmiştir (Spearman's ρ = ,99, p = 0,00, ρ^2 =0,98). İki değişken arasında yüksek düzeyde bir ilişki vardır ve öğrencilerin U3 puanları yükseldikçe C* puanları da yükselme eğilimindedir. U3 ve H^t indeksi arasında negatif bir korelasyon gözlemlenmiştir (Spearman's ρ = -,90, p = 0,00, ρ^2 =0,81). İki değişken arasında yüksek düzeyde bir ilişki vardır ve öğrencilerin H^t puanları yükseldikçe U3 puanları düşme eğilimindedir.

Ülke ve formların indeks puan ortalaması karşılaştırmaları için iki yönlü ANOVA testi yapılmıştır. Anlamlılık düzeyi $p<0,05$ kabul edilmiştir. Pratik anlamlılık kavramı için sadece p değerine bakmak yerine bağımsız değişkenlerin bağımlı değişkenler üzerinde oluşturduğu etkiyi belirlemek için etki büyüklükleri incelenmiştir (Kirk, 1996). Kısmı η^2 etki büyüklüğü için 0,01 küçük, 0,06 orta ve 0,14 düzeyindeki η^2 ise büyük (large) etki olarak alınmıştır (Cohen, 1992).

ANOVA sonuçlarına göre anlamlı farkların hangi ülke ortalamaları arasında olduğunu belirlemek için varyansların ve örneklemlerin eşit olmadığı durumda yapılması önerilen Games–Howell Post-Hoc testi yapılmıştır (Kayri, 2009). Post-Hoc testi sonuçlarında anlamlılık değerine Bonferonni düzeltmesi uygulanarak p 0,003 olarak hesaplanmıştır. Anormal yanıt davranışı gösterdiği tespit edilen öğrenci sayısı ile ülkeler arasında anlamlı bir ilişkinin var olup olmadığını test etmek için χ^2 ilişki testi yapılmıştır.

Bulgular

Veri analiz süreci sonunda araştırma sorularına yönelik elde edilen bulgular aşağıda tablolaştırılarak sunulmuştur.

PISA 2018 Araştırmasına Amerika Birleşik Devletleri, Çin, Güney Kore, Japonya, Rusya Ve Türkiye'den Katılan Ve İlk Beş Formu Alan Öğrencilerin H^t Birey Uyum İndeks Puan Ortalamaları Farklılık Göstermekte Midir?

Birinci araştırma sorusu için örneklemdeki her öğrencinin H^t indeks puanları hesaplanmıştır. Öncelikle tüm öğrencilerin puanları incelenmiş ardından indeks puanlarına kesme puanı uygulanarak kesme puanından daha düşük indeks puanına sahip öğrencilerin yanıt desenleri anormal olarak tespit edilmiş ve ardından anormal yanıt desenine sahip öğrenci sayısı için analizler yapılmıştır.

Ülkelerin ve Formların H^t İndeks Puan Ortalamaları Üzerindeki Etkileşim Etkisi Nedir? Öğrencilerin H^t İndeks Puan Ortalamaları Ülkelere ve Formlara Göre Farklılık Göstermekte Midir?

Öğrencilerin H^t indeks puanlarının mensup oldukları ülkeye, sınavda aldıkları forma ve her ikisinin etkileşiminin ortak etkisine göre farklılaşmasını incelemek için tüm öğrencilerin indeks puanlarına iki yönlü ANOVA testi yapılmış ve sonuçlar Tablo 23'te raporlanmıştır. İndeks puanlarına ait varyansları karşılaştırmak için yapılan Levene Testi $F(29, 7643)=5,420$ $p=0,00$ anlamlıdır. Varyanslar homojendir.

Tablo 23

H^t İndeks Ortalamalarına Ait İki-Yönlü ANOVA Tablosu

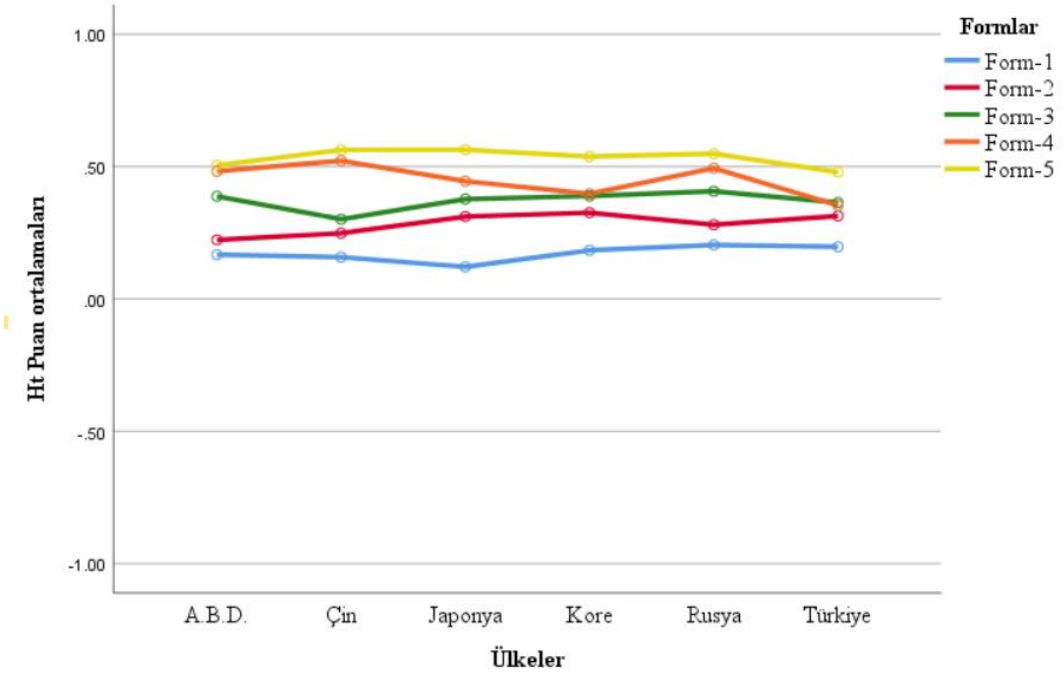
Varyansın kaynağı	Kareler toplamı	sd	Kareler ortalaması	F	p	η^2
Ülke	1,3	5	,26	9,79	,00*	,01
Form	109,02	4	27,26	1030,38	,00*	,35
Ülke*Form	11,6	20	,58	21,93	,00*	,05
Hata	202,17	7643	,03			
Toplam	1341,6	7673				

*p<0,05

Tablo 23'den de görüldüğü gibi H^t puanına göre ülke ve formun ortak etkisi anlamlıdır (F(20, 7643)=21,92, p<,05). Ülke*Form etkileşiminin η^2 değeri 0,05 olarak saptanmıştır ve orta düzeyde öneme sahip olduğunu söylemek mümkündür.

H^t puanına göre ülkeler arasında anlamlı bir farklılaşma bulunmaktadır (F(5, 7643)=9,79, p<,05). Ülkelerin η^2 değeri 0,01 olarak saptanmıştır ve düşük düzeyde öneme sahip olduğunu söylemek mümkündür.

Ülke ayrımı yapılmaksızın formlar arasında da anlamlı bir farklılaşma bulunmaktadır (F(4, 7643)=1030,37, p<,05). Formların η^2 değeri 0,35 olarak saptanmıştır dolayısıyla formların H^t puanlarının farklılaşmasındaki etkisinin yüksek düzeyde öneme sahip olduğunu söylemek mümkündür.



Şekil 9. H^t İndeks Puan Ortalamaları

H^t indeksinden elde edilen puan ortalamalarına ait grafik incelendiğinde Form-5'ten elde edilen puanlara ait ortalamaların tüm ülkeler için diğer formalardan elde edilen ortalamalara göre daha yüksek olduğu, Form-1'den elde edilen puanlara ait ortalamaların ise tüm ülkelerde formlara ait en düşük ortalamalar olduğu gözlenmiştir.

Tablo 24

H^t İndeks Puanları Ülkelere Ait Post-Hoc Sonuçları

H^t	ABD	Çin	Japonya	Kore	Rusya	Türkiye
ABD						
Çin	0,10					
Japonya	0,66	0,66				
Kore	0,79	0,83	1,00			
Rusya	0,00*	0,00*	0,04	0,00*		
Türkiye	0,49	0,10	0,01	0,02	0,00*	

*p<0,003

H^t indeks puanlarında ülkeler arasındaki farklılaşmanın kaynağını bulmak için yapılan post-hoc testi sonucu tablodan incelendiğinde; ABD-Rusya, Çin-Rusya, Kore-Rusya ve Türkiye-Rusya indeks puan ortalamaları arasındaki fark anlamlıdır. Japonya haricindeki tüm ülkelerin indeks puan ortalamaları Rusya'nın ortalamasından farklıdır.

Tablo 25

H^t İndeks Puanları Formlara Ait Post-Hoc Sonuçları

	Form-1	Form-2	Form-3	Form-4	Form-5
Form-1					
Form-2	0,00*				
Form-3	0,00*	0,00*			
Form-4	0,00*	0,00*	0,00*		
Form-5	0,00*	0,00*	0,00*	0,00*	

*p<0,003

H^t indeks puanlarında formlar arasındaki farklılaşmanın kaynağını bulmak için yapılan post-hoc testi sonucu tablodan incelendiğinde; tüm formlar arasındaki fark anlamlıdır.

H^t İndeksi Tarafından Yanıt Deseni Anormal Olarak Belirlenen Öğrenci Sayıları Ülkelere Göre Farklılık Göstermekte Midir?

H^t indeks puanları için bulundukları örnekleme göre belirlenmiş kesme puandan düşük indeks puanına sahip öğrenci sayıları tespit edilmiş ve ülkelere göre farklılaşmasını incelemek için χ^2 ilişki testi yapılmıştır.

Tablo 26

H^t İndeksi ülkeler arasındaki χ^2 İlişki Testi

H^t	ABD	Çin	Japonya	Kore	Rusya	Türkiye	Toplam	χ^2	p
Anormal Yanıt Deseni Tespit Edilen Öğrenci Sayısı	67	158	81	64	75	105	550		
Normal Yanıt Desenine Sahip Öğrenci Sayısı	845	2161	1095	846	956	1220	7123	1,80	,88

H^t indeksine göre anormal yanıt davranışı gösterdiği tespit edilen öğrenci sayısı en fazla olan ülke Çin, en az olan ülke ise Kore'dir ve toplamda 7673 öğrenciden 550 öğrencinin anormal yanıt desenine sahip olduğu tespit edilmiştir. H^t indeksine göre anormal yanıt davranışı gösterdiği tespit edilen öğrenci sayısı ile ülkeler arasında istatistiksel açıdan anlamlı bir ilişki yoktur ($\chi^2_{(5)}$= 1,80, p = 0,88).

Ht İndeksi Tarafından Yanıt Deseni Anormal Olarak Belirlenen Öğrenci Sayıları Formlara Göre Farklılık Göstermekte Midir?

H^t indeks puanları için anormal yanıt desenine sahip öğrenci sayıları tespit edilmiş ve formlara göre farklılaşmasını incelemek için χ^2 ilişki testi yapılmıştır.

Tablo 27

H^t İndeksi formlar arasındaki χ^2 İlişki Testi

H^t	Form-1	Form-2	Form-3	Form-4	Form-5	Toplam	χ^2	p
Anormal Yanıt Deseni Tespit Edilen Öğrenci Sayısı	96	112	109	113	120	550		
Normal Yanıt Desenine Sahip Öğrenci Sayısı	1423	1415	1447	1423	1415	7123	2,82	,59

H^t indeksine göre anormal yanıt davranışı gösterdiği tespit edilen öğrenci sayısı en fazla olan Form-5 en az olan ise Form-1'dir ve toplamda 7673 öğrenciden 550 öğrencinin anormal yanıt desenine sahip olduğu tespit edilmiştir. H^t indeksine göre anormal yanıt davranışı gösterdiği tespit edilen öğrenci sayısı ile formlar arasında istatistiksel açıdan anlamlı bir ilişki yoktur ($\chi^2_{(4)}$= 2,82, p = 0,59).

PISA 2018 Araştırmasına Amerika Birleşik Devletleri, Çin, Güney Kore, Japonya, Rusya ve Türkiye'den Katılan Ve İlk Beş Formu Alan Öğrencilerin U3 Birey Uyum İndeks Puan Ortalamaları Farklılık Göstermekte Midir?

İkinci araştırma sorusu için örneklemdeki her öğrencinin U3 indeks puanları hesaplanmıştır. Öncelikle tüm öğrencilerin puanları incelenmiş ardından indeks puanlarına kesme puanı uygulanarak kesme puanından daha düşük indeks puanına sahip öğrencilerin yanıt desenleri anormal olarak tespit edilmiş ve ardından anormal yanıt desenine sahip öğrenci sayısı için analizler yapılmıştır.

Ülkelerin ve Formların U3 İndeks Puan Ortalamaları Üzerindeki Etkileşim Etkisi Nedir? Öğrencilerin U3 İndeks Puan Ortalamaları Ülkelere ve Formlara Göre Farklılık Göstermekte Midir?

Öğrencilerin U3 indeks puanlarının mensup oldukları ülkeye, sınavda aldıkları forma ve her ikisinin etkileşiminin ortak etkisine göre farklılaşmasını incelemek için tüm öğrencilerin indeks puanlarına iki yönlü ANOVA testi yapılmış ve sonuçlaro Tablo 28'de raporlanmıştır. İndeks puanlarına ait varyansları karşılaştırmak için yapılan Levene Testi F(29, 7643)=10,879 p=0,00 anlamlıdır. Varyanslar homojendir.

Tablo 28

U3 İndeks Ortalamalarına Ait İki-Yönlü ANOVA Tablosu

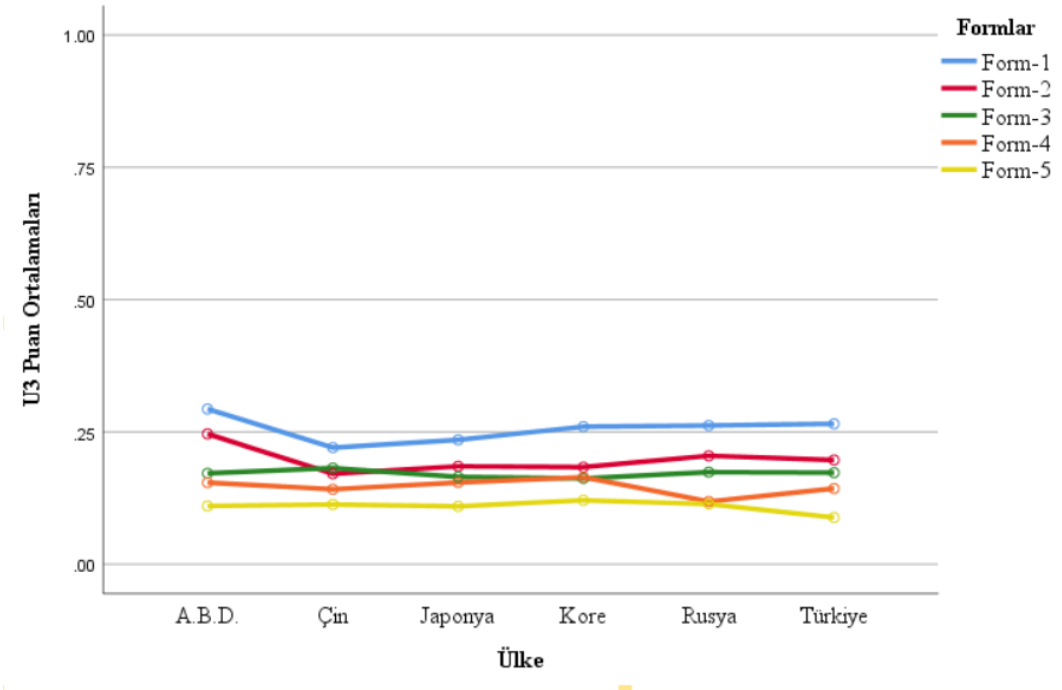
Varyansın kaynağı	Kareler toplamı	sd	Kareler ortalaması	F	p	η^2
Ülke	,61	5	,12	7,76	,00*	,01
Form	16,87	4	4,22	267,16	,00*	,12
Ülke*Form	1,58	20	,08	4,99	,00*	,01
Hata	120,66	7643	,02			
Toplam	370,57	7673				

*p<0,05

Tablo 28 incelendiğinde U3 puanına göre ülke ve formun ortak etkisinin anlamlı olduğu (F(20, 7643)=4,98, p<,05) tablodan anlaşılmaktadır. Ülke*Form etkileşiminin η^2 değeri 0,013 olarak saptanmıştır dolayısıyla Ülke*Form etkileşiminin U3 puanlarının farklılaşmasındaki etkisinin orta düzeyde öneme sahip olduğunu söylemek mümkündür.

U3 puanına göre ülkeler arasında anlamlı bir farklılaşma bulunmaktadır (F(5, 7643)=7,76, p<,05). Ülkelerin η^2 değeri 0,01 olarak saptanmıştır dolayısıyla ülkelerin U3 puanlarının farklılaşmasındaki etkisinin düşük düzeyde öneme sahip olduğunu söylemek mümkündür.

Ülke ayrımı yapılmaksızın formlar arasında da anlamlı bir farklılaşma bulunmaktadır (F(4, 7643)=267,16, p<,05). Formların η^2 değeri 0,12 olarak saptanmıştır dolayısıyla formların ve U3 puanlarının farklılaşmasındaki etkisinin orta düzeyde öneme sahip olduğunu söylemek mümkündür.



Şekil 10. U3 İndeksine Ait Puan Ortalamaları

U3 indeksinden elde edilen puan ortalamalarına ait grafik incelendiğinde Form-5'ten elde edilen puanlara ait ortalamaların tüm ülkeler için diğer formlardan elde edilen ortalamalara göre daha düşük olduğu, Form-1'den elde edilen puanlara ait ortalamaların ise tüm ülkelerde formlara ait en yüksek ortalamalar olduğu gözlemlemlenmiştir. U3 indeksinden elde edilen puan ortlamalarına ait grafikte en yüksek olan Form-1 ortalamasının H^t indeksinden elde edilen puan ortlamalarına ait grafikte en düşük olması indeksler arasında hesaplanan negatif yönlü güçlü korelasyonun varlığını işaret etmektedir.

Tablo 29

U3 İndeks Puanları Ülkelere Ait Post-Hoc Sonuçları

U3	ABD	Çin	Japonya	Kore	Rusya	Türkiye
ABD						
Çin	0,00*					
Japonya	0,00*	0,98				
Kore	0,10	0,08	0,49			
Rusya	0,09	0,41	0,91	0,98		
Türkiye	0,01	0,64	0,99	0,84	0,99	

*p<0,003

Tablo 29 incelendiğinde, U3 indeksine göre ortalamaları arasındaki fark anlamlı olan ülkeler ABD-Çin, ABD-Japonya, ABD-Türkiye'dir.

Tablo 30

U3 İndeks Puanları Formlara Ait Post-Hoc Sonuçları

	Form-1	Form-2	Form-3	Form-4	Form-5
Form-1					
Form-2	0,00*				
Form-3	0,00*	0,00*			
Form-4	0,00*	0,00*	0,00*		
Form-5	0,00*	0,00*	0,00*	0,00*	

*p<0,003

U3 indeks puanlarında formlar arasındaki farklılaşmanın kaynağını bulmak için yapılan post-hoc testi sonucu tablodan incelendiğinde; tüm formlar arasındaki fark anlamlıdır.

U3 İndeksi Tarafından Yanıt Deseni Anormal Olarak Belirlenen Öğrenci Sayıları Ülkelere Göre Farklılık Göstermekte Midir?

U3 indeks puanları için bulundukları örnekleme göre belirlenmiş kesme puandan düşük indeks puanına sahip öğrenci sayıları tespit edilmiş ve ülkelere göre farklılaşmasını incelemek için χ^2 ilişki testi yapılmıştır.

Tablo 31

U3 İndeksi ülkeler arasındaki χ^2 İlişki Testi

U3	ABD	Çin	Japonya	Kore	Rusya	Türkiye	Toplam	χ^2	p
Anormal Yanıt Deseni Tespit Edilen Birey Sayısı	58	150	71	54	65	93	491		
Normal Yanıt Desenine Sahip Birey Sayısı	854	2169	1105	856	966	1232	7182	1,47	,92

U3 indeksine göre anormal yanıt davranışı gösterdiği tespit edilen öğrenci sayısı en fazla olan ülke Çin, en az olan ülke ise Kore'dir ve toplam 7673 öğrenciden 491 öğrencinin anormal yanıt desenine sahip olduğu tespit edilmiştir. U3 indeksine göre anormal yanıt davranışı gösterdiği tespit edilen öğrenci sayısı ile ülkeler arasında istatistiksel açıdan anlamlı bir ilişki yoktur ($\chi^2_{(5)}$= 1,47, p = 0,92).

U3 İndeksi Tarafından Yanıt Deseni Anormal Olarak Belirlenen Öğrenci Sayıları Formlara Göre Farklılık Göstermekte Midir?

U3 indeks puanları için anormal yanıt desenine sahip öğrenci sayıları tespit edilmiş ve formlara göre farklılaşmasını incelemek için χ^2 ilişki testi yapılmıştır.

Tablo 32

U3 İndeksi formlar arasındaki χ^2 İlişki Testi

U3	Form-1	Form-2	Form-3	Form-4	Form-5	Toplam	χ^2	p
Anormal Yanıt Deseni Tespit Edilen Öğrenci Sayısı	88	90	102	109	102	491		
Normal Yanıt Desenine Sahip Öğrenci Sayısı	1519	1527	1556	1536	1535	7673	3,05	,55

U3 indeksine göre anormal yanıt davranışı gösterdiği tespit edilen öğrenci sayısı en fazla olan Form-4 en az olan ise Form-1'dir ve toplam 7673 öğrenciden 491 öğrencinin anormal yanıt desenine sahip olduğu tespit edilmiştir. U3 indeksine göre anormal yanıt davranışı gösterdiği tespit edilen öğrenci sayısı ile formlar arasında istatistiksel açıdan anlamlı bir ilişki yoktur ($\chi^2_{(4)}$= 3,05, p = 0,55).

PISA 2018 araştırmasına Amerika Birleşik Devletleri, Çin, Güney Kore, Japonya, Rusya ve Türkiye'den Katılan Ve İlk Beş Formu Alan Öğrencilerin C* Birey Uyum İndeks Puan Ortalamaları Farklılık Göstermekte Midir?

Birinci araştırma sorusu için örneklemdeki her öğrencinin C* indeks puanları hesaplanmıştır. Öncelikle tüm öğrencilerin puanları incelenmiş ardından indeks puanlarına kesme puanı uygulanarak kesme puanından daha düşük indeks puanına sahip öğrencilerin yanıt desenleri anormal olarak tespit edilmiş ve ardından anormal yanıt desenine sahip öğrenci sayısı için analizler yapılmıştır.

Ülkelerin ve Formların C İndeks Puan Ortalamaları Üzerindeki Etkileşim Etkisi Nedir? Öğrencilerin C* İndeks Puan Ortalamaları Ülkelere ve Formlara Göre Farklılık Göstermekte Midir?*

Öğrencilerin U3 indeks puanlarının mensup oldukları ülkeye, sınavda aldıkları forma ve her ikisinin etkileşiminin ortak etkisine göre farklılaşmasını incelemek için tüm öğrencilerin indeks puanlarına iki yönlü ANOVA testi yapılmış ve sonuçlaro Tablo 33'te raporlanmıştır. İndeks puanlarına ait varyansları karşılaştırmak için yapılan Levene Testi F(29, 7643)=10,536 p=0,00 anlamlıdır. Varyanslar homojendir.

Tablo 33

C İndeks Ortalamalarına Ait İki-Yönlü ANOVA Tablosu*

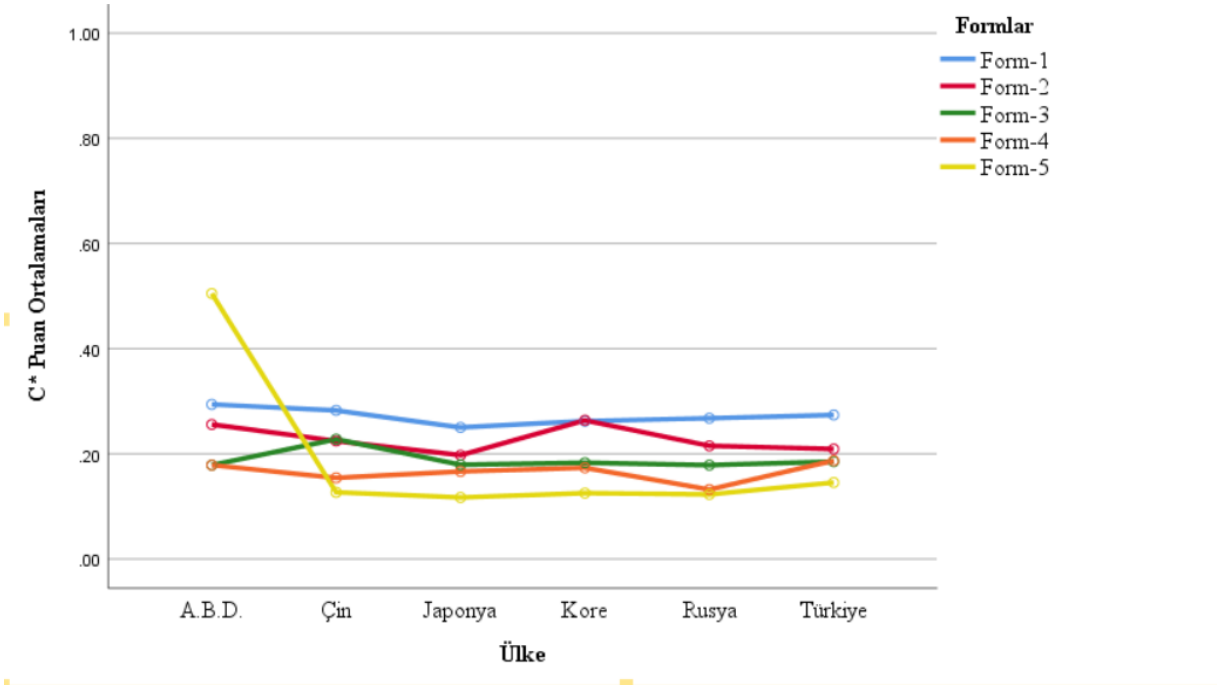
Varyansın kaynağı	Kareler toplamı	sd	Kareler ortalaması	F	p	Kısmi Eta Kare (η^2)
Ülke	6.61	5	1,32	67,1	,00*	,04
Form	9.55	4	2,39	121,3	,00*	,06
Ülke*Form	18.78	20	,94	47,7	,00*	,11
Hata	150.48	7643	,02			
Toplam	513.54	7673				

*p<0,05

Tablo 33 incelendiğinde, C* indeks puanına göre ülke ve formun ortak etkisinin anlamlı olduğu anlaşılmaktadır (F(20, 7643)=47,7, p<,05). Ülke*Form etkileşiminin η^2 değeri 0,11 olarak saptanmıştır dolayısıyla formların ve Ülke*Form etkileşiminin C* puanlarının farklılaşmasındaki etkisinin orta düzeyde öneme sahip olduğunu söylemek mümkündür.

C* indeks puanına göre ülkeler arasında anlamlı bir farklılaşma bulunmaktadır (F(5, 7643)=67,1, p<,05). Ülkelerin η^2 değeri 0,04 olarak saptanmış olup dolayısıyla ülkelerin C* puanlarının farklılaşmasındaki etkisinin düşük düzeyde öneme sahip olduğunu söylemek mümkündür.

Ülke ayrımı yapılmaksızın formlar arasında da anlamlı bir farklılaşma bulunmaktadır (F(4, 7643)=121,3, p<,05). Formların η^2 değeri 0,06, olarak saptanmıştır dolayısıyla formların C* puanlarının farklılaşmasındaki etkisinin orta düzeyde öneme sahip olduğunu söylemek mümkündür.



Şekil 11. C* İndeksine Ait Puan Ortalamaları

C* indeksinden elde edilen puan ortalamalarına ait grafik incelendiğinde Form-5'ten elde edilen ABD'ye ait indeks ortalamasını yüksek iken diğer ülkelerin indeks ortalamaları görece düşüktür. Form-1'den elde edilen indeks ortalamalarının ise Form-5 ABD haricindeki ortalamalardan daha yüksek olduğu gözlemlenmiştir.

İki yönlü ANOVA tablosu incelendiğinde C* indeks puanına göre ülkeler arasında bulunan anlamlı farklılaşmanın kaynağını incelemek amacıyla yapılan post-hoc testi sonuçlarına ait anlamlılık değerleri Tablo 34'te verilmiştir.

Tablo 34

C İndeks Puanları Ülkelere Ait Post-Hoc Sonuçları*

C*	ABD	Çin	Japonya	Kore	Rusya	Türkiye
ABD						
Çin	0,00*					
Japonya	0,00*	0,00*				
Kore	0,00*	1,00	0,01			
Rusya	0,00*	0,02	1,00	0,03		
Türkiye	0,00*	0,96	0,02	0,10	0,06	

*p<0,003

C* indeksine göre ortalamaları arasındaki fark anlamlı olan ülkeler ABD-Çin, ABD-Japonya, ABD-Kore, ABD-Rusya, ABD-Türkiye, Japonya-Çin, Rusya-Çin'dir. ABD ile tüm ülkelerin indeks puan ortalamaları farklıdır.

Tablo 35

C İndeks Puanları Formlara Ait Post-Hoc Sonuçları*

	Form-1	Form-2	Form-3	Form-4	Form-5
Form-1					
Form-2	0,00*				
Form-3	0,00*	0,00*			
Form-4	0,00*	0,00*	0,00*		
Form-5	0,00*	0,00*	0,03	0,42	

*p<0,003

C* indeks puanlarında formlar arasındaki farklılaşmanın kaynağını bulmak için yapılan post-hoc testi sonucu tablodan incelendiğinde; Form-3 ve Form-5 arasında ve Form-4 ile Form-5 ortalamaları arasındaki fark haricindeki tüm formlar arasındaki fark anlamlıdır.

C İndeksi Tarafından Yanıt Deseni Anormal Olarak Belirlenen Öğrenci Sayıları Ülkelere Göre Farklılık Göstermekte Midir?*

C* indeks puanları için bulundukları örnekleme göre belirlenmiş kesme puanından düşük indeks puanına sahip öğrenci sayıları tespit edilmiş ve ülkelere göre farklılaşmasını incelemek için χ^2 ilişki testi yapılmıştır.

Tablo 36

C İndeksi ülkeler arasındaki χ^2 İlişki Testi*

C*	ABD	Çin	Japonya	Kore	Rusya	Türkiye	Toplam	χ^2	P
Anormal Yanıt Deseni Tespit Edilen Öğrenci Sayısı	57	151	70	53	66	89	486		
Normal Yanıt Desenine Sahip Öğrenci Sayısı	855	2168	1106	857	965	1236	7187	1,16	,95

C* indeksine göre anormal yanıt davranışı gösterdiği tespit edilen öğrenci sayısı en fazla olan ülke Çin, en az olan ülke ise Kore'dir ve toplam 7187 öğrenciden 486 öğrencinin anormal yanıt desenine sahip olduğu tespit edilmiştir. C* indeksine göre anormal yanıt davranışı gösterdiği tespit edilen öğrenci sayısı ile ülkeler arasında istatistiksel açıdan anlamlı bir ilişki yoktur ($\chi^2_{(5)}$= 1,16, p = 0,95).

C İndeksi Tarafından Yanıt Deseni Anormal Olarak Belirlenen Öğrenci Sayıları Formlara Göre Farklılık Göstermekte Midir?*

C* indeks puanları için anormal yanıt desenine sahip öğrenci sayıları tespit edilmiş ve formlara göre farklılaşmasını incelemek için χ^2 ilişki testi yapılmıştır.

Tablo 37

C İndeksi Formlar Arasındaki χ^2 İlişki Testi*

C*	Form-1	Form-2	Form-3	Form-4	Form-5	Toplam	χ^2	P
Anormal Yanıt Deseni Tespit Edilen Öğrenci Sayısı	87	95	86	118	100	486		
Normal Yanıt Desenine Sahip Öğrenci Sayısı	1432	1432	1470	1418	1435	7187	7,47	,11

C* indeksine göre anormal yanıt davranışı gösterdiği tespit edilen öğrenci sayısı en fazla olan Form-4 en az olan ise Form-3'tür ve toplam 7187 öğrenciden 486 öğrencinin anormal yanıt desenine sahip olduğu tespit edilmiştir.C* indeksine göre anormal yanıt davranışı gösterdiği tespit edilen öğrenci sayısı ile formlar arasında istatistiksel açıdan anlamlı bir ilişki yoktur ($\chi^2_{(4)}$= 7,47, p = 0,11).

BÖLÜM IV

SONUÇ, TARTIŞMA VE ÖNERİLER

Bu kısımda öncelikle araştırmanın sonucunda elde edilen bulgular özetlenmiş ve literatürdeki benzer çalışmalar eşliğinde bulgular tartışılmıştır. Son olarak bu alanda çalışan araştırmacılara yönelik öneriler sunulmuştur.

Sonuç ve Tartışma

Tablolar ayrı ayrı yorumlandıktan sonra genel yorumlara ulaşmak için bu kısımda bulguların özet sonuçları sunulmuştur.

1. Birey uyum analizi sonucuna göre ülke ve formlara göre anormal yanıt desenine sahip öğrenci oranı %2 ile %11 arasındadır. Emmen (2011), araştırmasında Likert tipi anket verilerine uyguladığı $U3_p$ indeksine göre anormal yanıt desenine sahip birey oranını %5 olarak tespit etmiş ve Aydoğan (2021), araştırmasında TIMMS verilerinden hesapladığı l_z indeksine göre anormal yanıt desenine sahip öğrenci oranını %6 olarak tespit etmiştir. Avcu (2022), çalışmasında Temel Kişilik Envanteri verilerinden hesapladığı l_z indeksine göre anormal yanıt desenine sahip birey oranını %10 olarak tespit etmiştir. Çalışmanın sonucunda ülke ve formlara göre tespit edilen anormal yanıt desenine sahip öğrenci oranı literatürle uyumludur.
2. Yapılan iki yönlü ANOVA testi sonuçlarına göre H^t, U3 ve C* indeks puan ortalamalarının farklılaşmasında ülke form etkileşiminin, öğrencinin ülkesinin ve sınavda aldığı formun anlamlı olduğu sonucuna varılmıştır. Literatür incelendiğinde

Harnisch ve Linn (1981), çalışmasında okuma ve matematik testlerinden elde ettiği veri seti üzerinde C* indeksini hesaplamış ve C* indeks puan ortalamalarının farklı okul ve eğitim sürecine bağlı olarak farklılaşabileceğini ifade etmiştir benzer şekilde bu çalışma içinde farklılaşmanın kaynağı okul ve eğitim süreci olabilir. Bunun yanı sıra ortalamalar arasındaki farklaşmanın Meijer, vd., (2008) ve Meijer & Tendeiro (2014) çalışmalarındaki gibi etnik farklılıklar veya dilsel farklılıklardan; Ferrando (2012), çalışmasındaki gibi farklı ülkelerde testi alma motivasyonunun düşük olmasıdan veya formun zor olmasının bir sonucu olarak Meijer (1996) çalışmasında ifade edildiği gibi öğrencinin yetiştiremesinden kaynaklı olabileceği düşünülmektedir.

3. İki yönlü ANOVA sonuçlarına göre ortalamalar arasındaki anlamlı farklılaşmaların etki büyüklükleri incelendiğinde;

H^t indeksine göre; Ülke*Form etkileşiminin η^2 değeri 0,05 olarak saptanmıştır ve orta düzeyde öneme sahip olduğunu söylemek mümkündür. Ülkelerin η^2 değeri 0,01 olarak saptanmıştır ve düşük düzeyde öneme sahip olduğunu söylemek mümkündür. Formların η^2 değeri 0,35 olarak saptanmıştır dolayısıyla formların H^t puanlarının farklılaşmasındaki etkisinin yüksek düzeyde öneme sahip olduğunu söylemek mümkündür.

U3 indeksine göre; Ülke*Form etkileşiminin η^2 değeri 0,013 olarak saptanmıştır dolayısıyla Ülke*Form etkileşiminin U3 puanlarının farklılaşmasındaki etkisinin orta düzeyde öneme sahip olduğunu söylemek mümkündür. Ülkelerin η^2 değeri 0,01 olarak saptanmıştır dolayısıyla ülkelerin U3 puanlarının farklılaşmasındaki etkisinin düşük düzeyde öneme sahip olduğunu söylemek mümkündür. Formların η^2 değeri 0,12 olarak saptanmıştır dolayısıyla formların ve U3 puanlarının farklılaşmasındaki etkisinin orta düzeyde öneme sahip olduğunu söylemek mümkündür.

C* indeksine göre; Ülke*Form etkileşiminin η^2 değeri 0,11 olarak saptanmıştır dolayısıyla formların ve Ülke*Form etkileşiminin C* puanlarının farklılaşmasındaki etkisinin orta düzeyde öneme sahip olduğunu söylemek mümkündür. Ülkelerin η^2 değeri 0,04 olarak

saptanmış olup dolayısıyla ülkelerin C* puanlarının farklılaşmasındaki etkisinin düşük düzeyde öneme sahip olduğunu söylemek mümkündür. Formların η^2 değeri 0,06, olarak saptanmıştır dolayısıyla formların C* puanlarının farklılaşmasındaki etkisinin orta düzeyde öneme sahip olduğunu söylemek mümkündür.

İndeks ortalamalarında tespit edilen farklılaşmaların etki büyüklüklerinin incelendiğinde; etkileşimin, ülklerin ve formların farklı önem düzeylerinde olduğu görülmektedir. Karabatsos (2003), çalışmasında, H^t indeksinin kopya çekme, yaratıcı yanıt verme ve şanslı tahmin davranışını tespit etmede en iyi performans gösteren indeks olduğunu; U3 ve C* indekslerinin ise rasgele yanıtlama davranışını tespit etmede en iyi indeksler olduğunu ifade etmiştir. Etki büyüklüklerindeki farklılaşmanın sebebinin indekslerin farklı davranış türlerini tespit etme performansı olduğu düşülmektedir.

4. Yapılan ANOVA testi sonucunda indeks puan ortalamalarındaki farklılaşmanın hangi ülkeler arasında olduğunu tespit etmek için post-hoc testi yapılmıştır. H^t indeksine göre ortalamaları arasındaki farkın anlamlı olduğu ülkeler ABD-Rusya, Çin-Rusya, Kore-Rusya ve Türkiye-Rusya'dır. Rusya ile, Japonya haricindeki tüm ülkelerin indeks puan ortalamaları farklıdır. U3 indeksine göre ortalamaları arasındaki fark anlamlı olan ülkeler ABD-Çin, ABD-Japonya, ABD-Türkiye'dir. C* indeksine göre ortalamaları arasındaki fark anlamlı olan ülkeler ABD-Çin, ABD-Japonya, ABD-Kore, ABD-Rusya, ABD-Türkiye, Japonya-Çin, Rusya-Çin'dir. ABD ile tüm ülkelerin indeks puan ortalamaları farklıdır. PISA araştırması kültürlerarası bir çalışma olup, bu çalışmaya katılan farklı etnik gruplardaki ve kültürlerdeki bireylerin motivasyonel inançları düşük veya yüksek olabilir (Pintrich, 2003). Örneğin Heine (2001), çalışmasında Japonların eksik yönlerine odaklanarak daha çok çalıştıklarını, Kuzey Amerikalıların ise güçlü yönlerine odaklanarak daha çok çalıştıklarını bulmuştur. Koreli öğrencilerin ise okumaya karşı güçlü bir ilgilerinin olduğunu tespit etmiştir. Literatürle uyumlu olarak öğrencilerin kültürel farklılıklarının ve çalışma motivasyonlarının ülkeler arasındaki farklılaşmaya sebep olabileceği düşünülmektedir.

PISA matematik testi başarı sıraları incelendiğinde; araştırmaya dört eyaletiyle (Pekin, Şangay, Jiangsu ve Zhejiang) katılan Çin 1. olmuştur. Japonya 6. ve Güney Kore ise 7. sıradadır. Son olarak Rusya 31., ABD 38. ve Türkiye 43. sırada bulunmaktadır. Ülkelerin başarı durumlarını göz önünde bulundurulduğunda ülkenin başarı ortalaması düşükse öğrenciler yanlış maddeler doğruyu götürmediği (düzeltme puanı uygulaması) için bilmedikleri maddelere rastgele yanıtlama eğilimde olabilirler. Buna bağlı olarak da başarısı görece düşük olan ülkelerin öğrencileri ile başarısı görece yüksek olan ülkelerin indeks puan ortalamaları arasındaki fark anlamlı olabilir.

5. Yapılan ANOVA testi sonucunda indeks puan ortalamalarındaki farklılaşmanın hangi formlar arasında olduğunu tespit etmek için post-hoc testi yapılmıştır. H^t ve U3 indeksine göre tüm formların ortalamaları arasındaki farkın anlamlıdır. C* indeksinde ise form-3 ve form-5 arasında ve form-4 ile form-5 ortalamaları arasındaki fark haricindeki tüm formlar arasındaki fark anlamlıdır. Tablo 9'da raporlandığı gibi form-5'in ortalama güçlüğü diğer formlardan daha düşük, Form-3'ün güçlüğünün ise Japonya haricindeki ülkelerde diğer formlardan daha yüksektir. Literatür incelendiğinde, güçlük indeksi düşük maddelerden oluşan testlerde anormal yanıt deseni gösteren bireylerin tespit edilmesinin daha kolay olduğu gözlemlenmiştir (Cole, 1993; Emons, Sijtsma & Meijer, 2004; Meijer, Molenaar & Sijtma, 1994). Formlar arasındaki farklılaşmanın formların ortalama güçlüğünden kaynaklandığı düşünülmektedir.

6. Anormal yanıt gösteren öğrenci sayısı ile ülkeler arasında bir fark olup olmadığını test etmek için yapılan χ^2 ilişki testi sonucunda anlamlı bir fark bulunamamıştır. ANOVA sonuçlarında raporlandığı gibi ülkelerin indeks puanları ortalamaları arasındaki farklar anlamlıyken, kesme puanları yardımıyla anormal yanıt desenine sahip olduğu tespit edilen öğrenci sayıları arasındaki farkın anlamlı çıkmamasının; kesme puanlarının her örneklem için farklı olmasından kaynaklandığı düşünülmektedir.

7. Anormal yanıt gösteren öğrenci sayısı ile sınav esnasında yanıtladıkları form arasında bir fark olup olmadığını test etmek için yapılan χ^2 ilişki testi sonucunda anlamlı bir fark bulunamamıştır. form-1 ve form-4 17 madde, form-2 ve form-3 18 madde, form-5 ise 16 adet madde içermektedir. Literatür incelendiğinde test uzunluğu arttıkça birey uyum indekslerinin tespit oranlarının arttığı gözlemlenmektedir (Emons, Sijtsma & Meijer, 2004; Karabatsos, 2003; Nering & Meijer, 1998). Formların test uzunluklarının birbirine yakın olduğundan tespit edilen öğrenci sayıları arasında bir fark bulunamadığı düşünülmektedir.
8. Literatür incelendiğinde %25 ve üzeri anormal yanıt desenine sahip birey oranının kabul edilebilirlik açısından eşik olduğu belirtilmiştir (Karabatsos, 2003). PISA 2018 Matematik testi ilk beş formdaki anormal yanıt desenine sahip öğrenci oranı %2 ve %11 arasında ve kabul edilebilir eşiğin altında olduğu bulunmuştur.
9. Son olarak gerçek veride çalışmanın bir sonucu olarak katılımcıların yanıtlarını etkileyen tüm süreçler kontrol edilememiştir. Anormal yanıt desenine sahip öğrencilerin anormal yanıt davranışı gösterme sebepleri hakkında genel yorumlar yapmak yanıltıcı olabilir.

Öneriler

Bu kısımda birey uyum indekslerine yönelik çalışmalar yapan araştırmacılara öneriler sunulmuştur.

Bilgisayarla uygulanmış olan PISA 2018 araştırmasında öğrencilerin test maddelerine verdikleri yanıtlarla ilgili ayrıntılı bilgiler elde etmek için test maddesinin ekranda görülme anı, öğrencilerin test maddesini yanıtlama süresi gibi zaman verilerinden yola çıkarak, anormal yanıt davranışının sebebine ait analizler yapılabilir.

PISA 2018 uygulamasında kısmi puanlama yapılan sorular veya Likert tipi ölçeklendirilmiş olan anket maddeleri bulunmaktadır, bu madde tipleri için Rstudio programı PerFit

Kütüphanesi'nde bulunan Izpoly (Drasgow, Levine, & Williams, 1985), Gpoly (Molenaar, 1991) ve U3poly (Emons, 2008) indeksleri kullanılarak birey uyum analizi yapılabilir.

Ülkelerin indeks puan ortalamaları arasındaki farkın kaynağını tespit etmek için yapılan ikili karşılaştırmalarda farklı indeksler için farklı ülkeler ve formlar arasında fark tespit edilmiştir fakat bu farka sebep olan davranışın türü tahmin seviyesinde kalmıştır. Kesin sonuçlar için Meijer ve Sitjtsma (2001) ve Rupp (2013) tarafından birey uyum analizinin nasıl gerçekleştirileceği formüle edilmiştir. Çalışmada, istatistiksel tespit adım, sayısal tablo adımı, grafiksel keşif adımı mevcut PISA verilerinden hesaplanmıştır fakat OECD, PISA araştırmasına katılan öğrencilerin kimlik bilgilerini paylaşmadığından nicel keşif adımı ve nitel açıklama adımı gerçekleştirememiştir. Araştırmaya katılan öğrencilerin kimlik bilgilerinin paylaşıldığı bir uygulamada anormal yanıt deseni gösteren bireyler ile görüşmeler yapılarak sesli düşünme protokolleri uygulanarak nicel keşif adımı ve nitel açıklama adımı gerçekleştirilerek anormal yanıt desenin sebeplerine ulaşılabilir.

İndeks puan ortalamaları arasındaki fark anlamlıyken, her örneklem için belirlenen kesme puanları yardımıyla tespit edilen anormal yanıt desenine sahip öğrenci sayısı arasında anlamlı bir fark bulunamamıştır. Bu durumun kesme puanları arasındaki farktan kaynaklandığı düşülmektedir gelecek çalışmalarda kesme puanlarının belirlenmesi konusunda deneysel verilerde çalışmaların yapılması önerilmektedir.

Bu araştırma kullanılan indeksler ve örnekleme dahil edilen ülkleler ve formlarla sınırlıdır. Aynı çalışma farklı ülkeler ve formlar seçilerek, farklı birey uyum indeksleri kullanılarak tekrarlanabilir.

KAYNAKLAR

AERA, APA & NCME. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Armstrong, R. D., & Shi, M. (2009). A parametric cumulative sum statistic for person fit. *Applied Psychological Measurement, 33*(5), 391-410.

Avcu, A. (2022). The association between person-fit and extreme response style. Pamukkale *University Journal of Education, 55*, 27-45.

Avşar, A. (2019). Comparison of person-fit statistics for polytomous items in different test conditions. *Journal of Measurement and Evaluation in Education and Psychology*, *10*(4), 377-393.

Avşar, A. Ş., & Emons, W. H. (2021). A cross-cultural comparison of non-cognitive outputs towards science between Turkish and Dutch students taking into account detected person misfit. *Studies in Educational Evaluation*, *70*, 101053.

Aydoğan, S. (2021). *Kişi Uyum İndekslerinin Testin Psikometrik Özelliklerine Etkisi*. Doktora Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Birenbaum, M. (1986). Effect of dissimulation motivation and anxiety on response pattern appropriateness measures. *Applied Psychological Measurement*, *10*, 167-174.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment, 6*(4), 284.

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

Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement, 6*, 475–494.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*(3), 98-101.

Cole, J. (1993). Development and validation of a Rasch-derived CES-D short form. *Psychological assessment, 16*(4), 360.

Cui, Y., & Li, J. (2015). Evaluating person fit for cognitive diagnostic assessment. *Applied Psychological Measurement, 39*, 223–238.

de la Torre, J., & Deng, W. (2008). Improving person-fit assessment by correcting the ability estimate and its reference distribution. *Journal of Educational Measurement, 45*, 159–177.

Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology, 38*, 67–86.

Donlon, T. F., & Fischer, F. E. (1968). An index of an individual's agreement with group-determined item difficulties. *Educational and Psychological Measurement, 28*, 105-113.

Emmen, P. (2011). *A person-fit analysis of personality data*. Master Thesis, The Vrije Universiteit Amsterdam, Amsterdam.

Emons, W. H. M., Meijer, R. R., & Sijtsma, K. (2002), Comparing simulated and theoretical sampling distributions of the U3 person-fit statistic. *Appl. Psychol. Meas, 26*, 88–108.

Emons, W. H. M. (2003). *Detection and diagnosis of misfitting item-score vectors.* Doktora Tezi, The Netherlands Tilburg University, Tilburg.

Emons, W. H. M. (2008). Nonparametric person-fit analysis of polytomous item scores. *Applied Psychological Measurement, 32*(3), 224–247.

Ercan, İ., & Kan, İ. (2004). Ölçeklerde Güvenirlilik ve Geçerlilik. *Uludağ Üniversitesi Tıp Fakültesi Dergisi, 30*(3), 211-216.

Ergin, D . (2013). 1. Ölçeklerde Geçerlik ve Güvenirlik. *Marmara University Atatürk Education Faculty Journal of Educational Sciences, 7*(7) , 125-148.

Felt, J. M., Castaneda, R., Tiemensma, J., & Depaoli, S. (2017). Using person fit statistics to detect outliers in survey research. *Frontiers in psychology, 8,* 863.

Ferrando, P. J. (2012). Assessing inconsistent responding in E and N measures: An application of person-fit analysis in personality. P*ersonality and Individual Differences, 52,* 718- 722.

Ferrara, S., & DeMauro, G. E. (2006). Standardized assessment of individual achievement in K-12. *Educational measurement, 5,* 579-621.

Fowler, H. M. (1954). An application of the Ferguson method of computing item conformity and person conformity. *Journal of Experimental Education, 22,* 237–245.

Glaser, R. (1949). A methodological analysis of the inconsistency of responses to test items. *Educational and Psychological Measurement, 9,* 721–739.

Glaser, R. (1950). Multiple operation measurement. *Psychological Review, 57,* 241–253.

Glaser, R. (1951). The applications of the concepts of multiple operation measurement to the response patterns on psychological tests. *Educational and Psychological Measurement, 11,* 322–382.

Glaser, R. (1952). The reliability of inconsistency. *Educational and Psychological Measurement, 12,* 60–64

Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review, 9,* 139–150.

Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Claussen (Eds.), *Measurement and prediction* (pp. 60–90). Princeton, NJ: Princeton University.

Hacettepe Üniversitesi (2020). *PISA ve Türkiye 2000-2018*. http://www.egitim.hacettepe.edu.tr/belge/pisaveturkiye.pdf

Harnisch, D. L. (1983). Item response patterns: Applications for educational practice. *Journal of Educational Measurement*, *20*, 191-206.

Harnisch, D. L., & Linn, R. L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. J*ournal of Educational Measurement, 18*, 133-146.

Harter, S. (1985). *Manual for the self-perception profile for children*. Denver, CO: University of Denver.

Heine, S. J., Kitayama, S., Lehman, D. R., Takata, T., Die, E., Leung, C. (2001). Divergent consequences of success and failure in Japan and North America: An investigation of self-improving motivations and malleable selves. *Journal of Personality and Social Psychology, 81*, 599–615.

Karabatsos, G. (2003). Comparing the Aberrant Response Detection Performance of Thirty-Six Person-Fit Statistics. *Applied Measurement In Education*, *16*(4), 277–298.

Kane, M. T., & Brennan, R. L. (1980). Agreement coefficients as indices of dependability for domain-referenced tests. *Applied Psychological Measurement, 4*, 105-126.

Kayri, M. (2009). Araştirmalarda gruplar arasi farkin belirlenmesine yönelik çoklu karşilaştirma (post-hoc) teknikleri. *Journal of Social Science, 55*, 22.

Kirk, D. (1996). Demographic transition theory. *Population studies, 50*(3), 361-387.

Kiliç, S. (2016). Cronbach's alpha reliability coefficient. *Psychiatry and Behavioral Sciences, 6*(1), 47.

Kogut, J. (1987). *Detecting aberrant item response patterns in the Rasch model* (Research Report 87- 3). Enschede: University of Twente, Department of Education.

Levine, M. V., & Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational Statistics, 4*, 269–290.

MEB. (2018). *PISA Nedir*? http://pisa.meb.gov.tr/?page_id=18 adresinden alınmıştır.

MEB (2019). *PISA 2018 ulusal ön raporu.* http://pisa.meb.gov.tr/wp-content/uploads/2020/01/PISA_2018_Turkiye_On_Raporu.pdf.

Meijer, R. R. (1994). The number of Guttman errors as a simple and powerful person-fit statistic. *Applied Psychological Measurement, 18*(4), 311–314.

Meijer, R. R. (1996). Person-fit research: An introduction. *Applied Measurement in Education, 9*(1), 3-8.

Meijer, R. R. (2003). Diagnosing item score patterns on a test using item response theorybased person-fit statistics. *Psychological Methods, 8*(1), 72–87.

Meijer, R. R., Egberink, I. J., Emons, W. H., & Sijtsma, K. (2008). Detection and validation of unscalable item score patterns using item response theory: An illustration with Harter's Self-Perception Profile for Children. J*ournal of Personality Assessment, 90*(3), 227–238.

Meijer, R. R., Muijtjens, A. M. M., & van der Vlueten, C. P. M. (1996). Nonparametric person-fit research: Some theoretical issues an empirical example. *Applied Measurement in Education, 9*(1), 77-89.

Meijer, R. R., Molanaar, I.W., & Sijtsma, K. (1994). Influence of test and person characteristics on nonparametric appropriateness measurement. *Applied Psychological Measurement, 18,* 111-120.

Meijer, R. R., Niessen, A. S. M., & Tendeiro, J. N. (2016). A practical guide to check the consistency of item response patterns in clinical research through person-fit statistics: Examples and a computer program. *Assessment, 23*(1), 52-62.

Meijer, R. R. & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement, 25*(2), 107–135.

Meijer, R. R., & Tendeiro, J. N. (2014). *The use of person-fit scores in high stakes educational testing: How to use them and what they tell us*. LSAC Research Report 14-03.

Mosier, C. I. (1940). Psychophysics and mental test theory: Fundamental postulates and elementary theorems. *Psychological Review, 47*, 355–366.

Mousavi, A., Tendeiro, J. N., & Younesi, J. (2016). Person fit assessment using the Perfit package in R. *The Quantitative Methods for Psychology, 12*(3), 232–242.

Mousavi, A., Cui, Y., & Rogers, T. (2019). An examination of different methods of setting cutoff values in person fit research. *International Journal of Testing, 19*(1), 1-22.

Nering, M. L. (1995). The distribution of person fit using true and estimated person parameters. *Applied Psychological Measurement, 19*, 121–129.

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2014). *Detection of unmotivated students in a realistic test setting*. Manuscript in preparation.

OECD, (2003). *Literacy Skill for the World of Tomorrow: Further Results from PISA 2000.* Paris: OECD.

OECD, (2009). *PISA 2009 Assessment Framework: Key competencies in reading, mathematics and science.* Paris: OECD.

OECD, (2014). *PISA 2012 Results: What Students Know and Can Do: Students Performance in Mathematics, Reading and Science (Volume I).* Paris: OECD.

OECD, (2018). *What is PISA? Organisation for Economic Co-operation and Development Home*. Paris: OECD.

OECD, (2019). *PISA 2018 Assessment and Analytical Framework.* Paris: OECD.

OECD, (2020). *PISA Participants*. Web Adresi: https://www.oecd.org/pisa/aboutpisa/pisa-participants.htm.

Olson, J., & Fremer, J. (2013). *TILSA test security guidebook: Preventing, detecting, and investigating test security irregularities*. Washington, DC: Council of Chief State School Officers.

Reise, S. P., & Flannery, W. P. (1996). Assessing person-fit on measures of typical performance. *Applied Measurement in Education, 9*, 9-26.

Reise, S. P., & Waller, N. G. (1993). Traidness and the assessment of response pattern scalability. *Journal of Personality and Social Psychology, 65,* 143- 151.

Reise, S. P., & Widaman, K. F. (1999). Assessing the fit of measurement models at the individual level: A comparison of item response theory and covariance structure approaches. *Psychological Methods, 4*, 3-21.

Rudner, L. M. (1983). Individual assessment accuracy. *Journal of Educational Measurement, 20,* 207–219.

Rudner, L. M., & Schafer, W. D. (2001). *Reliability*. ERIC Digest. ED 458213.

Rupp, A. A. (2013). A systematic review of the methodology for person fit research in item response theory: lessons about generalizability of inferences from the design of simulation studies. *Psychological Test and Assessment Modeling, 55,* 3–38.

Sato, T. (1975). *The construction and interpretation of s-p tables*. Tokyo: Mei.

Sherif, M., & Cantril, H. (1945). The psychology of attitudes. *Psychological Review, 52,* 259–319.

Sherif, M., & Cantril, H. (1946). The psychology of attitudes. *Psychological Review, 53,* 1–24.

Sijtsma, K. (1986). A coefficient of deviance of response patterns. *Kwantitative Methoden, 7,* 131-145.

Sijtsma, K., & Meijer, R. R. (1992). A method for investigating the intersection of item response functions in Mokken's non-parametric IRT model. *Applied Psychological Measurement, 16,* 149–157.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3,* 271–295

St-Onge C., Valois P., Abdous B., & Germain S. (2011). Accuracy of person-fit statistics: a monte carlo study of the ınfluence of aberrance rates. *Applied Psychological Measurement, 35*(6), 419–432.

Tatsuoka, K. K., & Tatsuoka, M. M. (1983). Spotting erroneous rules of operation by the individual consistency index. *Journal of Educational Measurement, 20,* 221–230.

Tatsuoka, K. K., & Tatsuoka, M. M. (1982). Detection of aberrant response patterns and their effect on dimensionality. *Journal of Educational Statistics, 7,* 215- 231.

Tellegen, A. (1988). The consistency in personality assessment. *Journal of Personality, 56,* 621–663.

Tendeiro, J. N., & Meijer, R. R. (2012). A CUSUM to detect person misfit: A discussion and some alternatives for existing procedures. *Applied Psychological Measurement, 36,* 420- 442.

Tendeiro, J. (2015). *Package "PersonFit".*,v. 1.3.1.http://cran.rproject.org/web/Packeges/PerFit/PerFit.pdf

Tendeiro, J. N., & Meijer, R. R. (2014). Detection of invalid test scores: The usefulness of simple nonparametric statistics. *Journal of Educational Measurement, 51*(3), 239-259.

Tendeiro, J. N., Meijer, R. R., & Niessen, A. S. M. (2016). PerFit: An R package for person-fit analysis in IRT. *Journal of Statistical Software, 74*(5), 1-27.

Thiessen, B. (2008). *Relationship between test security policies and test score manipulations*. (Unpublished Doctorial Thesis), University of Iowa, Iowa. Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3347249).

Thissen, D., & Wainer, H. (Eds.) (2001). *Test scoring*. Mahwah, NJ: Erlbaum.

Trabin, T. E., & Weiss, D. J. (1983). The person response curve: Fit of individuals to item response theory models. In D. J. Weiss (Ed.), *New horizons in testing*. New York: Academic.

Uçar, E., & Karsak, E. E. (2021). Educational Performance Assessment of OECD Countries Using PISA 2018 Data. *Proceedings of IAC 2021 in Vienna, 1*, 64.

Uçar, A. (2021). *Kopya belirlemede benzerlik indekslerinin birey-uyum istatistikleri aracılığıyla aşamalı kullanımının I. tip hatalarının ve gücünün incelenmesi*. Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Walker, A. A., Jennings, J. K., & Engelhard Jr, G. (2018). Using person response functions to investigate areas of person misfit related to item characteristics. *Educational Assessment*, *23*(1), 47-68.

Wright, B. D., & Stone, M. H. (1979). *Best test design. Rasch measurement.* Chicago: Mega.

Yüksel, M. (2022). PISA 2018 Araştırma Sonuçlarına Göre Ülkelerin Bileşik PISA Performans Sıralaması . *Muğla Sıtkı Koçman Üniversitesi Eğitim Fakültesi Dergisi, 9*(2) , 788-821.

Yormaz, S. (2019). *Test güvenliği açısından bireyler arasındaki olası iş birliğinin incelenmesi.* Doktora Tezi, Mersin Üniversitesi Eğitim Bilimleri Enstitüsü, Mersin.

EKLER

Ek-1.

```
data(FORM1ABD)

library(PerFit)

Cstar(FORM1ABD, NA.method = "Pairwise", Save.MatImp = FALSE,
IP = NULL, IRT.PModel = "2PL", Ability = NULL, Ability.PModel
= "ML", mu = 0, sigma = 1)

Cstar.out<-Cstar(FORM1ABD)

cutoff(Cstar.out, ModelFit = "NonParametric", Nreps=1000,
     IP=x$IP, IRT.PModel=x$IRT.PModel, Ability=x$Ability,
     Ability.PModel=x$Ability.PModel, mu=0, sigma=1, Blvl =
     0.05, Breps = 1000, CIlvl = 0.95, UDlvl = NA)

Cstar.cut<- cutoff(Cstar.out)

flagged.resp(Cstar.out, cutoff.obj=NULL, scores=TRUE,
ord=TRUE,ModelFit="NonParametric", Nreps=1000,IP=x$IP,
IRT.PModel=x$IRT.PModel, Ability=x$Ability,
Ability.PModel=x$Ability.PModel, mu=0, sigma=1,Blvl = 0.05,
Breps = 1000, CIlvl = 0.95,UDlvl=NA)

PRFplot (FORM1ABD, respID=25)
```

```
Ht(FORM1ABD, NA.method = "Pairwise", Save.MatImp = FALSE, IP
= NULL, IRT.PModel = "2PL", Ability = NULL, Ability.PModel =
"ML", mu = 0, sigma = 1)

Ht.out<-Ht(FORM1ABD)

cutoff(Ht.out,  ModelFit  =  "NonParametric",  Nreps=1000,
     IP=x$IP,                          IRT.PModel=x$IRT.PModel,
     Ability=x$Ability,Ability.PModel=x$Ability.PModel, mu=0,
     sigma=1, Blvl = 0.05, Breps = 1000, CIlvl = 0.95,UDlvl =
     NA)

Ht.cut<- cutoff(Ht.out)

flagged.resp(Ht.out,      cutoff.obj=NULL,      scores=TRUE,
ord=TRUE,ModelFit="NonParametric",        Nreps=1000,IP=x$IP,
IRT.PModel=x$IRT.PModel,                   Ability=x$Ability,
Ability.PModel=x$Ability.PModel,  mu=0,  sigma=1,Blvl  =  0.05,
Breps = 1000, CIlvl = 0.95,UDlvl=NA)

PRFplot (FORM1ABD, respID=)

U3(FORM1ABD, NA.method = "Pairwise", Save.MatImp = FALSE, IP
= NULL, IRT.PModel = "2PL", Ability = NULL, Ability.PModel =
"ML", mu = 0, sigma = 1)

U3.out<-U3(FORM1ABD)
```

```
cutoff(U3.out, ModelFit = "NonParametric", Nreps=1000,
     IP=x$IP, IRT.PModel=x$IRT.PModel, Ability=x$Ability,
     Ability.PModel=x$Ability.PModel, mu=0, sigma=1, Blvl =
     0.05, Breps = 1000, CIlvl = 0.95,UDlvl = NA)


U3.cut<- cutoff(U3.out)


flagged.resp(U3.out, cutoff.obj=NULL, scores=TRUE,
ord=TRUE,ModelFit="NonParametric", Nreps=1000,IP=x$IP,
IRT.PModel=x$IRT.PModel, Ability=x$Ability,
Ability.PModel=x$Ability.PModel, mu=0, sigma=1,Blvl = 0.05,
Breps = 1000, CIlvl = 0.95,UDlvl=NA)


PRFplot (Data, respID=)
```

Ek-2.

Form-2 Madde Güçlük İndeksleri						
Maddeler	ABD	Çin	Japonya	Kore	Rusya	Türkiye
1	0,33	0,77	0,55	0,53	0,35	0,43
2	0,33	0,69	0,56	0,48	0,39	0,47
3	0,35	0,54	0,44	0,4	0,47	0,32
4	0,25	0,66	0,41	0,46	0,17	0,53
5	0,51	0,61	0,46	0,26	0,45	0,38
6	0,68	0,97	0,77	0,87	0,72	0,85
7	0,67	0,89	0,77	0,76	0,64	0,73
8	0,55	0,89	0,8	0,86	0,63	0,62
9	0,12	0,64	0,35	0,37	0,17	0,2
10	0,4	0,68	0,44	0,53	0,47	0,33
11	0,58	0,81	0,73	0,71	0,65	0,56
12	0,48	0,89	0,66	0,68	0,56	0,55
13	0,39	0,8	0,45	0,62	0,51	0,39
14	0,43	0,81	0,58	0,5	0,46	0,4
15	0,40	0,67	0,51	0,57	0,46	0,28
16	0,33	0,6	0,38	0,31	0,23	0,33
17	0,82	0,95	0,93	0,96	0,92	0,93
18	0,21	0,37	0,12	0,16	0,26	0,25

Form-3 Madde Güçlük İndeksleri						
Maddeler	ABD	Çin	Japonya	Kore	Rusya	Türkiye
1	0,32	0,72	0,51	0,48	0,43	0,38
2	0,3	0,53	0,35	0,34	0,14	0,36
3	0,55	0,61	0,38	0,29	0,53	0,45
4	0,71	0,92	0,77	0,82	0,73	0,71
5	0,66	0,86	0,78	0,81	0,53	0,53
6	0,10	0,68	0,36	0,48	0,2	0,23
7	0,55	0,93	0,62	0,8	0,65	0,6
8	0,82	0,97	0,95	0,97	0,92	0,95
9	0,22	0,36	0,1	0,15	0,39	0,28
10	0,5	0,84	0,57	0,56	0,6	0,39
11	0,39	0,8	0,6	0,68	0,28	0,36
12	0,62	0,88	0,42	0,77	0,72	0,6
13	0,85	0,81	0,79	0,8	0,85	0,8
14	0,38	0,45	0,39	0,57	0,23	0,22
15	0,4	0,67	0,51	0,57	0,46	0,28
16	0,33	0,6	0,38	0,31	0,23	0,33
17	0,82	0,95	0,93	0,96	0,92	0,93
18	0,21	0,37	0,12	0,16	0,26	0,25

Form-4 Madde Güçlük İndeksleri						
Maddeler	ABD	Çin	Japonya	Kore	Rusya	Türkiye
1	0,06	0,29	0,23	0,14	0,04	0,04
2	0,54	0,86	0,65	0,55	0,64	0,49
3	0,87	0,95	0,93	0,84	0,89	0,62
4	0,55	0,76	0,63	0,57	0,61	0,38
5	0,25	0,69	0,51	0,41	0,29	0,16
6	0,41	0,81	0,66	0,67	0,25	0,33
7	0,64	0,83	0,45	0,77	0,69	0,57
8	0,62	0,88	0,85	0,78	0,65	0,48
9	0,27	0,88	0,7	0,68	0,44	0,22
10	0	0,08	0,05	0,06	0,03	0,01
11	0,83	0,84	0,82	0,8	0,82	0,74
12	0,36	0,37	0,42	0,49	0,18	0,25
13	0,53	0,77	0,74	0,64	0,64	0,65
14	0,38	0,73	0,74	0,49	0,35	0,31
15	0,79	0,9	0,82	0,75	0,67	0,51
16	0,09	0,52	0,24	0,26	0,11	0,09
17	0,31	0,58	0,48	0,49	0,34	0,32

Form-5 Madde Güçlük İndeksleri						
Maddeler	ABD	Çin	Japonya	Kore	Rusya	Türkiye
1	0,03	0,3	0,16	0,12	0,09	0,02
2	0,71	0,89	0,83	0,76	0,74	0,63
3	0,88	0,94	0,94	0,85	0,88	0,66
4	0,48	0,68	0,62	0,57	0,54	0,33
5	0,24	0,64	0,48	0,41	0,26	0,13
6	0,77	0,94	0,82	0,83	0,81	0,83
7	0,41	0,54	0,31	0,35	0,45	0,37
8	0,45	0,81	0,57	0,63	0,53	0,42
9	0,03	0,5	0,08	0,16	0,1	0
10	0,55	0,87	0,82	0,78	0,61	0,53
11	0,32	0,86	0,69	0,67	0,41	0,23
12	0,40	0,81	0,52	0,56	0,42	0,3
13	0,66	0,88	0,76	0,75	0,75	0,54
14	0,21	0,52	0,22	0,24	0,19	0,06
15	0,04	0,07	0,05	0,05	0,04	0,01
16	0,29	0,56	0,48	0,42	0,31	0,32

Ek-3.

Form-2	C* $\bar{x}$	C* SS	H^t $\bar{x}$	H^t SS	U3 $\bar{x}$	U3 SS
ABD	0,25	0,15	0,22	0,14	0,24	0,14
Çin	0,22	0,16	0,24	0,15	0,17	0,12
Japonya	0,19	0,12	0,31	0,14	0,18	0,12
Kore	0,26	0,15	0,32	0,15	0,18	0,12
Rusya	0,21	0,13	0,28	0,14	0,20	0,13
Türkiye	0,20	0,13	0,31	0,15	0,19	0,14

Form-3	C* $\bar{x}$	C* SS	H^t $\bar{x}$	H^t SS	U3 $\bar{x}$	U3 SS
ABD	0,17	0,11	0,38	0,15	0,17	0,12
Çin	0,22	0,17	0,30	0,17	0,18	0,13
Japonya	0,17	0,13	0,37	0,17	0,16	0,12
Kore	0,18	0,12	0,38	0,15	0,16	0,11
Rusya	0,17	0,12	0,40	0,16	0,17	0,13
Türkiye	0,18	0,12	0,36	0,15	0,17	0,11

Form-4	C* $\bar{x}$	C* SS	H^t $\bar{x}$	H^t SS	U3 $\bar{x}$	U3 SS
ABD	0,13	0,11	0,51	0,18	0,16	0,13
Çin	0,15	0,13	0,52	0,19	0,14	0,12
Japonya	0,16	0,12	0,44	0,16	0,15	0,11
Kore	0,17	0,11	0,39	0,17	0,16	0,11
Rusya	0,13	0,11	0,49	0,19	0,11	0,10
Türkiye	0,18	0,14	0,35	0,18	0,14	0,10

Form-5	C* $\bar{x}$	C* SS	H^t $\bar{x}$	H^t SS	U3 $\bar{x}$	U3 SS
ABD	0,14	0,12	0,50	0,18	0,11	0,09
Çin	0,12	0,13	0,56	0,18	0,11	0,12
Japonya	0,11	0,10	0,56	0,17	0,10	0,09
Kore	0,12	0,10	0,53	0,18	0,12	0,10
Rusya	0,12	0,09	0,54	0,16	0,11	0,09
Türkiye	0,14	0,12	0,47	0,17	0,08	0,07

Ek-4.

Form-2	C*	H^t	U3	C*∩ H^t ∩ U3
ABD	11	10	10	8
Çin	33	35	28	18
Japonya	13	17	16	10
Kore	9	13	9	9
Rusya	13	18	15	11
Türkiye	14	18	15	8

Form-3	C*	H^t	U3	C*∩ H^t ∩ U3
ABD	11	15	14	10
Çin	21	32	26	17
Japonya	16	16	15	14
Kore	10	13	13	8
Rusya	12	17	14	12
Türkiye	16	16	19	14

Form-4	C*	H^t	U3	C*∩ H^t ∩ U3
ABD	13	18	12	8
Çin	40	32	28	22
Japonya	17	16	16	11
Kore	9	9	11	6
Rusya	17	17	16	12
Türkiye	19	23	26	19

Form-5	C*	H^t	U3	C*∩ H^t ∩ U3
ABD	14	16	16	13
Çin	29	38	24	20
Japonya	14	17	11	11
Kore	13	16	8	6
Rusya	11	14	11	11
Türkiye	18	23	19	12

Ek-5.

Form-2 Japonya Anormal yanıt desenine sahip bireyler																		
R Studio Programının Maddeleri Analize Dâhil Ettiği Sıra																		
Öğrenci Sıra No	17 ,92	8 ,79	6 ,77	7 ,73	11 ,66	12 ,58	14 ,55	2 ,54	1 ,51	15 ,45	5 ,45	13 ,44	10 ,44	3 ,41	4 ,37	16 ,35	9 ,11	C* istatistik puanı
1	0	0	0	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0,55
43	1	0	0	0	1	1	1	0	1	0	1	1	1	0	0	0	1	0,42
55	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1	0,42
70	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0,57
80	0	1	1	1	1	1	0	1	0	1	0	1	1	1	1	0	1	0,52
102	1	1	0	1	1	0	1	1	1	1	0	1	1	0	1	0	1	0,43
125	1	1	0	1	1	1	1	0	1	1	0	1	1	1	0	1	1	0,46
145	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0,49
169	0	1	1	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0,44
183	0	1	0	0	0	1	1	1	0	0	1	0	0	1	0	0	0	0,46

Form-3 Türkiye'de Anormal yanıt desenine sahip bireyler																			
R Studio Programının Maddeleri Analize Dâhil Ettiği Sıra																			
Anormal yanıt veren bireyler	8	4	5	7	13	17	12	15	11	16	14	10	1	6	2	18	3	9	U3 istatistik puanı
	,97	,81	,80	,80	,80	,77	,77	,71	,67	,64	,56	,55	,47	,47	,33	,31	,28	,14	
5	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0,42
7	1	1	1	1	0	1	1	0	1	0	1	1	0	1	0	1	0	1	0,41
10	1	0	1	1	1	1	0	1	0	1	1	0	1	0	1	1	1	0	0,41
49	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0,39
87	0	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0,38
92	1	1	1	0	1	1	1	1	0	1	0	0	1	1	1	1	1	0	0,37
94	1	1	1	1	1	0	1	1	1	0	0	1	1	1	1	1	1	1	0,56
97	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0,37
134	1	0	1	1	1	1	1	1	0	1	1	0	0	1	0	1	1	1	0,44
144	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0,50
158	1	1	1	0	1	1	0	1	1	0	0	1	0	1	1	1	1	0	0,40
162	1	0	1	1	1	1	1	0	1	0	0	0	1	1	0	0	1	1	0,38



GAZİLİ OLMAK AYRICALIKTIR…