

T.C.
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI



BİR BAŞARI TESTİNİN MADDE AYIRT EDİCİLİK
PARAMETRESİ VE GÜVENİRLİĞİNİN İKİ KURAM
ALTINDA KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

AYSU KORKMAZ

TEZ DANIŞMANI
Prof. Dr. İbrahim Alper KÖSE

BOLU, MAYIS - 2024

KABUL VE ONAY SAYFASI

Aysu KORKMAZ tarafından hazırlanan “Bir Başarı Testinin Madde Ayırt Edicilik Parametresi ve Güvenirliğinin İki Kuram Altında Karşılaştırılması” adlı tez çalışması jürimiz tarafından Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı’nda Yüksek Lisans Tezi olarak oy birliğiyle kabul edilmiştir. 27/05/2024

Jüri Üyeleri

İmza

Danışman
Prof. Dr. İbrahim Alper KÖSE
Bolu Abant İzzet Baysal Üniversitesi

.....

Üye
Dr. Öğr. Üyesi Alperen YANDI
Bolu Abant İzzet Baysal Üniversitesi

.....

Üye
Dr. Öğr. Üyesi Ayşe BİLİCİOĞLU-GÜNEŞ
Bartın Üniversitesi

.....

Lisansüstü Eğitim Enstitüsü Onayı

Prof. Dr. İbrahim KÜRTÜL

Lisansüstü Eğitim Enstitüsü Müdürü

ETİK BEYAN

Bolu Abant İzzet Baysal Üniversitesi, Lisansüstü Eğitim Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu bildirir,

aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Teze ilişkin Turnitin adlı programında enstitü müdürlüğünce belirlenen filtrelemeler uygulanarak alınmış olan benzerlik raporuna göre, tezin benzerlik oranı %30'u geçmemektedir.

Bu çalışma için Sosyal Bilimlerde İnsan Araştırmaları Etik Kurulundan 2023/04 sayısı ile etik izin alınmıştır.

.....

AYSU KORKMAZ

ÖZET

BİR BAŞARI TESTİNİN MADDE AYIRT EDİCİLİK PARAMETRESİ VE GÜVENİRLİĞİNİN İKİ KURAM ALTINDA KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

AYSU KORKMAZ

BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANABİLİM DALI

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI

(TEZ DANIŞMANI: PROF. DR. İBRAHİM ALPER KÖSE)

BOLU, MAYIS - 2024

XI + 60

Bu araştırmada Parametrik Olmayan Madde Tepki Kuramı yaklaşımı, Klasik Test Kuramı yaklaşımına alternatif olarak kullanılabilir mi?" sorusuna yanıt aranmıştır. Araştırmanın genel amacı; Klasik Test Kuramı (KTK) ve Parametrik Olmayan Madde Tepki Kuramı (POMTK) yöntemlerine dayalı olarak kestirilen madde ayırt edicilik indeksleri ve güvenirlik katsayıları arasındaki ilişkiyi ortaya koymaktır. Bu amaç doğrultusunda öğretmen yapımı bir başarı testi olan ve ikili olarak puanlanan Eğitimde Ölçme ve Değerlendirme testindeki maddelere ait madde ayırt edicilik indeksleri KTK ve POMTK yöntemleri kullanılarak kestirilmiştir. Araştırma betimsel bir araştırma olup korelasyonel araştırma desenindedir. Araştırma kapsamında Bolu Abant İzzet Baysal Üniversitesinde öğrenim gören 592 katılımcıya ulaşılmıştır. Klasik Test Kuramına dayalı olarak madde ayırt edicilik çift serili korelasyon katsayısıyla, Parametrik Olmayan Madde Tepki Kuramına dayalı olarak madde ayırt edicilik H_1 katsayılarıyla incelenmiştir. Madde güvenirliği ise Cronbach Alpha, eşdeğer yarılar güvenirlik katsayısı, MS (Molenaar- Sijtsma) ve LCRC güvenirlik katsayısıyla incelenmiştir. Analizler R paket programından (mokken 2.8.11) ve SPSS 21.0 paket programından faydalanılarak gerçekleştirilmiştir. Elde edilen bulgular, KTK ve POMTK yöntemine dayalı olarak kestirilen madde ayırt edicilik indeksleri arasında yüksek ve manidar bir ilişki olduğunu, güvenirlik katsayılarının da benzer olduğunu göstermiştir. Bu bulgular doğrultusunda sıralama amaçlı ölçmelerde, kısa testlerde ve küçük örneklemelerde, ölçek geliştirme çalışmalarında POMTK yaklaşımının kullanılabileceği söylenebilir. POMTK yaklaşımının, KTK yaklaşımına göre daha veri odaklı ve keşfedici olması, gruba ve testte yer alan maddelere bağımlı olmaması gibi avantajları göz önüne alındığında verileri daha derinden incelemek amacıyla POMTK, KTK'dan daha çok tercih edilen bir yaklaşım olabilir.

ANAHTAR KELİMELELER: Parametrik Olmayan Madde Tepki Kuramı, Klasik Test Kuramı, İkili Puanlama, Mokken Ölçeklendirme, R Paket Programı.

ABSTRACT

COMPARISON OF ITEM DISCRIMINATION PARAMETER AND RELIABILITY OF AN ACHIEVEMENT TEST ACCORDING TO TWO THEORIES

MSC THESIS

AYSU KORKMAZ

BOLU ABANT İZZET BAYSAL UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF EDUCATIONAL SCIENCES

(SUPERVISOR: PROF. DR. İBRAHİM ALPER KÖSE)

BOLU, MAY 2024

XI + 60

The study aimed to investigate the relationship between item discrimination indices and reliability coefficients estimated based on CTT and NIRT methods. In order to achieve this objective, the item discrimination indices of the items in the Measurement and Evaluation in Education test, which is a teacher-made achievement test and scored dichotomously, were estimated using CTT and NIRT methods. The study employed a descriptive research approach and a correlational research design. The study was conducted with a sample of 592 participants studying at Bolu Abant İzzet Baysal University. Item discrimination was examined using CTT, with the two-series correlation coefficient, and nonparametric item response theory, with Hi coefficients. Item reliability was examined using Cronbach's Alpha, equivalent halves reliability coefficient, MS (Molenaar- Sijtsma) and LCRC reliability coefficients. The analyses were conducted using the R package program (mokken 2.8.11) and SPSS 21.0 package program. The findings indicated a high and significant correlation between the item discrimination indices estimated based on the CTT and NIRT methods, as well as similar reliability coefficients. These findings suggest that the NIRT approach can be employed in scale development studies for ranking purposes, short tests, and small samples. The NIRT approach offers several advantages over the CTT approach. It is more data-oriented and exploratory, and it is not dependent on the group or items in the test. Therefore, it may be a more preferred approach for examining data in greater depth.

KEYWORDS: Nonparametric Item Response Theory, Classical Test Theory, Binary Scoring, Mokken Scaling, R Package Program.

İÇİNDEKİLER

Sayfa

ÖZET	iv
ABSTRACT	v
İÇİNDEKİLER	vi
ŞEKİL LİSTESİ.....	viii
TABLO LİSTESİ.....	ix
KISALTMA VE SEMBOLLER LİSTESİ	x
TEŞEKKÜR	xi
1.GİRİŞ	1
1.1. Problem Durumu	1
1.2. Araştırmanın Amacı	5
1.3. Araştırmanın Önemi	5
1.4. Varsayımlar	7
1.5. Sınırlılıklar.....	7
2.ARAŞTIRMANIN KURAMSAL TEMELİ VE İLGİLİ ARAŞTIRMALAR	8
2.1. Klasik Test Kuramı (KTK).....	8
2.1.1. Klasik Test Kuramına Dayalı Olarak Kestirilen Madde Parametreleri (Madde Geçerliği ve Madde Güvenirliği)	9
2.1.2. Madde Ayırıcılık Gücü İndeksi (Madde Geçerlik Katsayısı).....	9
2.1.3. Madde Güçlük ve Güvenirlik İndeksi.....	13
2.1.4. KTK’de Test Güvenirliği.....	13
2.2. Madde Tepki Kuramı (MTK).....	16
2.3. Parametrik Olmayan Madde Tepki Kuramı (POMTK)	21
2.3.1. Parametrik Olmayan Madde Tepki Kuramı Varsayımları.....	23
2.3.2. Parametrik Olmayan Madde Tepki Kuramı Modelleri.....	30
2.3.3. Parametrik Olmayan Madde Tepki Kuramı Modellerinde Geçerlik ve Güvenirlik İncelemeleri.....	31
2.4. İlgili Araştırmalar	33

3.YÖNTEM	37
3.1. Araştırmanın Modeli	37
3.2. Çalışma Grubu.....	37
3.3. Veri Toplama Aracı.....	37
3.4. Verilerin Toplanması ve Analizi	38
4.BULGULAR	42
4.1. Klasik Test Kuramında Dayalı Olarak Kestirilen Madde Ayırt Edicilik İndeksleri	42
4.2. POMTK Ölçeklenebilirlik Katsayılarının Hesaplanması ve Modelin Belirlenmesi.....	43
4.3. POMTK ve KTK Kuramına Dayalı Olarak Kestirilen Madde Ayırt Edicilik İndeksleri Arasındaki İlişkinin Belirlenmesi	48
5.TARTIŞMA	51
6.SONUÇ VE ÖNERİLER	54
7.KAYNAKLAR	55
8.EKLER	58

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1. Rasch modelinden madde karakteristik eğrileri.....	17
Şekil 2.2. İki parametrelili modelden üç maddeye ait madde karakteristik eğrileri...	18
Şekil 2.3. Üç parametrelili modelden üç maddeye ait madde karakteristik eğrileri...	19
Şekil 2.4. POMTK modeline dayalı olarak kestirilen MKE'ler.....	22
Şekil 2.5. Monoton artan ve kesişmeyen madde karakteristik eğrileri.....	28
Şekil 3.1. Yamaç grafiği.....	40



TABLO LİSTESİ

Sayfa

Tablo 4.1. Eğitimde ölçme değerlendirme testi toplam puanına ilişkin çarpıklık basıklık değerleri.....	42
Tablo 4.2. KTK'ye dayalı olarak kestirilen madde ayırt edicilik indeksleri.....	43
Tablo 4.3. Madde ve ölçek düzeyinde hesaplanan ölçeklenebilirlik katsayıları...	44
Tablo 4.4. Veri setinin Otomatik Madde Seçim İşlemi Yöntemi ile tek boyutluluğunun incelenmesi.....	45
Tablo 4.5. Monoton artma varsayımının incelenmesinde faydalanan değerler.....	46
Tablo 4.6. Madde kalan puanı grubu yöntemiyle yapılan değişmez madde sıralaması incelemesi.....	47
Tablo 4.7. POMTK ve KTK madde ayırt edicilik indeksleri.....	48
Tablo 4.8. KTK ve POMTK madde ayırt ediciliği arasındaki benzerliklerin Spearman Brown Sıra Farkları Korelasyon Katsayısından faydalanarak incelenmesi.....	49
Tablo 4.9. Güvenirlik incelemeleri.....	50

KISALTMA VE SEMBOLLER LİSTESİ

DMS	: Değişmez Madde Sıralaması
İMM	: İkili Monotonluk Modeli
KTK	: Klasik Test Kuramı
MHM	: Monoton Homojenlik Modeli
MKE	: Madde Karakteristik Eğrisi
MTK	: Madde Tepki Kuramı
OMSİ	: Otomatik Madde Seçim İşlemi
POMTK	: Parametrik Olmayan Madde Tepki Kuramı

TEŞEKKÜR

Beni Eğitimde Ölçme ve Değerlendirme alanıyla tanıştıran ve bana bu alanı sevdiren, uzakta da olsak kalbimde sevgisini, güvenini ve desteğini her zaman hissettiğim çok kıymetli hocam Dr. Öğr. Üyesi Emine Burcu Tunç’a,

Yüksek lisans yolculuğum boyunca bu yolu güzelleştiren, sabırla beni dinleyen, bana güvenen, her zaman desteğini sunan ve bilgileriyle yolumu aydınlatan tez danışmanım Prof. Dr. İbrahim Alper Köse’ye,

Ders dönemim boyunca her sorumu bıkmadan usanmadan cevaplayan ve beni farklı araştırmalar yaparak öğrenmeye teşvik eden hocam Dr. Öğr. Üyesi Alperen Yandı’ya,

Hayatım boyunca aldığım tüm kararlarda sabır, sevgi ve şefkatle yanımda olan sevgili ailem Cansu, Sibel, Emine ve Aytekin KORKMAZ’a,

Her yolculukta olduğu gibi yüksek lisans yolculuğumda da bana tüm kalpleriyle eşlik eden değerli dostlarım Simay SACAR ve Sena UÇMUŞ’a,

Bana bu yolculuğa başlama cesareti verdiği için ve yolculuğum boyunca sunduğu destek için Muhammed Ramazan İNCE’ye teşekkür ederim.

1.GİRİŞ

1.1.Problem Durumu

Eğitimciler ve ruh sağlığı uzmanları karmaşık yapıları çözümlmek, tanımlamak ve kategorize etmek için çeşitli ölçme araçları ile psikolojik yapıları ölçmektedir. Başarı, ilgi, tutum, yetenek gibi bireyde var olduğu bilinmesine karşın direkt olarak ölçülemeyen özellikler, psikolojik yapılar olarak tanımlanmaktadır (Volainiuk ve Bokovets, 2017). Psikolojik yapıların doğrudan ölçülememesi testlerin kullanılmasını gerekli kılmıştır. Ölçülen psikolojik yapılar, bireylerin davranışları hakkında oluşturulan kuramların temel yapı taşıdır (Crocker ve Algina, 1986).

Ölçmenin toplumdaki yerinin ne kadar önemli olduğunu gören ölçme uzmanları, geçmişten bugüne değin çeşitli çalışmalar yapmıştır. Gerek yapılan seçimlerin adil olması için gerek vasıflı bireylerin işe seçilmesi için uygun ölçme araçlarına ihtiyaç duyulmuştur. Ölçme aracının istenilen nitelikte olması için ise sağlam bir teorik alt yapısı olan, deneysel verilere uyum sağlayan ve yaşamda uygulanabilirliği olan bir test teorisi gerekli görülmüştür (Lee, 2007).

Test teorisinin tarihsel gelişimi incelendiğinde, Avrupa ve Amerika Birleşik Devletleri'ndeki ruh sağlığı uzmanlarının psikoloji ve eğitim alanında yaptığı araştırmalar sonucunda test teorilerinin geliştiği görülmektedir (Jones ve Thissen, 2006). Psikolojik bir yapıyı ölçmede çoğu araştırmacının sıklıkla karşılaştığı bazı problemler, ölçme teorilerinin gelişmesine öncü olmuştur. Bu problemlere aşağıdaki maddelerde yer verilmiştir (Crocker ve Algina, 1986) :

- Yapının ölçülmesinde tek bir yaklaşımın herkes tarafından kabul görmemesi,
- Psikolojik yapıların ölçülmesinin davranışlara dayalı olması,
- Davranışın örneklendirilmesinin ölçüm hatalarına sebep olması,
- Ölçmedeki birimlerin iyi tanımlanmamış olması,
- Ölçümlerin farklı değişkenlerle ilişkilerinin gösterilmesi şartıyla ölçümlerin anlamlı olması.

Psikolojinin gelişimiyle beraber gelişen test teorileri, bu problemlerden yola çıkarak çeşitli yöntemler geliştirmeyi ve yukarıda yer verilen bu problemleri en aza indirmeyi amaçlamaktadır.

Bireylerin herhangi bir test maddesine verdiği tepkinin altında yatan örtük özelliği açıklamayı amaçlamak üzere oluşturulmuş ilk ölçme kuramı Klasik Test Kuramıdır (KTK). KTK'nin temellerini ilk olarak Gulliksen (1962), 1950 senesinde bir kitapta toplamıştır. KTK'nin istatistik temelli sağlam bir model olmasının ardından Lord ve Novick (1968) bu kuramı açıklayan bir kitap yazmıştır (Akt. Baykul, 2021).

KTK, testten alınan puanın, gerçek puan ve hata puanının toplamına eşit olduğu matematiksel bir formüle dayanmaktadır (Lord ve Novick, 1968). Araştırmacılar çoğunlukla analizlerinde KTK'yi kullanma eğilimindedir. Araştırmacıların çoğunlukla KTK'yi kullanma sebepleri arasında testin uygulanmasının ve parametrelerin kestirilmesinin kolay olması, KTK'nin az sayıda varsayıma sahip olması ve KTK üzerine çok sayıda araştırma yapılmış olması sayılabilir.

KTK' nin, gruba ve testte yer alan maddelere bağımlı olması, güvenilirliğin paralel testler arasındaki korelasyona dayalı olması, bireylerin test maddesine verdiği tepki ve başarı arasında doğrusal bir ilişkinin olması, tüm bireyler için sabit standart hatanın kestirilmesi, test eşitleme ve bilgisayar tabanlı testlere uygun olmaması gibi sınırlılıkları Madde Tepki Kuramının (MTK) geliştirilmesine yol açmıştır. MTK, KTK kadar yaygın olarak kullanılmamakla beraber 1950'li yılından itibaren giderek yaygınlaşan bir kuram olmuştur (Kelecioğlu, 2001).

MTK'nin geliştirilmesiyle beraber KTK'nin sınırlılıklarına çözüm getirilmiştir. Madde ve yetenek parametrelerinin değişmezlik özelliği, MTK ve KTK arasındaki en büyük farktır (Hambleton ve diğerleri, 1991). Değişmezlik özelliği; bireyin yetenek kestiriminin, testte yer alan maddelerden ya da testin uygulandığı gruptan bağımsız olduğu anlamına gelmektedir (Reise ve diğerleri, 2005).

Değişmezlik özelliği, teste yanıt veren bireyin yeteneğinin daha keskin ve daha az ölçme hatasıyla ortaya konmasına, farklı örneklemelerde karşılaştırmalar yapılabilmesine ve kestirilen yetenek düzeylerinin farklı alt test maddelerinde de değişmez kalmasına olanak sağlaması yönlerinden MTK'yi, KTK'den daha kullanışlı hale getirmektedir. Teste yanıt veren bireylerin yetenek/yeterlilik seviyeleri (proficiency levels) karşılaştırılmak istendiğinde MTK, KTK'den daha doğru bir tercih olacaktır.

KTK’de standart hata sabittir ve KTK’ye dayalı olarak geliştirilen ve farklı madde gruplarında yer alan iki testten elde edilen puanların karşılaştırılabilmesi zordur (Lee, 2007).

KTK’ye dayalı olarak geliştirilen bir testte, teste giren bireylerin yeteneklerinin ölçülmesi için testte yer alan tüm maddelere yanıt vermesi beklenmektedir. KTK’de teste yanıt veren bireylerin özellikleriyle, test özelliklerini birbirinden ayırmak mümkün değildir. Bireyin özellikleri ve test özellikleri birbirine bağlı olarak yorumlanabilmektedir (Hambleton ve diğerleri, 1991). MTK bireyin yetenek düzeyinde bir maddeye yanıt vermesine olanak sağlamaktadır çünkü madde karakteristik eğrisi, bireyin bir test maddesine nasıl tepki vereceğini tahmin etmeye olanak sağlar. Dolayısıyla test uzunluğu ve daha kesin ölçme sonuçları için MTK’nin, KTK’den daha uygun bir tercih olduğu söylenebilir (Lee, 2007).

MTK, KTK’den farklı olarak bilgisayar ortamında uygulanan testlere de uygundur. KTK’ye dayalı kâğıt-kalem testlerinde, bireylerin yetenek düzeyleri göz önüne alınmaksızın tüm bireylerden zordan kolaya testteki her maddeye yanıt vermesi beklenmektedir. Bu durum yetenek düzeyleri ortalamanın dışında kalan bireylerin ölçüm hatasını arttırmaktadır. MTK’ye dayalı bilgisayar tabanlı testlerde ise MTK’nin değişmezlik özelliği kullanılarak testi yanıtlayan bireye özgü, onun yetenek düzeyinde maddeler sunulmaktadır. Diğer bir deyişle test, testi yanıtlayan bireye göre özelleştirilmektedir (Embretson ve Reise, 2000).

MTK, KTK’nin sınırlılıklarına çözümler getirmekle birlikte birçok varsayımın da karşılanmasını gerektirmektedir. Bu varsayımlar “tek boyutluluk” ve “yerel bağımsızlık” varsayımlarıdır (Crocker ve Algina, 1986). Varsayımların karşılanmasına ek olarak MTK, KTK’den farklı olarak model-veri uyumunun değerlendirilmesini ve sonrasında uygun modelin seçilmesini gerektirir (Hambleton ve diğerleri, 1991). Ayrıca MTK modellerinde normallik varsayımının sağlanması gerekmemesine karşın normallik sağlanamadığında yetenek kestirimlerinde sapmalar meydana gelebilmektedir (DeMars, 2010). Dolayısıyla MTK’nin üstünlüğü, varsayımların ne derece karşılandığıyla doğru orantılıdır. MTK varsayımlarının her zaman karşılanamaması farklı MTK uygulamalarının gelişmesine sebep olmuştur. Bu MTK uygulamalarından bir tanesi de Parametrik Olmayan Madde Tepki Kuramıdır (POMTK).

Örneklemin ya da madde sayısının az olması ve normallik varsayımının karşılanamaması durumunda MTK yerine POMTK yaklaşımı ile analizler gerçekleştirilmektedir (Mor-Dirlik, 2021).

POMTK, veri odaklı ve keşfedicidir. Madde karakteristik eğrilerinin yalnızca bazı özelliklerini değil, özelliklerinin çoğunu göstermesi sebebiyle MTK'ye göre "daha veri odaklı" kabul edilmektedir (Sijtsma, 2005). MTK'den farklı olarak POMTK'de madde tepki fonksiyonları belirli bir parametrik fonksiyondan bağımsız olarak kestirilmektedir. Bu sebeple parametrik MTK'ye göre POMTK'nin daha esnek bir çerçeve sağladığı söylenebilir (Junker ve Sijtsma, 2001). Junker ve Sijtsma'ya (2001) göre POMTK'nin kullanışlı olmasının sebepleri; POMTK modelleriyle MTK modellerini daha derin olarak inceleyebilmek, MTK'nin gereklilikleri veya model-veri uyumu sağlanamadığında POMTK'nin daha esnek bir perspektif sunması ve madde sayısının ya da teste yanıt veren bireylerin sayısı az olduğunda POMTK'nin kullanabilmesidir (Akt. Molenaar, 2001). POMTK modelleri, sıralama amaçlı ölçmelerde oldukça fayda sağlamaktadır. Psikoloji, siyaset, sosyoloji gibi uygulamalı çeşitli alanlarda sağladığı faydalar kanıtlanmıştır (Sijtsma, 2005). POMTK modelleri; tutum, başarı, yetenek ölçekleri oluşturmak amacıyla kullanılabileceği gibi basit ve çoğunluk tarafından da uygulanabilirdir (Sijtsma ve Molenaar, 2002).

KTK ve POMTK, MTK'ye kıyasla daha az sayıda ve katı olmayan varsayımları olması, daha fazla veri setine uyumlu olması, örneklem ya da madde sayısının az olduğu durumlarda kullanılabilmesi açısından benzerlik göstermektedir. İki kuram arasındaki benzerlik birinin diğerine alternatif olup olamayacağı sorusunu akla getirmektedir.

Tüm bu bilgiler doğrultusunda, KTK ve POMTK'nin benzerlikleri göz önüne alınarak KTK ve POMTK yöntemlerine dayalı olarak kestirilen madde ayırt edicilik parametresi ve güvenilirlik katsayıları arasındaki ilişkinin araştırılması gerekli görülmüştür. Yukarıda değinilen konular doğrultusunda bu araştırma, "POMTK yaklaşımı, KTK yaklaşımına alternatif olarak kullanılabilir mi?" sorusuna yanıt aramaktadır.

1.2.Araştırmanın Amacı

Araştırmanın genel amacı, KTK'ye dayalı olarak geliştirilen, ikili olarak puanlanan ve başarı testi olan Eğitimde Ölçme ve Değerlendirme testindeki maddelere ait madde ayırt edicilik indeksleri ve güvenilirlik katsayılarını KTK ve POMTK yöntemleri kullanarak kestirmek ve iki yöntemden elde edilen kestirimler arasındaki ilişkiyi belirlemektir. Bu genel amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır:

- Eğitimde Ölçme ve Değerlendirme testinin uygun olduğu Mokken modeline dayalı olarak kestirilen madde çiftleri arasındaki ölçeklenebilirlik katsayıları kaçtır?
- Eğitimde Ölçme ve Değerlendirme testinin uygun olduğu Mokken modeline dayalı olarak kestirilen madde ve ölçek düzeyindeki ölçeklenebilirlik katsayıları kaçtır?
- Eğitimde Ölçme ve Değerlendirme testi Mokken modellerinden hangisine uyum göstermektedir?
- POMTK ve KTK'ya dayalı olarak kestirilen madde ayırt edicilik parametreleri arasında manidar bir ilişki var mıdır?
- POMTK ve KTK'ya dayalı olarak hesaplanan güvenilirlik katsayıları ne düzeydedir?

1.3.Araştırmanın Önemi

Alanyazında çeşitli test teorilerinin ve bu test teorilerine dayalı olarak geliştirilen yöntemlerin karşılaştırıldığı birçok araştırma bulunmaktadır fakat bu araştırmalar çoğunlukla KTK ve parametrik MTK kuramlarını karşılaştırmaktadır (Fan, 1998; McDonald ve Paunonen, 2002; Çelen ve Aybek, 2013; Courville, 2005; Bulut, 2018; Sayın ve Koğar, 2019).

KTK ve POMTK'yi karşılaştıran araştırmalar incelendiğinde ulusal alanyazında yalnızca bir araştırma bulunduğu görülmektedir (Mor-Dirlik, 2021). Mor-Dirlik (2021) duyuşsal bir testten elde ettiği verileri kullandığı araştırmasında, iki yaklaşıma dayalı olarak kestirilen madde ayırt edicilik indeksleri arasında yüksek ve manidar bir ilişki olduğu sonucuna ulaşmıştır.

Uluslararası alanyazında ise KTK ve POMTK'yi karşılaştıran ve bilişsel bir testin kullanıldığı yalnızca bir araştırma olduğu görülmüştür. Delevar ve diğerleri (2006), KTK ve POMTK yaklaşımlarından elde edilen madde parametrelerini katılımcıların özellikleri açısından karşılaştırmıştır. İki farklı grubun bulunduğu araştırmadan elde edilen bulgular, KTK yaklaşımına göre kestirilen madde istatistiklerinin gruplara göre değişirken POMTK yaklaşımına göre kestirilen madde istatistiklerinin gruplara göre değişmediğini göstermiştir.

Bilişsel bir testin kullanıldığı ve POMTK ve KTK yaklaşımları arasında karşılaştırma yapılan bir araştırmaya hem ulusal hem uluslararası alanyazında rastlanmamıştır. Bu araştırma, başarı testi kullanarak KTK ve POMTK'ye dayalı madde parametrelerini ve güvenirliği karşılaştırması yönünden diğer araştırmalardan farklılaşmaktadır.

POMTK yaklaşımının, MTK modellerine göre daha esnek bir perspektif sunması ve madde sayısının ya da teste yanıt veren bireylerin sayısı az olduğunda kullanılabilmesi, sıralama amaçlı ölçmelerde fayda sağlaması ve uygulamalı alanlarda kullanılabilmesi gibi avantajları bulunmaktadır. Ayrıca, KTK yaklaşımıyla benzer olarak daha az sayıda ve katı olmayan varsayımları olması ve daha fazla veri setine uyumlu olması da diğer avantajlarındandır.

POMTK'nin avantajları ve KTK ile kullanım kolaylığı açısından benzerliği göz önüne alındığında, iki kuramın karşılaştırılmasının alanyazına katkı sağlayacağı ve araştırmacıları bu konuda daha fazla araştırma yapmaya teşvik edebileceği, bununla birlikte POMTK araştırmalarının da önünü açabileceği düşünülmektedir. Aynı zamanda bu araştırmanın, eğitim ortamlarında POMTK'nin uygulanabilirliğini keşfetmeleri ve beraberinde POMTK'yi okul ortamına entegre etmeleri konusunda eğitimcileri cesaretlendirmesi umulmaktadır.

POMTK gelişmekte olan bir alandır. Bu araştırmanın yeni gelişen bir yaklaşıma katkı sunabileceği düşünülmektedir. Tezin sonunda veri analizinde kullanılan kodlar paylaşılmıştır. Diğer yandan yeni araştırmacılar için bu kodların POMTK'ye ve R kullanım sürecine bir giriş niteliğinde olabileceği umulmaktadır.

1.4.Varsayımlar

- Araştırmaya katılan bireylerin, teste verdikleri tepkiler yeteneklerini yansıtmaktadır.

1.5.Sınırlılıklar

- Araştırmada POMTK modellerinden yalnızca Monoton Homojenlik Modeli (MHM) kullanılmıştır.
- Araştırma bulguları, araştırma kapsamında ulaşılan bireylerin yanıtlarıyla sınırlıdır.
- Araştırmada yalnızca ikili puanlanan bilişsel bir başarı testi kullanılmıştır.
- Araştırmada kullanılan veriler, tez danışmanı tarafından geliştirilen testle sınırlı kalmaktadır.
- Araştırmada kullanılan başarı testi sonuçları sadece Bolu'da üniversite okuyan üniversite öğrencilerine aittir.
- Araştırmada kullanılan Eğitimde Ölçme ve Değerlendirme testinin geliştirilmesi için büyük madde havuzları kullanılamamıştır. Araştırma geliştirilen 40 madde ile sınırlıdır.

2. ARAŞTIRMANIN KURAMSAL TEMELİ VE İLGİLİ ARAŞTIRMALAR

Hem eğitim alanında hem de psikoloji alanında, geçerli ve güvenilir ölçümler yapılabilmesi için uygun istatistiksel, matematiksel ve teorik alt yapılara dayanan ölçme araçlarına ihtiyaç duyulmaktadır (Keeves, 1988). Bireylerin herhangi bir test maddesine verdiği tepkinin altında yatan örtük özelliği açıklamak amacıyla çeşitli ölçme kuramları oluşturulmuştur. Bu kuramlar: Klasik Test Kuramı, Madde Tepki Kuramı, Bilişsel Tanı Modelleri ve Genellenebilirlik Kuramıdır. Bu bölümde, araştırma kapsamında Klasik Test Kuramı, Madde Tepki Kuramı ve Parametrik Olmayan Madde Tepki Kuramı açıklanmıştır.

2.1.Klasik Test Kuramı (KTK)

KTK; testten alınan puanın, gerçek puan ve hata puanının toplamına eşit olduğu matematiksel bir formüle dayanmaktadır. KTK'nin matematiksel gösterimine Eşitlik.2.1'de yer verilmiştir (Lord ve Novick, 1968):

$$X = T + E \quad (2.1)$$

Eşitlik.2.1.'de yer alan matematiksel gösterimde; X testten alınan puanı (gözlenen puan), T gerçek puanı ve E hata puanını ifade etmektedir. Gerçek puan, araştırmacının ölçmeyi hedeflediği yapının “gerçek değerini” ifade etmektedir. Gerçek puan dolaylı yollarla elde edilmekte ve gözlenen puanlardan kestirilmektedir (Baykul, 2021).

Hata puanı ise ölçme yapan bireyden, ölçme yapılan ortamdan ya da çeşitli başka sebeplerden kaynaklanan ve kontrol edilemeyen puanı ifade etmektedir. Hata puanları, ölçmenin yapıldığı koşullara bağlı olmakla beraber hata puanının hangi koşula bağlı olarak ortaya çıktığı kesin olarak bilinmemektedir (Lord ve Novick, 1968). Bu hata türü, ölçmede “tesadüfi hata” olarak adlandırılmaktadır. Diğer hata türlerinde (sabit ve sistematik hatalar) hatayı yapan, hata miktarı ve hata yönü bellidir. Tesadüfi hatada ise kaynak, yön ve miktar belirsizdir dolayısıyla diğer hata türleri gibi matematiksel işlemler uygulayarak düzeltilememektedir. Ölçmede hata teorileri, tesadüfi hata üzerine kurulmaktadır (Baykul, 2021).

Her ölçme kuramı, belirli varsayımlara dayanmaktadır. KTK'nin dört varsayımı bulunmaktadır. Bu varsayımlar aşağıda sunulmuştur (Lord ve Novick, 1968):

- Sonsuz ölçümlerde hata puanlarının ortalaması sıfırdır.
- Gerçek puan ve hata puanı arasında ilişki yoktur.
- Bir ölçüme ait hata puanı ve başka bir ölçüme ait gerçek puan arasında ilişki yoktur.
- İki farklı ölçümdeki hata puanlarının arasında ilişki yoktur.

2.1.1. Klasik Test Kuramına Dayalı Olarak Kestirilen Madde Parametreleri (Madde Geçerliği ve Madde Güvenirliği)

Bir test, testin puanlanabilen en küçük birimi olan maddelerden oluşmaktadır. Her bir maddenin özellikleri bulunmakla beraber bu özellikler, nicel olarak ifade edilebilmektedir. Maddenin özelliklerini gösteren bu nicel ifade ise madde parametreleri olarak adlandırılmaktadır (Baykul, 2021). Madde güçlük indeksi, madde varyansı, madde standart sapması, madde çarpıklık/basıklık katsayısı, madde güvenirlilik indeksi, madde ayırt edicilik (güvenirlilik) indeksi madde parametreleridir.

2.1.2. Madde Ayırtıcılık Gücü İndeksi (Madde Geçerlik Katsayısı)

Madde ayırt ediciliği bir maddenin, testte ölçülen davranışın gözlendiği ve gözlenmediği bireyleri hangi derecede ayırt edebildiğini göstermektedir (Turgut ve Baykul, 2019). Madde ayırt ediciliği katsayısı, -1 ve +1 arasında değerler almaktadır. Negatif değerler, maddelerin ters yönde çalıştığını göstermektedir. Başka bir deyişle negatif değer alan maddeler, aracın ölçmeyi amaçladığı özelliğin tam tersini ölçmektedir (Turgut ve Baykul, 2019). Bu sebeple ölçme aracında bu maddelere yer verilmemelidir.

Madde ayırt edicilik indeksi 0.19 değerinden düşük ise o maddenin zayıf bir madde olduğu ve testten çıkarılması gerektiği, 0.19-0.29 aralığında bir değer alırsa maddenin revize edilmesi gerektiği, 0.30-0.39 aralığında bir değer alırsa iyi bir madde olduğu fakat bazı değişiklikler yapılması gerektiği, 0.40 ve üzerinde değer alırsa ise iyi bir madde olduğu belirtilmektedir (Koğar, 2022). Madde ayırt edicilik gücü; a) madde-toplam test korelasyonlarıyla, b) alt-üst %27 grup ortalamalarının karşılaştırılmasıyla ve c) regresyon analiziyle hesaplanabilmektedir (Erkuş, 2003).

Madde ayırıcılık gücü indeksi ise madde puanları ve test puanları arasındaki ilişkiyi tanımlamaktadır. Diğer bir deyişle testten alınan toplam puan, madde ayırt ediciliği için ölçüttür (Baykul, 2021). Madde ayırıcılık gücü indeksi, madde-toplam test korelasyonlarıyla hesaplanırken bazı katsayılardan yararlanılmaktadır. Bu katsayılar: Nokta çift serili, çift serili, dörtlü ve tetrakorik katsayılarıdır. Aşağıda bu katsayıların formülleri ve hangi durumda, hangi korelasyon katsayısının kullanılması gerektiği açıklanmaktadır.

Nokta Çift Serili Korelasyon Katsayısı: Test puanlarının sürekli olduğu ve testte ikili puanlanan (gerçek iki kategorili) maddelerin yer aldığı durumlarda “nokta çift serili korelasyon katsayısı” kullanılmaktadır (Baykul, 2021). Nokta çift serili korelasyonun formülüne Eşitlik.2.2’de yer verilmiştir (Crocker ve Algina, 1986):

$$p_{pbis} = [(\mu_+) - (\mu_x) / \sigma_x] \cdot \sqrt{p/q} \quad (2.2)$$

μ_+ : Maddeyi doğru cevaplayanların ortalama test puanı

μ_x : Tüm grubun ortalama test puanı

σ_x : Standart sapma

p : Madde güçlük indeksi

q : 1-p

Testte yer alan madde sayısı 25’ten az olduğunda madde puanının, testi cevaplayan her bireyin toplam puanına katkısının olması sebebiyle nokta çift serili korelasyon katsayısı yanıltıcı olabilmektedir. Böyle bir durumda Crocker ve Algina (1986), düzeltme yapılmasını önermektedir. Eşitlik.2.3’te düzeltme formülüne yer verilmiştir:

$$p_{i(X-i)} = p_{Xi} \sigma_X - \sigma_i / \sqrt{\sigma_i^2 + \sigma_X^2 - 2p_{Xi}\sigma_X\sigma_i} \quad (2.3)$$

Eşitlik.2.3’te yer alan $p_{i(X-i)}$, bir madde puanı ile o maddenin çıkarıldığı toplam puan arasındaki korelasyon, σ_x toplam standart sapma ve σ_i madde standart sapmasıdır.

Çift Serili Korelasyon Katsayısı: Sürekli bir değişken (testten elde edilen puanlar) ve yapay iki kategorili bir değişken arasındaki ilişki hesaplanırken “çift serili korelasyon” kullanılmaktadır. Çift serili korelasyon katsayısı, toplam puan ve doğru-yanlış cevaplar arasındaki ilişkiyi bulmaya yaramaktadır (Büyüköztürk ve diğerleri, 2017). Thorndike’a (1982) göre madde ayırt edicilik indeksinin hesaplanması gerektiğinde bu katsayıdan yararlanılabilir (Akt. Baykul, 2021). Eşitlik.2.4’te çift serili korelasyon katsayısının formülüne yer verilmiştir (Crocker ve Algina, 1968):

$$p_{bis} = [(\mu_+) - (\mu_x) / \sigma_x] \cdot (p/Y) \quad (2.4)$$

Eşitlik.2.4’te yer alan çift serili korelasyon formülünde, Eşitlik.2.4’te yer alan nokta çift serili korelasyon katsayısı formülünden farklı olarak “Y” değeri yer almaktadır. Y, standart normal dağılım eğrisinde P ve Q alanlarını ayıran ordinatları ifade etmektedir.

Çift serili korelasyonla hesaplanan ayırt edicilik indeksi, nokta çift serili korelasyonla hesaplanan ayırt edicilik indeksinden daha büyüktür. Lord ve Novick’e (1968) göre bu büyüklük orta zorluktaki maddeler için bir problem oluşturmazken 0,25’in altındaki değerlerde ve 0,75’in üzerindeki değerlerde iki korelasyon katsayısı arasındaki fark keskin bir şekilde artmaktadır (Crocker ve Algina, 1986).

Nokta çift serili ve çift serili korelasyon katsayılarından yararlanılarak madde ayırt ediciliğinin yorumlanması için normallik varsayımının karşılanması gerekmektedir. İki yöntemin de avantajları ve dezavantajları bulunmakla beraber testin kullanım amacına ve elde edilen ölçümlerin durumuna göre uygun yöntemin araştırmacı tarafından tercih edilmesi gerekmektedir (Baykul, 2021).

Dörtlü (Phi) Katsayısı ve Tetrakorik Katsayısı: Dörtlü (Phi) katsayısı süreksiz iki değişkenin arasındaki ilişkiyi hesaplamak için kullanılmaktadır (Baykul, 2021). Tetrakorik katsayısı ise test geliştirme sürecinde araştırmacı tarafından bir değişkenin ikiye bölünmesi ve oluşturulan bu yapay iki değişken arasındaki ilişkinin gücünün araştırmak istenmesi durumunda kullanılmaktadır (Crocker ve Algina, 1986).

İkili puanlanan maddeler için yukarıda açıklanan korelasyon katsayılarından hangisinin seçilmesi gerektiğine ilişkin Crocker ve Algina (1968) bazı önerilerde bulunmuştur:

- Maddeler orta güçlükte ise ayırt edicilik korelasyon katsayıları benzer sonuçlar vermektedir.
- Araştırmacı, madde güçlüklerinin ucunda yer alan maddeleri seçmek istiyorsa nokta çift serili korelasyon katsayısı tercih edilmelidir. Aynı zamanda normal dağılım varsayımı da karşılanmalıdır.
- Test geliştiren araştırmacı, örneklem grubunun yetenek bakımından diğer gruplardan (analizlerin gerçekleştirildiği gruptan farklı bir grup) farklı olacağını düşünüyorsa çift serili korelasyon kullanmalıdır.
- Test geliştiren araştırmacı, örneklem grubunun yetenek bakımından diğer gruplarla aynı olacağını düşünüyorsa nokta çift serili korelasyon kullanmalıdır.
- Madde ve ölçüt değişken ikili olarak puanlanıyorsa dörtlü veya tetrakorik korelasyon katsayısı kullanılmalıdır.

KTK, testin uygulanması ve parametrelerin kestirilmesi açısından kolay olması, az sayıda varsayıma sahip olması ve üzerine çok sayıda araştırma yapılmış olması açısından sıklıkla tercih edilmektedir (Kelecioğlu, 2001). Bunların yanı sıra KTK'nin bazı sınırlılıkları bulunmaktadır. Bu sınırlılıklar test sonuçlarının gruba bağımlı olması ve testte yer alan maddelere bağımlı olmasıdır.

KTK bağlamında “yetenek”, bireye uygulanan testte gözlenen puanın beklenen değeri olarak ifade edilebilir. Bu kavramdan yola çıkarak yeteneğin, bir test bağlamında açıklandığı görülmektedir. Örneğin bir testin zor olması durumunda bireyin yeteneği de düşük olarak yorumlanmaktadır. Testin kolay olması durumunda ise tam tersi bir yorum söz konusu olacaktır. Dolayısıyla yetenek, teste giren bireylerin yeteneğine yani gruba ve testte yer alan maddelerin zor/kolay olmasına bağımlı olacaktır. Grubun ve test maddelerinin değiştirilmesi durumunda bireyin yeteneği de değişecektir. KTK'nin bu sınırlılığı, farklı teste yanıt veren bireyleri karşılaştırmayı zorlaştırmaktadır.

KTK'de maddelerden ziyade teste odaklanılmaktadır. KTK, “Testi cevaplayan bireyler, maddelere nasıl tepki verdi?” sorusunu yanıtlanamamaktadır. Bu sebeple KTK, bir bireyin bir testte ne derece başarılı olacağı hakkında bilgi verememektedir (Hambleton ve diğerleri, 1991).

2.1.3. Madde Güçlük ve Güvenirlik İndeksi

Madde güçlük düzeyi, maddenin hangi gruba zor ya da kolay geldiğiyle ilgilidir. Dolayısıyla, başarı ya da yetenek gibi testlerde kullanılır. Madde güçlüğü, maddeye verilen doğru yanıtların bütün gruba oranlamasıyla bulunur ve bu oran “madde güçlük indeksi” olarak adlandırılır. Bu oran 0-1 aralığında değişmekte olup oranın 1’e yaklaşması maddenin gruba kolay geldiğini göstermektedir (Erkuş, 2003). Araştırmacılara önerilen, 0.50 civarı madde güçlük indeksine sahip maddelerle test geliştirmektir (Ghiselli ve diğerleri, 1981).

Madde güvenirlik indeksi ise madde ayırtıcılık gücü indeksi ile madde standart sapmasının çarpımıyla hesaplanmaktadır. Madde güvenirliğinin hesaplandığı formül Eşitlik.2.5’de yer almaktadır (Baykul, 2021):

$$\rho_j = \rho_{jx} \cdot \sigma_j \quad (2.5)$$

Eşitlik.2.5’te yer alan ρ_{jx} , madde ayırt edicilik indeksini; σ_j , madde standart sapmayı ifade etmektedir.

2.1.4. KTK’de Test Güvenirliği

Ölçme araçlarından, gerçek puana olabildiğince yakın gözlenen puanlar elde etmesi beklenmektedir. Hata puanları azaldıkça, gözlenen puanlar gerçek puanlara yaklaşmaktadır. Ölçme araçlarından elde edilen puanlar, tesadüfi hatalar ne kadar azsa o kadar güvenilir olmaktadır. Turgut’a (1977) göre güvenirlik, “ölçme sonuçlarının tesadüfi hatalardan arınlık derecesi” olarak tanımlanmaktadır (Akt. Baykul, 2021).

Tesadüfi hataların kaynağı tam olarak bilinmese dahi nelerin bu tür hatalara sebep olabileceği tahmin edilebilmektedir. Güvenirlik, hata kaynaklarına ya da tesadüfi hata sebeplerine göre farklı anlamlarda kullanılmaktadır. Aşağıda KTK’de güvenirliğin anlamlarına yer verilmiştir (Baykul, 2021):

- **Duyarlılık:** Ölçme araçlarının/sonuçlarının birim büyüklüğünün, güvenirliğe olan etkisi duyarlılık kavramıyla açıklanmaktadır. Birim küçüldükçe, duyarlılık artmakta dolayısıyla ölçme sonuçları daha güvenilir olmaktadır.
- **Kararlılık:** Bir nitelik, aynı ölçme aracıyla tekrar tekrar ölçüldüğünde aynı sonuç elde ediliyorsa ölçme sonuçlarının kararlı olduğu ifade edilmektedir. Kararlı ölçme sonuçları ise güvenilirdir.

- **Tutarlılık:** Psikoloji ve eğitim alanlarında güvenilirlik, genel olarak ölçme araçlarıyla ilgilidir. Ölçme aracını oluşturan her bir maddenin birbiriyle bütünlük içerisinde olması beklenmektedir. Bu güvenilirlik ise “tutarlılık” ya da “iç tutarlılık” olarak adlandırılmaktadır. Cronbach Alfa (α), KR-20 ve KR-21 katsayıları testin iç tutarlılığına dair bilgi vermektedir. KR-20 ve KR-21 katsayıları, ikili puanlanan maddelerin (çoktan seçmeli, doğru-yanlış) iç tutarlılığına dair bilgi verirken, Cronbach Alfa katsayısı ikili puanlanan maddeler dışındaki maddelerin de iç tutarlılığı hakkında bilgi vermektedir.

Bir test, aynı kişilere benzer koşullarda tekrar uygulandığında, aynı sonuçları vermelidir. Testten elde edilen sonuçların istenen tutarlılığı ya da tekrarlanabilirliği (reproducibility) “güvenirlik” olarak adlandırılmaktadır (Crocker ve Algina, 1986).

Güvenirlik indeksinin doğrudan hesaplanamaması, güvenilirliği sayısal olarak ifade etmeye yarayan güvenilirlik katsayısı kavramını ortaya çıkarmıştır. Güvenirlik katsayısı, birbiriyle paralel olan ölçmeler arasındaki korelasyon katsayısıdır. Bu katsayı, teorik olarak ± 1 arasında değişiyor olsa da $p(X, X') = p^2$ (X, T) eşitliğinden dolayı güvenilirlik katsayısı negatif değerler alamaz. Dolayısıyla güvenilirlik katsayısı 0-1 değerleri arasında değerler alabilmektedir. Grubun homojen ya da heterojen olması, test uzunluğu ve testi yanıtlayan bireylerle ilgili faktörler testin güvenilirliğini etkilemektedir.

Güvenirlik katsayısı çeşitli yöntemlerle hesaplanabilmektedir. Bu yöntemler: Test-tekrar test, paralel formlar, eşdeğer yarılar, maddelerin testin bütünüyle korelasyonu ve varyans analizine dayanan yöntemlerdir. Aşağıda test-tekrar test, paralel formlar ve eşdeğer yarılar yöntemleri açıklanmıştır (Baykul, 2021).

Test-Tekrar Test Yöntemi: Test-Tekrar Test yönteminde; aynı test, aynı gruba farklı zamanlarda uygulanmakta ve iki uygulama arasındaki korelasyon hesaplanmaktadır (Özgüven, 2020). Bu yöntemden elde edilen güvenilirlik katsayısına “kararlılık katsayısı” da denmektedir.

İki uygulama arasındaki korelasyonun düşük olmasının temelde iki sebebi bulunmaktadır: Bunlardan ilki ölçme aracından, ikincisi testi cevaplayan bireylerden kaynaklanan sebeplerdir (Thorndike ve Thorndike-Christ, 2017).

Bireylerden kaynaklanan sebepler, iki ölçme arasındaki zamanın ayarlanmasında göz önünde bulundurulmaktadır. Aynı zamanda test maddelerinin hatırlanması ve yeni öğrenmeler gibi etkenler de bu zamanın belirlenmesinde önem arz etmektedir. Bu zamanın ne kadar olması gerektiği hakkında kesin bir görüş bulunmamakla beraber yeni öğrenmelerin, hatırlamaların ve pratiğin olmayacağı kadar uzun; cevaplayan bireylerin ölçülen özelliğinde değişim olmayacak kadar kısa olması tavsiye edilmektedir. Bu yöntemin; ölçülmesi amaçlanan değişkende, zamana bağlı olarak değişimlerin olmayacağı ya da minimum düzeyde olacağı ön görüldüğü durumlarda ve ölçme aracının ne derece kararlı olduğu saptanmak istendiğinde kullanması önerilmektedir.

Paralel Formlar Yöntemi: Aynı özelliği ölçmeyi amaçlayan testler, paralel/eşdeğer formlar olarak ifade edilmektedir. Gün içerisinde farklı oturumlarda ya da farklı günlerde uygulanacak olan testlerde, teste ait maddelerin teste girmeyenler tarafından öğrenilmesini önlemek amacıyla kullanılabilen bir yöntemdir. Bazı zamanlarda, katılımcıların teste verdiği yanıtların birbirinden etkilenmesini (kopya gibi durumların önüne geçmek) önlemek amacıyla da kullanılabilmektedir.

Paralel formlar yönteminde, aynı özelliği ölçmeyi amaçlayan iki test geliştirilir. Geliştirilen iki test, aynı gruba uygulanır. Ardından iki gruptan elde edilen sonuçlar arasındaki korelasyon hesaplanır (Thorndike ve Thorndike-Christ, 2017). Paralel formların ortalama ve standart sapmalarının eşit olması gerekmektedir. Dolayısıyla iki testin de ortalama ve standart sapma değerlerinin kontrol edilmesi önem arz etmektedir.

Eşdeğer Yarılar Yöntemi: Eşdeğer yarılar yönteminde, testte yer alan maddeler iki alt teste ayrılmaktadır. Testi iki yarıya bölerken; 1) Maddeler rastgele iki alt teste ayrılabilir, 2) Tek ve çift maddeler olarak iki alt test oluşturulabilir, 3) Madde güçlük indeksleri sıralandıktan sonra tek ve çift olarak maddeler ikiye ayrılabilir, 4) Aynı kapsamdaki maddelerden iki ayrı alt test oluşturulabilir (Turgut ve Baykul, 2019).

Testin iki alt teste bölünmesinin ardından iki alt test arasındaki korelasyon hesaplanmaktadır. Testin tamamına ait güvenilirlik katsayısı için ise Sperman-Brown formüllerinden faydalanılmaktadır.

2.2.Madde Tepki Kuramı (MTK)

KTK'nin sınırlılıkları, MTK'nin gelişmesine sebep olmuştur. Tucker'ın 1940 yılında “madde karakteristik eğrisi” kavramını ilk defa kullanmasının ardından 1950 yıllarında Lord, MTK ile ilgili çeşitli çalışmalar yapmıştır. Lord'un çalışmaları, MTK'nin temellerini oluşturmuştur (Akt. Doğan ve Tezbaşaran, 2003). Teknolojinin gelişmesiyle beraber matematiksel hesaplamaların daha kolay yapılabilmesi, başlangıçta yavaş ilerleyen MTK'nin gelişim sürecini hızlandırmıştır.

MTK'nin dayandığı iki önerme bulunmaktadır: 1) Bireyin uygulanan bir testteki performansı; gizil özelliklerle ya da yetenek adı verilen bazı faktörlerce açıklanabilmektedir. 2) Teste giren bireylerin, bir maddeye verdikleri tepkilerin ve bu tepkiye yol açan özelliklerin ilişkisi, “madde karakteristik eğrisi” adı verilen bir fonksiyonla açıklanmaktadır. Madde karakteristik eğrisi, testin ölçtüğü yetenek ve madde özelliklerini temel alarak bir maddeye doğru yanıt verme olasılığının matematiksel bir ifadesi olarak tanımlanabilir (Hambleton ve diğerleri, 1991).

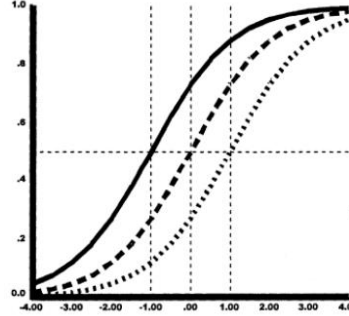
Madde karakteristik eğrisi fonksiyonun farklı sayıda parametreyi kapsayan çeşitli modelleri bulunmaktadır. MTK modellerinin tümünde her parametre kullanılmamaktadır (Spencer, 2004). İkili yanıt verilerinde en çok kullanılan modeller bir, iki ve üç parametrelili lojistik modellerdir. Aşağıda çoğunlukla kullanılan bu üç model açıklanmıştır.

Tek Parametrelili Lojistik Model: Modelde sabit tutulan şans parametresi (c) sıfır yerine bire sabitlendiğinde bu model, Rasch modeli olarak da adlandırılmaktadır (Baker, 2001). Tek parametrelili lojistik modelin formülü Eşitlik.2.6'da yer almaktadır:

$$P_i(\Theta) = e^{(\Theta - b_i)} / 1 + e^{(\Theta - b_i)} \quad (2.6)$$

Formülde yer alan Θ , maddeye tepki veren bireyin yeteneğini/yeterliğini göstermektedir (Spencer, 2004). b parametresi yani güçlük parametresi ise bir bireyin testte yer alan bir maddeyi, yüzde 50 doğru yanıtlama olasılığının olduğu noktayı ifade etmektedir. ± 2 değerleri arasında yer alan b parametresinin değerinin artması, teste yanıt veren bireyin o maddeyi yanıtlamak için gerekli yetenek düzeyinin de artması anlamına gelmektedir. Yani, b değeri ikiye yaklaştıkça testte yer alan maddeleri yanıtlamanın zorlaştığı söylenebilir (Hambleton ve diğerleri,

1991). Şekil.2.1’de Rasch modelinden üç maddeye ait madde karakteristik eğrileri gösterilmiştir (Embretson ve Reise, 2000):



Şekil.2.1. Rasch modelinden madde karakteristik eğrileri

Şekil.2.1’de yer alan Rasch modelinden üç madde için gösterilen madde karakteristik eğrileri (MKE) incelendiğinde, olasılıkların her madde için kademeli olarak arttığı görülmektedir. Burada yer alan maddeler arasındaki tek fark, farklı zorluk düzeyinde olmalarıdır. Dolayısıyla eğrilerin tümünün eğimi eşittir. MKE’lerinin bükülme noktası, bir maddenin doğru yanıtlanma olasılığı .50 olduğunda ortaya çıkmaktadır. Şekil.2.1’de yer alan ve zorluk seviyesi 1.00 olan 3.maddenin doğru cevaplanma olasılığı .50’dir.

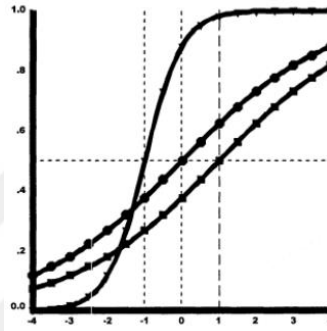
Tek parametrelili lojistik model, iki ve üç parametrelili lojistik modelden farklı olarak şu özelliklere sahiptir: 1) Testte yer alan maddelerin her birinin aynı ayırt edicilikte olduğunu varsaymaktadır. 2) Tahmin yöntemi olarak “koşullu maksimum olabilirlik yöntemi (conditional maximum likelihood)” kullanmaktadır. 3) Dayandığı ölçüm felsefesi farklıdır (Lee, 2007).

İki Parametrelili Lojistik Model: Bu model, tek parametrelili modele ek olarak a parametresinin ve D ölçekleme faktörünün fonksiyona eklenmesiyle oluşur. Formülde yer alan D 1,702 ölçek sabitini göstermektedir. Eşitlik.2.7’de iki parametrelili lojistik modelin formülü yer almaktadır:

$$P_i(\Theta_j) = P(x_i=1 | \alpha_i, \beta_i, \Theta_j) = 1 / 1 + e^{-D a_i (\Theta_j - \beta_i)} \quad (2.7)$$

Eşitlik.2.7’de yer alan formüldeki a parametresi, madde ayırt edicilik parametresi olarak adlandırılmakta olup; MKE’nin b noktasındaki eğimiyle orantılıdır.

Maddenin eğiminin daha dik olması, daha iyi ayırt edicilik anlamına gelmektedir. Daha az ayırt edici olan maddeler, daha düz bir eğime sahip olup farklı yetenek düzeyindeki bireylerin bir maddeye verecekleri tepkinin benzer olasılıkta olduğunu göstermektedir (Spencer, 2004). Teoride $\pm\infty$ aralığında değer alabilen a parametresi değeri, uygulamada genellikle 0-2 ya da 0-3 aralığında değer almaktadır (DeMars, 2010). Negatif değerler, testte ölçülmek özelliğın zıttı bir yönde çalıştığından testten çıkarılmaktadır. Şekil.2.2’de iki parametrelı modelden üç maddeye ait madde karakteristik eğrileri gösterilmiştir (Embretson ve Reise, 2000):



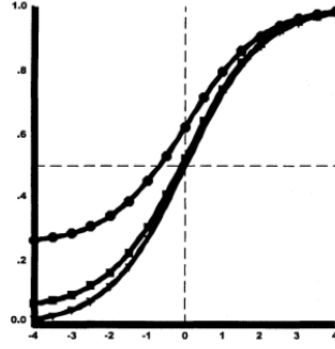
Şekil.2.2. İki parametrelı modelden üç maddeye ait madde karakteristik eğrileri.

Şekil.2.2’de yer alan madde karakteristik eğrileri, zorlukları aynı olan fakat farklı eğimlere sahip olan üç maddeyi göstermektedir. En düşük yetenek düzeyinde bile maddenin doğru yanıtlanma olasılığının sıfırdan büyük olduğu görülmektedir.

Üç Parametrelı Lojistik Model: Bu modelde, b ve a parametrelerine ek olarak fonksiyona c parametresi eklenir. C parametresi, şans parametresi olarak adlandırılmakta ve yeteneğı düşük olan bireylerin bir maddeyi doğru yanıtlanma olasılığını göstermektedir. Eşitlik.2.8’de üç parametrelı lojistik modelin formülü yer almaktadır (Hambleton ve diğerleri, 1991) :

$$P_i(\Theta_j) = P(x_i=1 | \alpha_i, \beta_i, c_i, \Theta_j) = c_i + (1-c_i) \cdot \frac{1}{1 + e^{-D_{ai}(\Theta_j - \beta_i)}} \quad (2.8)$$

Eşitlik.2.8’de görüldüğü gibi, iki parametrelı lojistik modelden farklı olarak üç parametrelı lojistik model formülüne yalnızca c parametresi eklenmiştir. Şekil.2.3’te üç parametrelı modelden üç maddeye ait madde karakteristik eğrileri yer almaktadır (Embretson ve Reise, 2000):



Şekil.2.3. Üç parametrelî modelden üç maddeye ait madde karakteristik eğrileri.

Şekil.2.3'te yer alan üç parametrelî modelden üç maddeye ait madde karakteristik eğrileri, farklı asimptotlara sahip fakat aynı zorluk seviyesindeki (0.00) maddeleri göstermektedir.

Üç parametrelî model, çoktan seçmeli maddelerden oluşan bir testin değerlendirilmesi için ikili modellerden (dichotomous models) en çok tercih edilenidir. Bunun sebebi düşük yetenek düzeyindeki bireylerin, doğru cevabı şans eseri seçme olasılıklarının olmasıdır.

İki parametrelî model ise ikili puanlanan tutum maddeleri (dichotomous attitudinal items) için ve etkili çeldiricilere sahip çoktan seçmeli testler için uygun bir seçim olabilir. Çeldiriciler etkili olduğunda, düşük yetenek düzeyindeki bireylerin çeldirici seçeneğin doğru olduğunu düşünme ihtimali daha yüksek olacaktır. Tek parametrelî modelde ise modele uygun olmayan maddelerin ölçüme fayda sağladığı kabul edilmemektedir. Bazı araştırmacıların daha basit olduğu için veya karmaşık olan bir modelin tahmin edilmesi için gerekli sayıda verinin olmadığı durumlarda bu modeli tercih etmektedir (DeMars, 2010).

Psikolojide farklı anlamlarda kullanılan model kavramı, psikolojik ölçüm kapsamında bağımlı değişkenin en geçerli tahminini elde etmek üzere bağımsız değişkenlerin sayısal ifadeleri olarak tanımlanabilir (Embretson ve Reise, 2000). MTK modelleri; boyut (tek boyutlu ve çok boyutlu modeller), doğrusallık (doğrusal modeller ve doğrusal olmayan modeller) ve testin puanlanması (1-0 puanlanan ve çoklu puanlanan tepki modelleri) olarak üç başlıkta incelenmektedir (McDonald, 1982).

MTK, KTK'den farklı olarak model-veri uyumunun değerlendirilmesini ve sonrasında uygun modelin seçilmesini gerektirir (Hambleton ve diğerleri, 1991). Bir MTK modelinin, verilerle olan uyumu “uyum iyiliği indeksleri (goodness of fit indices)” ile değerlendirilir. Uyumun sağlanmadığı bir modelde, değişmez madde ve yetenek parametreleri de sağlanamaz (Spencer, 2004).

Çoğu MTK modeli iki varsayıma dayanmaktadır. Bu varsayımlar “tek boyutluluk” ve “yerel bağımsızlık” varsayımlarıdır (Sijtsma ve Junker, 2006). Tek boyutluluk varsayımı, testte yer alan tüm maddelerin tek bir özelliği ölçmesi anlamına gelmektedir. Bu varsayım, testin uygulandığı anda bireylerin motivasyonlarından, sınav kaygılarından vb. gibi sebeplerden etkilenebileceğinden dolayı her zaman karşılanmayabilir. Tek boyutluluk varsayımının yeterince karşılanabilmesi için bir özelliğin, başat özellik göstermesi gerekmektedir. Bu başat özellik, testin ölçmeyi amaçladığı yetenektir (Hambleton ve diğerleri, 1991).

MTK'de yapı, gizil özellik olarak adlandırılmaktadır. Gizil özellik, test maddelerine verilen tepkilerin altında yatan, bu tepkileri doğrudan etkileyen ve testte ölçülmesi hedeflenen özellik olarak ifade edilebilir (Reise ve diğerleri, 2005). Tek boyutluluk varsayımının sağlanamaması, MTK modelinin yanlış uygulanmasına, yanlış parametre tahminlerine ve standart hatalara sebep olabilmektedir (Spencer, 2004; DeMars, 2010).

Tek boyutluluk varsayımı birçok yöntemle test edilmektedir. En yaygın olanları “Maddeler arası korelasyon matrisinin öz değer analizi, Stout temel tek boyutluluk testi ve tek boyutlu çözümlemeden kalan atık varyanslara dayanan indekstir (DeMars, 2010).

Yerel bağımsızlık varsayımı bireylerin, testte yer alan her bir maddeye verdikleri yanıtların birbiriyle ilişkili olmadığını ifade eder (Embretson ve Reise, 2000). Bireylerin testte yer alan maddelere vermiş oldukları tepkileri etkileyen şey yetenekleridir. Yetenek, parça parça düşünüldüğünde maddelere verilen tepkilerin birbirine bağlı olmadığı düşünülebilir. Buradan hareketle yerel bağımsızlık varsayımına, koşullu bağımsızlık varsayımı da demek mümkündür (Hambleton ve diğerleri, 1991). Testte yer alan bir maddeyi doğru yanıtlamak için başka bir maddenin gerekli olması ya da bir maddenin başka bir maddeyi doğru yanıtlamak için gerekli bilgiyi içeriyor olması, yerel bağımsızlığın ihlal edilmesine sebep olmaktadır (Embretson ve Reise, 2000).

MTK’de ölçüm hatası ve güvenilirlik “bilgi fonksiyonu” ile hesaplanmaktadır. Bilgi fonksiyonu, madde bilgisi fonksiyonlarının toplamına eşit olup madde seviyesinde tanımlanabilmektedir. Test bilgisi, ölçülen özelliğin ve testte yer alan maddelerin bir fonksiyonu olup yetenek düzeylerine göre değişmektedir. Bilginin artması, standart hatayı azaltmakta ve güvenilirliği arttırmaktadır. Bir teste giren bireylerin yetenek düzeylerinden güvenilirlik tahmini, madde parametrelerine ve testin ölçtüğü özelliğin gruptaki dağılımına göre hesaplanmaktadır (DeMars, 2010).

Testte yer alan maddelerin tanımlanmasında ve seçilmesinde madde bilgi fonksiyonundan yararlanılmaktadır. Madde güçlük düzeyi (b) ve madde ayıricılığının (a), madde bilgi fonksiyonu düzeyine yakın değerler alması, c parametresinin ise sıfıra yaklaşması bilgiyi de arttırmaktadır (Köse, 2010).

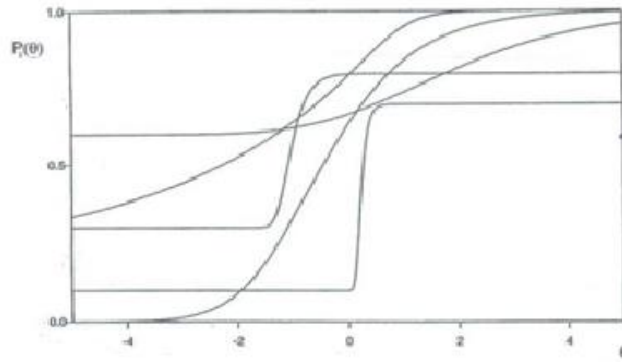
2.3.Parametrik Olmayan Madde Tepki Kuramı (POMTK)

Matematiksel ölçüm modeli ve testlere yanıt veren bireylerin karmaşık dünyası arasında bir boşluk vardır. Testi geliştirirken geçerlik, yorumlama kolaylığı ve uygulama kolaylığı gibi unsurlar her zaman istenilen şekilde sonuçlanmayabilir. Bir testin nitelikli olabilmesi için tüm bu unsurların olabildiğince bir arada olması gerekmektedir. MTK, test geliştirme süreci için hassas bir çerçeve sunmaktadır. POMTK ise daha esnek varsayımlara sahip olması ve geniş bir yöntem sınıfına dönüşmesi sebebiyle özellikle faydalı olabilmektedir (Molenaar, 2001).

Mokken’in (1971) ikili puanlanan maddeler için geliştirdiği POMTK, Molenaar’in (1991, 1997) de çalışmalarıyla çok kategorili puanlanan maddeler için de kullanıma uygun bir yaklaşım haline gelmiştir (Akt. Mor-Dirlik, 2017). Sijtsma ve Molenaar (2002), yayınladıkları “Parametrik Olmayan Madde Tepki Kuramına Giriş (Introduction to Nonparametric Item Response Theory)” kitabında POMTK’yi tanıtmış ve POMTK’nin kuramsal temellerini okuyuculara aktarmıştır.

Meijer ve Baneke (2004) ise POMTK’nin psikolojik özelliklerin ölçülmesinde kullanılan ölçeklerde uygulanabilirliğini tartışmış ve POMTK modellerini, MTK modelleriyle karşılaştırmıştır. Ayrıca POMTK modellerinin kolaylıkla nasıl uygulanacağını da göstermişlerdir. Psikoloji, siyaset, sosyoloji gibi uygulamalı çeşitli alanlarda sağladığı faydaların kanıtlandığı (Sijtsma, 2005) POMTK modelleri, giderek araştırmacıların dikkatini daha fazla çekmeye başlamıştır.

POMTK, parametrik MTK modellerinden ayıran özelliklerden biri de veri setinden madde yanıt fonksiyonlarının tahmin edilmesi ve varsayılan monotonluk durumunun değerlendirilmesidir (Sijtsma, 2005). POMTK’de MTK’den farklı olarak madde karakteristik eğrilerinin parametrik bir formu yoktur. Parametrik modellere göre varsayımlarının daha esnek olmasına rağmen POMTK’nin ölçüm özellikleri istenen şekildedir (Sijtsma ve Junker, 2006). Daha esnek olması sayesinde gizil özelliğin boyutluluğu daha kolay değerlendirilmektedir (Junker ve Sijtsma, 2001). POMTK modellerini belirli bir lojistik forma sahip olmaması, bu modellerin parametrik olmayan modeller olarak adlandırılmasına sebep olmuştur. Şekil.2.4’te POMTK modeline dayalı olarak kestirilen MKE’lere yer verilmiştir (Sijtsma ve Molenaar, 2002):



Şekil.2.4. POMTK modeline dayalı olarak kestirilen MKE’ler

Şekil.2.4’te yer alan MKE’ler incelendiğinde, MKE’lerin lojistik olabileceği gibi aynı zamanda kısmen doğrusal veya üstel olabileceği de görülmektedir. Diğer bir deyişle MKE’lerde sabit/oransal artış ya da sabit/oransal azalma görülmeyebilir.

Bir fonksiyon, tüm tanım aralığında kesin artan ya da azalan ise bu fonksiyonun monoton olduğu söylenebilir. POMTK’de MKE’ler monoton olarak azalmadığı sürece düzensiz, ayırık ve basamaklı olabilmektedir. Şekil.2.4’te yer alan eğriler, aynı zamanda monotonluk özelliğine sahip MKE’lerin de örneklerini sunmaktadır (Sijtsma ve Molenaar, 2002).

POMTK modelleri, MTK’ye kıyasla daha fazla veriye hitap etmektedir. POMTK verilere bazı kısıtlamalar getirmekle beraber MTK’ya kıyasla daha az kısıtlayıcıdır. Bu özelliğiyle POMTK, veri odaklı ve keşfedicidir (Sijtsma, 2005).

POMTK'nin keşfedici özelliğinden, MTK analizi öncesinde verilerin boyutunu incelemek amacıyla yararlanılabilir. Ayrıca MTK için gereken model veri uyumunun sağlanamaması durumunda da POMTK alternatif olabilir (Sijtsma, 2005).

POMTK modelleri, veri odaklı olmakla beraber işe uygun adayın seçilmesi gibi sıralama düzeyinde ölçümler için kullanışlıdır (Sijtsma, 2005). POMTK, bireyleri sıralamak için uygun bir modeldir (Sijtsma ve Molenaar, 2002; Sijtsma ve diğerleri, 2011). POMTK analizleriyle her bir madde ve testin tamamı için bilgi elde etmek mümkündür. Maddelerin, düşük ve yüksek yetenek düzeyindeki bireyleri ayırt ediciliği ve bir maddenin, testte yer alan diğer maddelerle aynı gizil özelliği ölçüp ölçmediği POMTK analizleriyle değerlendirilebilir (Sijtsma, 2005).

Junker ve Sijtsma'ya (2001) göre POMTK'nin kullanışlı olmasının sebepleri; POMTK modelleriyle parametrik MTK modellerini daha derin olarak incelenebilmesi, MTK'nin gereklilikleri veya model-veri uyumu sağlanamadığında POMTK'nin daha esnek bir perspektif sunması ve madde sayısının ya da teste yanıt veren bireylerin sayısı az olduğunda POMTK'nin kullanabilmesidir (Molenaar, 2001). POMTK küçük örneklemelerde de kullanılabilir (Junker ve Sijtsma, 2001).

2.3.1. Parametrik Olmayan Madde Tepki Kuramı Varsayımları

İkili puanlanan maddeler için yapılacak POMTK analizlerinde varsayımların test edilmesinden önce izlenmesi gereken adımlar araştırmacılar tarafından şu şekilde belirtilmektedir (Sijtsma ve diğerleri, 2011) :

- İlk olarak verilerin model için uygunluğu değerlendirilmelidir.
- Veri seti POMTK modeliyle uyumlu ise madde çiftleri, madde ve toplam ölçek puanları H_{jk} , H_j ve H katsayılarıyla yorumlanmaktadır.

Veri setinin POMTK modeline uygunluğu değerlendirildikten sonra POMTK varsayımları kontrol edilmelidir. POMTK'nin dört varsayımı bulunmaktadır. Bu varsayımlar: 1) Tek boyutluluk, 2) Yerel bağımsızlık, 3) Gizil monoton artma/monotonluk ve 4) Kesişmeyen MKE'ler/değişmez madde sıralamasıdır. Dördüncü varsayımın ikili monotonluk modelinde sağlanması gerekirken, monoton homojenlik modelinde ilk üç varsayımın karşılanması yeterlidir.

Tek Boyutluluk (Unidimensionality): Tek boyutluluk varsayımı, testte yer alan maddelerin tamamının aynı gizil özelliği ölçmesi anlamına gelmektedir. MTK kuramı kapsamında da tek boyutluluk varsayımı, verinin yapısını açıklayan tek bir özelliğin olması gerektiğini ifade etmektedir. Tek boyutluluk varsayımı, eğitimde ve psikolojide testten elde edilen sonuçların yorumlanmasını kolaylaştırmaktadır fakat bazı gizil özelliklerin psikolojik olarak daha karmaşık olduğu ve başka gizil özelliklerle beraber yorumlanması gerekebileceği de unutulmamalıdır. Bu noktada araştırmacıların ölçmek istediği yeteneği veya özelliği ölçen ölçme araçlarını seçerken dikkatli olması gerekmektedir (Sijtsma ve Molenaar, 2002).

Test boyutluluk çoğunlukla iki şekilde değerlendirilmektedir: 1) Tek boyutluluk doğrulanır veya çürütülür. 2) Gerekli olması durumunda çok boyutlu yapı ayrıntılı şekilde açıklanır. Gerek MTK'ye dayalı analizlerde gerek POMTK'ye dayalı analizlerde veri setinin tek boyutlu modele uyduğu varsayılmaktadır. Bu durum tek boyutluluğun doğrulanmasını gerekli kılar. POMTK kapsamında tek boyutluluk varsayımının sınanması amacıyla kullanılan çeşitli yöntemler bulunmaktadır. Bunlardan birisi özel bir tahmin prosedürü olan "DETECT" yöntemidir. Bu yöntemde maddeler gruplara bölünmekte, bölünen gruplar arasındaki ikili kovaryanslara bakılmakta ve madde çiftlerinin pozitif değerler alması beklenmektedir (Stout ve diğerleri, 1996).

Test boyutluluğunu incelerken kullanılan bir diğer yöntem ise Otomatik Madde Seçim İşlemi (OMSİ)'dir. Test ve anketleri oluşturmak için kullanılan Mokken Ölçek Analizi, ölçeğin içerisinde yer alan ve ortak tek bir özelliği ölçen maddeleri seçerken otomatik bir seçim prosedürü kullanmaktadır.

Madde seçiminin maddeler arası kovaryanslara bağlı olarak yapıldığı bu yöntem; POMTK modellerinden, Monoton Homojenlik Modeline dayanmaktadır. OMSİ'de maddeler arası kovaryansın negatif olmaması beklenmekte ve değerlerin negatif olup olmadığı ölçeklenebilirlik katsayılarından H katsayısı ile değerlendirilmektedir (Kujipers ve diğerleri, 2016). Maddeler arası kovaryansların pozitif olması tek boyutluluk varsayımını karşılamak için yeterli olmamakta, aynı zamanda H katsayısının 0.3'e eşit ya da 0.3'ten büyük olması beklenmektedir. Çoklu ya da ikili olarak puanlanan maddelerde kullanılabilen OMSİ, faktör analizine bir alternatif olarak görülmektedir (Sijtsma ve Molenaar, 2002).

Yerel Bağımsızlık (Local Independence): Yerel bağımsızlık varsayımı, bireylerin bir maddeye verdikleri yanıtın, diğer maddelere verdikleri yanıtlardan etkilenmediği anlamına gelmektedir. Bu varsayım güçlü bir varsayım olmakla beraber ihlal edilmesi de oldukça kolaydır. Eşitlik.2.9’da yerel bağımsızlığın matematiksel gösterimi yer almaktadır (Sijtsma ve Molenaar, 2002):

$$P(X = x | \Theta) = \prod_{i=1}^k P(X_i = x_i | \Theta) \quad (2.9)$$

Eşitlik.2.9’da yer alan formül doğrultusunda yerel bağımsızlık, sabit bir theta değeri için i ve j maddelerinin arasındaki kovaryansın sıfıra eşit olması olarak da tanımlanabilir.

Yerel bağımsızlık ve tek boyutluluk varsayımları ilişkili olmakla beraber aynı anlama gelmemektedir (Sijtsma ve Molenaar, 2002). DeMars (2010) tek boyutluluk ve yerel bağımsızlık arasındaki ilişkiye dikkat çekmiş, tek boyutluluk varsayımının sağlandığı fakat yerel bağımsızlığın sağlanmadığı durumlarda bağımlılığa başka bir boyutun sebep olabileceğini ifade etmiştir. Dolayısıyla tek boyutluluk varsayımının sağlanması, yerel bağımsızlık varsayımının karşılanması için ön koşulu olarak görülmektedir. Yerel bağımsızlığın test edilmesinde maddeler arasındaki bağımlılık incelenmektedir. Bu bağımlılıkların büyük bir madde grubunu etkilemediği zamanlarda ayrı boyutlar ortaya çıkmayabilir. Bu durumda da yerel bağımsızlığın ihlali, tek boyutluluğu incelemeye yönelik yöntemlerle ortaya çıkmayabilir. Sonuç olarak yerel bağımsızlık varsayımını incelemek için farklı yöntemler geliştirilmiştir (DeMars, 2010).

Yerel bağımsızlık varsayımının incelenmesinde, Q_3 analizi (Yen, 1984) ve Stout analizi kullanılmaktadır. Yen’in (1984) önerdiği test, Stout analizinden farklı olarak her bir madde çiftini ayrı ayrı incelemektedir. Bu yönüyle daha güçlü bir test olduğu söylenebilir (DeMars, 2010). Faktör analizinden faydalanarak da yerel bağımsızlığın karşılanıp karşılanmadığı kısmen anlaşılabilmektedir (Embretson ve Reise, 2000). Madde çiftleri arasındaki korelasyon katsayıları incelenerek yerel bağımsızlığın test edilmesi mümkündür. Hambleton ve diğerleri (1991) de faktör analizinin yerel bağımsızlığı test etmekte kullanılabileceğini belirtmiştir.

Strout (1990) ise yerel bağımsızlık yerine yerel bağımsızlıktan daha az kısıtlayıcı olan “temel bağımsızlık” (essential independence) kavramının özelliklerini ve performansını araştırdığı çalışmasında, belirli yetenek düzeylerinde maddeler arası kovaryansların sıfıra eşit olmasının yerel bağımsızlık varsayımının karşılandığını gösterdiğini ifade etmiştir.

Bilişsel testlerde yerel bağımsızlığın ihlal edilmesine sebep olabilecek bazı durumlar söz konusudur. Bir maddeyi doğru yanıtlayabilmek için başka bir maddenin gerekmesi, farklı maddelerde aynı ifadelerin yer alması ya da aynı köke atıfta bulunan ifadelerin yer alması yerel bağımsızlığın ihlal edilmesiyle sonuçlanabilir. Duyuşsal testler de ise bireyler, orta derecede olumsuz bir maddenin önünde olumsuz ifadelerin yer alması durumunda o maddeyi de olumsuz işaretleme eğiliminde olabilirler (Embretson ve Reise, 2000).

Gizil Monoton Artma/Monotonluk (Monotonicity of IRFs): POMTK’de bireylerin bir maddeye doğru yanıt verme olasılıkları, gizil yeteneğin monoton bir şekilde azalmadığı bir fonksiyon olarak tanımlanmaktadır (Sijtsma ve Molenaar, 2002). Monoton olarak azalmayan bir fonksiyon, monoton olarak artabilir ya da sabit kalabilir. Monoton artma, bireylerin ölçülen özelliğe ilişkin yetenek düzeylerinin artmasıyla beraber maddeye doğru yanıt verme olasılıklarının da artması ya da aynı kalması anlamına gelmektedir. Bir birey yetenek, beceri, tutum ya da özellik yönünden başka bir bireyden daha yüksek seviyedeysen, maddeleri olumlu yanıtlama olasılığı da daha yüksektir (Meijer ve diğerleri, 1990).

Eşitlik.2.10’da monoton artma varsayımının matematiksel ifadesine yer verilmiştir:

$$\begin{aligned} \Theta_a &< \Theta_b, \\ P_i(\Theta_a) &\leq P_i(\Theta_b) \end{aligned} \quad (2.10)$$

Eşitlik.2.10’da yer alan matematiksel ifade, a yeteneğindeki bireyin maddeye yanıt verme olasılığının, b yeteneğindeki bireyin maddeye yanıt verme olasılığından küçük olması durumunda madde karakteristik eğrilerinin gizil yeteneğin azalmayan bir fonksiyonu olduğunu ifade etmektedir.

İki parametrelili lojistik modele ait madde karakteristik eğrileri ve Monoton Homojenlik Modeline ait madde karakteristik eğrileri karşılaştırıldığında (Sijtsma ve Meijer, 2007):

- POMTK'ye ait madde karakteristik eğrilerinin, S şeklinde olması hatta düzgün olması gerekmemektedir.
- POMTK'ye ait madde karakteristik eğrileri birden fazla bükülme noktasına sahip ve birkaç Θ aralığında sabit olabilmektedir.
- POMTK'de madde karakteristik eğrileri, 0'dan büyük alt asimptotlara veya minimum değerlere ve 1'den küçük üst asimptotlara veya maksimum değerlere sahip olabilmektedir.

Monotonluk varsayımının sağlanması POMTK modelleri için gerekli olmakla beraber monoton olan MKE, her zaman doğru tanımlamayı yapmayabilir. Bazı durumlarda monoton olmayan MKE ile daha uygun bir tanımlama yapılabilir. Örneğin “Eşine şiddet gösteren bir bireyin, 1 yıl hapis cezası alması yeterlidir” ifadesine katılmayanlar ve ceza süresini yetersiz bulanlar düşük, kararsız kalanlar ise yüksek olasılıkla bu ifadeyi destekleyebilir. Böyle bir durumda, MKE'nin çan şeklinde olması yanıt olasılığını tanımlamak için daha uygun olabilmektedir (Sijtsma ve Molenaar, 2002).

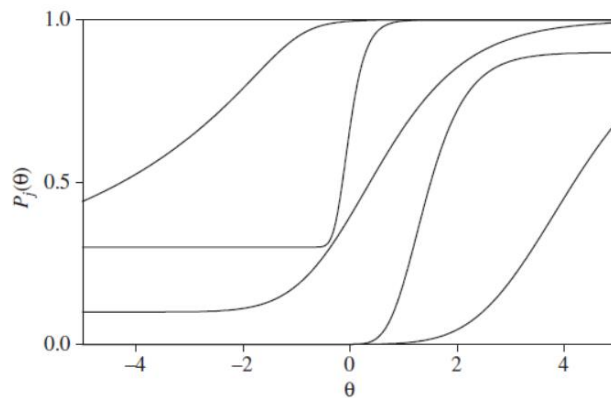
Gizil monoton artma varsayımı, R paket programında bulunan “mokken” paketi kullanılarak sınanabilmektedir. Mokken paketinde “madde kalan puanı yöntemiyle” varsayım ihlal sayısı ve bu ihlallerin istatistiksel olarak anlamlı olup olmadığı incelenmektedir. Boyutluluğun test edilmesi sırasında H_j (maddelerin her biri için hesaplanan ölçeklenebilirlik katsayısı) için seçilen alt sınır (c) araştırmacının isteğine bağlı olarak seçilebilmekle beraber varsayılan değer 0.3'tür. Madde ölçeklenebilirlik katsayılarının en az 0.3 olarak seçilmesi sonucunda, madde karakteristik eğrileri genellikle monoton olma eğiliminde olur. H_j değeri için 0.3'ten küçük bir alt sınır belirlemek, gizil monoton artma varsayımını sağlamayan maddelerin seçilmesiyle sonuçlanabilmektedir. Bu değer (c değeri) tek başına varsayımın karşılanması için bir ölçüt olmamakla beraber 0.3 değeri altında olan maddeler ölçeğin dışında tutulmaktadır (Wismeijer ve diğerleri, 2008).

Değişmez Madde Sıralaması (DMS) (Nonintersecting IRFs): Tek boyutluluk, yerel bağımsızlık ve monotonluk varsayımlarının test edilmesi, bireylerin sıralandığı ölçmeler ve birçok POMTK uygulaması için yeterlidir fakat maddelerin zorluk düzeylerine göre sıralanması durumu söz konusu olduğunda “değişmez madde sıralaması” varsayımı da test edilmelidir. Değişmez madde sıralaması varsayımı, “kesişmeyen madde karakteristik eğrileri” varsayımı olarak da adlandırılmaktadır. Monoton homojenlik modelinden farklı olarak, ikili monotonluk modelinde bu varsayımın karşılanması gerekmektedir.

Değişmez madde sıralaması, MKE’lerin yetenek düzeyi boyunca kesişmediği anlamına gelmektedir (Sijtsma ve Molenaar, 2002). Başka bir ifadeyle değişmez madde sıralaması varsayımı, “gizil özellik ölçeğinin her bir değeri için aynı olan maddelerin sıralaması” şeklinde tanımlanabilir. Gizil özellik ölçeği, POMTK modellerinin ilgili kuramsal yapı için sağladığı sıralı ölçeği ifade etmektedir (Sijtsma, 2005). Eşitlik.2.11’de değişmez madde sıralaması varsayımının matematiksel ifadesine yer verilmiştir:

$$P_1(\Theta) \leq P_2(\Theta) \leq \dots \leq P_k(\Theta) \quad (2.11)$$

Eşitlik.2.11’de yer alan denklem; MKE’lerin kesişmediğini, maddeler arasından en zor olan maddenin 1.madde olduğunu ve 1.maddeyi 2.maddenin takip ettiğini ve belirli theta değerlerinde zorluk sıralamasının bağlantı olasılığının olduğunu göstermektedir. Şekil.2.5’te monoton artma ve değişmez madde sıralaması varsayımlarının sağlandığı madde karakteristik örneği sunulmuştur (Koğar, 2014):



Şekil.2.5. Monoton artan ve kesişmeyen madde karakteristik eğrileri

Şekil.2.5 incelendiğinde MKE'lerin yetenek düzeyi boyunca kesişmediği görülmektedir. Dolayısıyla değişmez madde sıralaması varsayımının karşılandığı söylenebilir.

Değişmez madde sıralaması (DMS) varsayımının sağlanması, bireylerin ve grupların karşılaştırıldığı durumlarda önem taşımaktadır. Aşağıda DMS'den pratikte faydalanılabilecek alanlara yönelik örnekler sunulmuştur (Sijtsma ve Molenaar, 2002):

- **Psikolojik teorilerin test edilmesi:** MTK modelleri, psikolojik yapılarla ilişkin teorilerin test edilmesinde kullanılabilir. “Gelişim sürekli ve belli aşamalarda gerçekleşir” ilkesine dayanarak geliştirilmiş bir testte, bireylerin içerisinde bulunduğu gelişim dönemine ve geçmiş gelişim dönemlerine ait görevleri yapabiliyor olmaları beklenir (Çelikkaleli, 2014). Dolayısıyla bu gibi testlerde, sabit bir sıralama ve maddelere karşılık gelen gelişim görevleri söz konusudur. Böyle bir testte, DMS varsayımının sağlanması önem taşımaktadır.
- **Başlatma ve durdurma kuralları:** Bilişsel yetenek ya da eğitsel başarı test edilirken, maddeler genellikle kolaydan zora doğru sıralanır. Böylece teste yanıt verenlerin, zor maddeleri cevaplayamadıkları için teste devam etme noktasında motivasyon kaybı yaşama ya da maddeleri yanıtlamayacaklarına dair kaygı yaşama olasılıkları azalır. Motivasyon ve kaygı gibi durumlar bireyi ve bireyin teste verdiği yanıtı etkilemektedir (Yıldırım, 2011). Diğer yandan bir test, bireyin birden fazla yanlış yanıt vermesi durumunda durdurulabilir. Maddelerin kolaydan zora doğru sıralanması, üst üste verilen yanlış yanıtlardan sonra doğru yanıt gelmeyeceği varsayımını beraberinde getirir. Farklı başlama ve bitirme kurallarıyla aynı madde sıralaması kullanılabilir. Wechsler Zekâ testi, başlatma-durdurma kurallarını uygulayan testlere bir örnektir (Özgüven, 2020). Özetle, zekâ testlerinin genellikle tanı koymaya yönelik olması ve madde sıralaması her bireye uygulandığını varsayması nedeniyle bu kuralların “değişmez madde sıralaması” na dayandırılması anlamlı olacaktır.

- **Anormal madde puanı modelleri (Abberant item score patterns):** MTK modeli, ikili puanlanan testlerde (1-0) madde zorlaştığında 1 puan yanıtlarının daha düşük olduğunu öngörmektedir fakat bazı veri setlerinde zor maddelerde çok sayıda doğru yanıt ya da tam tersi kolay maddelerde çok sayıda yanlış yanıtın olduğu görülür. Bu cevaplar kopya, dikkat eksikliği gibi sebeplerden kaynaklanabilmektedir (Turgut ve Baykul, 2019). Dolayısıyla şüpheli yanıtlardır. Diğer yandan teste yanıt veren tüm bireylerin yanıtları şüpheli yanıtlar değildir. MTK'da MKE'ler kesiştiği için böyle bir durum ortaya çıkar fakat İkili Monotonluk Modelinde (İMM), kesişmeyen MKE'ler yani değişmez madde sıralaması vardır. Dolayısıyla İMM ile bireye uygun sonuçların yorumlanması daha kolaydır.
- **Değişen Madde Fonksiyonu (DMF) :** Bir maddeye ait MKE'lerin alt gruplarda farklı olması, iki alt grubun maddeyi doğru yanıtlama olasılığının farklı olduğunu göstermektedir. Örneğin, testte yer alan bir maddeyi doğru yanıtlama olasılığı erkek bireylerde, kadın bireylere göre daha fazla olabilir. Bu durumda aynı madde için ayrı ayrı olasılıklar değerlendirilir (Atalay ve diğerleri, 2012). Test, farklı gruplardan eş zamanlı seçimler yapılması amacıyla kullanılacaksa DMM, bu eşitsizliği ortaya çıkarabilir.

2.3.2. Parametrik Olmayan Madde Tepki Kuramı Modelleri

Monoton Homojenlik Modeli (MHM): POMTK modellerinden Monoton Homojenlik Modeli (MHM), parametrik MTK modelini tahmin etmek için örneklemin yetersiz olması durumunda bile ölçeklendirme, puanlama ve bireye uygun soruların incelenebilmesine yönelik olarak geliştirilmiştir (Sijtsma ve Junker, 2006).

Mokken (1971) ikili puanlanan veriler için bireylerin yeteneklerinin ölçümü ve hem bireylerin hem de maddelerin ölçümü için iki farklı model önermiştir (Sijtsma ve diğerleri, 2011). Bu modellerden ilki olan MHM, POMTK modellerinin en genel olan modelidir ve üç varsayımı bulunmaktadır. Bu varsayımlar: 1) Tek boyutluluk 2) Yerel bağımsızlık ve 3) Gizil monoton artma varsayımlarıdır (Sijtsma ve diğerleri, 2011). MHM modeli, bireylerin sıralandığı ölçmeler için uygun bir MTK modelidir (Sijtsma ve Molenaar, 2002).

Eşitlik.2.12’de MHM modelinin denklemine yer verilmiştir:

$$P(X_+ \geq x_+ | \Theta = \Theta_a) \leq P(X_+ \geq x_+ | \Theta = \Theta_b), \Theta_a < \Theta_b \quad (2.12)$$

Eşitlik.2.12’de yer alan denklemde X_+ toplam puanı, Θ (theta) ise yetenek düzeyini ifade etmektedir. Ayrıca denklem bireylerin theta (yetenek düzeyi) tarafından sıralandığını ifade etmektedir. Theta ve testten beklenen toplam puan doğru orantılıdır (Sijtsma ve Molenaar, 2002).

Meijer ve diğerleri (1990) birçok test uygulamasında bireylerin yeteneklerinin sıralamasını bilmenin yeterli olduğunu, bu nedenle MHM modelinin kullanılmasının uygun olduğunu ifade etmiştir çünkü MHM, sıralı bir ölçümü garanti etmekle beraber diğer modeller kadar kısıtlayıcı değildir.

İkili Monotonluk Modeli (İMM): İkili Monotonluk Modelinde (İMM), MHM’den farklı olarak tek boyutluluk, yerel bağımsızlık ve gizil monoton artma varsayımlarının yanı sıra madde karakteristik eğrilerinin kesişip kesişmediği de incelenmektedir. Diğer bir deyişle, İMM değişmez madde sıralaması varsayımının sağlanmasını da gerektirir (Sijtsma ve diğerleri, 2011).

MTK modellerinden, Rasch modeli kesişmeyen MKE’lere sahip tek model olarak bilinmektedir (Sijtsma ve Molenaar, 2002). İMM, sıralı ve değişmez madde sıralaması özelliğine sahip olması sebebiyle Rasch modelinin parametrik olmayan karşılığı olarak görülmektedir (Van Schuur, 2003).

İMM, MHM’nin özel bir durumu olduğundan İMM tarafından açıklanan her veri setinin MHM tarafından da açıklanabildiği söylenebilir fakat MHM tarafından açıklanan her veri seti, İMM tarafından açıklanamayabilir (Sijtsma ve Molenaar, 2002).

2.3.3. Parametrik Olmayan Madde Tepki Kuramı Modellerinde Geçerlik ve Güvenirlilik İncelemeleri

POMTK kapsamında geliştirilen H_i , H_{ij} ve H ölçeklenebilirlik katsayıları madde ayırt ediciliği/geçerliği ve testin güvenirliliği hakkında bilgi vermektedir. POMTK modellerine dayalı analizlerde H_i katsayısından faydalanılarak maddenin ayırt ediciliği değerlendirilir. Teoride negatif ve pozitif değerler alabilen H_i katsayısı, MHM’de 0-1 değerleri arasında değerler alabilmektedir. Negatif H değeri, örnekleme ilgili bir problem ya da MHM’nin ihlal edilmesi sonucunda ortaya çıkmaktadır.

Bir madde için hesaplanan H_i katsayısı 0.00-0.30 değerleri arasında yer alıyorsa, o maddenin zayıf ayırt etme gücüne sahip olduğu ve bireyleri sıralamak için yeterli bilgiyi vermeye neredeyse katkıda bulunmadığı söylenir. Dolayısıyla bu maddelerin teste alınmaması gerektiği savunulur (Sijtsma ve Molenaar, 2002).

H_{ij} katsayısı, bir maddenin diğer maddelerden hangisiyle daha iyi uyum gösterdiğini ve hangi maddelerin bireyleri doğru şekilde sıralamaya en fazla katkıda bulunduğunu belirlemeyi sağlayan katsayıdır. H_i ile benzer şekilde yorumlanmaktadır.

H katsayısı ise H_i katsayılarının ağırlıklı ortalaması olmakla birlikte H_i katsayılarının verdiği bilgilerin bir özetini sunmaktadır. H katsayısı, testin tamamı için bireylerin ne derece sıralanabileceğine dair kanıt sağlayan bir katsayıdır. H katsayısı da H_i katsayısı gibi yorumlanmakla beraber H katsayısının en küçük H_i katsayısından büyük olduğu göz önünde bulundurulduğunda $H \geq 0.3$ 'tür ve 0.30-1.00 aralığında değişmektedir (Sijtsma ve Molenaar, 2002).

Molenaar (2001) H katsayısını, Mokken analizinin temel yapısı olarak tanımlamıştır. H katsayısı, testin güvenilirliği hakkında da bilgi vermektedir. Gutmann hatasına dayalı olan H katsayısının bire eşit olması, "Gutmann hatasının hiç yapılmadığını" göstermektedir (Sijtsma ve Molenaar, 2002). Bazı araştırmacılar H katsayısını varsayım sınamasında değerlendirmektedir. H katsayısı, değişmez madde sıralamasının değerlendirilmesinde kullanılmamalıdır. Öncelikle tüm maddelerin MKE'lerine veya madde adım yanıt fonksiyonlarına bakılarak monotonluğu değerlendirilmelidir. H katsayısının pozitif olması sonuca varmak için tek başına yeterli olmamakla beraber MKE'lerin monoton olarak artıp artmadığı ayrı ayrı değerlendirilmeli ve karşılanmalıdır (Sijtsma ve diğerleri, 2011).

Parametrik MTK'de bilgi fonksiyonu, KTK'deki güvenilirlik katsayısıyla benzer bir role sahiptir. KTK'deki güvenilirlik katsayıları genel bir değerlendirme yapılmasına imkân sağlarken MTK'deki bilgi fonksiyonu, daha derin bir inceleme fırsatı sunmaktadır. POMTK modellerinde güvenilirlik incelemeleri MTK'den farklı olarak bilgi fonksiyonlarıyla değerlendirilmemektedir. POMTK modelleriyle, bilgi fonksiyonunu tahmin etmek için gerekli olan kişi ve madde parametrelerinin sayısal kestirimleri yapılamamaktadır. Bu nedenle güvenilirlik, KTK'den elde edilen güvenilirlik katsayısıyla belirlenmektedir (Meijer ve diğerleri, 1990).

POMTK kapsamında yapılan güvenilirlik incelemelerinde MS (Molenaar-Sijtsma) katsayısından ve LCRC (Latent Class Reliability Coefficient) katsayısından da faydalanılmaktadır. Van der Ark ve arkadaşlarının (2011) gizil sınıf analizi amacıyla geliştirdiği LCRC katsayısından POMTK yaklaşımı güvenilirlik incelemelerinde de faydalanılabilmektedir (Akt. Mor-Dirlik, 2017). POMTK kapsamında yapılan güvenilirlik incelemeleri için R programında yer alan “mokken” paketinden faydalanılmaktadır.

2.4.İlgili Araştırmalar

Bu bölümde POMTK ile ilgili ulusal ve uluslararası alanyazında bulunan araştırmalara yer verilmiştir. POMTK’nin gelişmekte olan bir alan olması sebebiyle uluslararası alanyazında daha çok POMTK’nin tanıtıldığı, varsayımlarının ve kuramsal temellerinin sınıandığı teorik araştırmalar yer almaktadır. Bunun yanı sıra daha az sayıda olmasına rağmen uygulamalı araştırmalar da uluslararası alanyazında mevcuttur. Ulusal alanyazında ise POMTK ile ilgili sınırlı sayıda araştırma olduğu görülmüştür. Bu sebeple ulusal alanyazın bölümünde POMTK ile ilgili yapılan araştırmaların tümüne yer verilmiştir.

Uluslararası alanyazın incelendiğinde, POMTK’yle ilgili teorik çalışmaların dışında, araştırmacıların iki kuramı (KTK’yi ve parametrik MTK’yi ya da POMTK ve MTK’yi) çeşitli değişkenler açısından karşılaştırdıkları görülmektedir.

Fan (1998) araştırmasında, gerçek verilerle MTK ve KTK’ye dayalı olarak kestirilen madde istatistiklerini incelemiştir. Araştırmanın bulguları iki yaklaşımdan elde edilen madde istatistiklerinin karşılaştırılabilir olduğunu göstermiştir. KTK ve MTK’ye dayalı yetenek tahminleri arasındaki ortalama korelasyonun tüm koşullar için 0.96’dan büyük olduğu ve zorluk tahminleri arasındaki ilişkilerin çoğu koşulda 0.90’ın üzerinde olduğu bulgusu elde edilmiştir.

McDonald ve Paunonen (2002) araştırmalarında Monte Carlo simülasyonu kullanarak KTK ve MTK’ye dayalı olarak kestirilen madde parametrelerini karşılaştırmıştır. Araştırmadan elde edilen bulgular, iki kuramdan elde edilen madde güçlük parametresi hesaplamaları arasında anlamlı ve yüksek ilişki olduğunu göstermiştir fakat KTK’ye dayalı olarak kestirilen madde ayırt edicilik hesaplamalarının, yalnızca belli simülasyon koşulları altında MTK’ye dayalı olarak kestirilen madde ayırt edicilik hesaplamalarıyla benzer olduğu görülmüştür.

Meijer ve Baneke (2004) arařtırmalarında Minnesota Çok Yönlü Kişilik Envanteri-2'nin Depresyon ölçeğinden elde edilen verilerle POMTK'nin psikoloji ölçeklerinde uygulanabilirliğini tartışmış ve POMTK modellerini, MTK modelleriyle karşılaştırmıştır. Elde edilen bulgular doğrultusunda, araştırmacılar psikolojik ölçeklerden elde edilen verilerin araştırılmasında MTK modellerinden önce POMTK modellerinin kullanılmasını önermişlerdir.

Courville (2005) araştırmasında çoktan seçmeli bir testten elde edilen verilerle KTK ve MTK'ye dayalı olarak kestirilen madde parametrelerini karşılaştırmış, iki yaklaşımın da benzer madde istatistikleri ürettiğini bulgulamıştır. Progar ve diğerleri (2008) araştırmalarında TIMSS 1995 verilerini kullanarak KTK ve MTK'ye dayalı olarak kestirilen madde parametrelerini karşılaştırmıştır. Araştırmadan elde edilen bulgular, iki yaklaşıma dayalı olarak kestirilen madde parametrelerinin karşılaştırılabilir ve benzer olduğunu göstermiştir.

Delevar ve diğerleri (2006) araştırmasında KTK ve POMTK yaklaşımlarından elde edilen madde parametrelerini katılımcıların özellikleri açısından karşılaştırmıştır. İki farklı grubun bulunduğu araştırmadan elde edilen bulgular, KTK yaklaşımına göre kestirilen madde istatistiklerinin gruplara göre değişirken POMTK yaklaşımına göre kestirilen madde istatistiklerinin gruplara göre değişmediğini göstermiştir.

Dyehouse (2009) beşli likert olarak puanlanan ölçekten elde edilen gerçek verilerle yaptığı araştırmasında POMTK ve MTK modellerini karşılaştırmış, Mokken modellerinden MHM modelinin birçok maddeyle uyumlu olduğunu ancak daha katı olan İMM modelinin her maddeyle uyumlu olmadığını bulgulamıştır. Zhao (2008) de araştırmasında gerçek verileri kullanarak POMTK ve MTK modellerini karşılaştırmıştır. Araştırmadan elde edilen bulgular, MTK modellerinden elde edilen sonuçların daha tutarlı olduğunu ancak POMTK modellerinden elde edilen sonuçların ölçek ve yapı ilişkisine dair daha fazla bilgi verdiğini göstermiştir.

Ulusal alanyazın incelendiğinde, araştırmacıların çoğunun parametrik MTK ve parametrik olmayan MTK arasında örneklem büyüklüğü, test uzunluğu, boyut sayısı, madde ve yetenek kestirimleri, kestirim hataları, parametreler, model uyumu gibi değişkenler açısından karşılaştırmalar yaptığı görülmektedir.

Koğar (2015) araştırmasında madde parametrelerini ve model veri uyumlarını simülasyon koşullarında örneklem büyüklüğü, test uzunluğu gibi değişkenler açısından MTK modellerini karşılaştırmıştır. Bulgular; POMTK yaklaşımının her simülasyon koşulunda model veri uyumunu düşük hatayla hesapladığını, örneklem büyüklüğünün artmasıyla beraber H_i ve p parametrelerine ait standart hata değerlerinin azaldığını göstermiştir.

Bıkmaz-Bilgen ve Doğan (2017), çok kategorili maddeler için MTK ve POMTK modellerini kullanarak kestirimler yapmış ve ilgili modelleri örneklem büyüklüğü, örneklem dağılımı, madde sayısı ve madde yanıt kategori sayısı değişkenleri açısından karşılaştırmıştır. Araştırmadan elde edilen bulgular, POMTK modelinde model veri uyumu hesaplamasının daha kolay olmasının MTK modeline göre avantajlı olduğunu göstermiştir. Diğer bulgular ise güvenilirlik hesaplamalarının iki model için benzer olduğunu, POMTK modelinde örneklem sayısının az ve test uzunluğunun daha kısa olduğu durumlarda parametre hesaplamalarının standart hatalarının MTK modeline kıyasla daha düşük olduğunu ve POMTK modelinin farklı koşullarda, standart hatalarının yakın değerler aldığını göstermiştir.

Avşar (2018) araştırmasında yapay veriler kullanılarak çok kategorili maddeler için kategori sayısının psikometrik özellikleri nasıl etkilediğini POMTK modeliyle belirlemeyi amaçlamıştır. Araştırmadan elde edilen bulgular, kategori sayısı ve model-veri uyumu arasında ilişki olmadığını ve farklı test uzunluklarında kategori sayısının güvenilirlik kestirimine etkisi olmadığını göstermiştir.

Mor-Dirlik ve Koç (2019) ise TIMMS 2011 verilerini kullanarak yaptıkları araştırmada, test uzunluğu ve örneklem büyüklüğü açısından POMTK ve MTK arasında karşılaştırma yapmış, farklı koşullarda hesaplanan madde ve yetenek parametrelerinin, iki yaklaşımda benzer sonuçlar elde edildiğini bulgulamıştır.

Koğar (2019), Belirti Tarama Listesi'nin psikometrik özelliklerini Mokken ölçekleme analizleriyle değerlendirmiş ve ölçeğin yüksek geçerlik ve güvenilirliğe sahip olduğunu bulgulamıştır. Kıbrıslıoğlu-Uysal (2020) araştırmasında MTK ve POMTK modellerini gerçek ve yapay veriler kullanarak farklı örtük özellik dağılımları, madde güçlük dağılımları, madde sayısı ve örneklem büyüklüğü değişkenleri açısından karşılaştırmıştır.

Elde edilen bulgular; b parametresinin yetenek kestirimlerine etkisinin olmadığını, örtük özellik normal dağılmadığında madde parametre kestirimlerini etkilediğini ve parametrik olmayan Bayesçi kestirim yönteminin en fazla hatalı kestirim yapan yöntem olduğunu, yapay verilerde ve gerçek verilerde benzer sonuçlara ulaşıldığını göstermektedir.

Reyhanlıoğlu ve Doğan (2020), tek boyutlu ve çok boyutlu MTK ve POMTK modellerinde örneklem büyüklüğüne göre kestirilmiş olan parametrelerin değişmezliğini incelemiştir. TEOG (2015) verileri kullanılan araştırmanın bulguları; Tek boyutlu POMTK'nin 200 örneklem büyüklüğünde madde parametresi değişmezliğini sağlarken, MTK için en az 1000 örneklem büyüklüğüne ihtiyaç duyulduğunu göstermiştir.

Mor-Dirlik (2021) araştırmasında, KTK temel alınarak kestirilen ve parametrik olmayan Madde Tepki Kuramı temel alınarak kestirilen madde ayırt edicilik parametrelerini karşılaştırmış ve iki kuram arasındaki benzerlikleri incelemiştir. Araştırmada duyuşsal özellikleri ölçen “Tek Tip Öğrenci Kıyafetine İlişkin Öğretmen Tutum Ölçeği” kullanılarak veriler toplanmış ve çalışma grubunu 226 öğretmen oluşturmuştur. Madde ayırt edicilik hesaplamaları, KTK kapsamında nokta çift serili korelasyon katsayısından ve POMTK kapsamında ölçeklenebilirlik katsayılarından faydalanılarak hesaplanmıştır. Elde edilen bulgular, KTK'ye dayalı hesaplamalar ve POMTK'ye dayalı hesaplamalar arasındaki ilişkinin yüksek ve anlamlı olduğunu göstermiştir. Bulgulardan hareketle KTK ve POMTK'nin madde ayırt edicilik hesaplamalarında birbirlerine alternatif olabileceği sonucuna ulaşılmıştır.

3. YÖNTEM

Bu bölümde araştırmanın modeli, çalışma grubu, veri toplama araçları ve verilerin toplanması ve analizi başlıklarına yer verilmiştir.

3.1.Araştırmanın Modeli

Bu araştırmada, çalışma grubuna uygulanan Eğitimde Ölçme ve Değerlendirme testinden elde edilen verilerin madde ayırt edicilik parametresi ve güvenirlik indeksleri KTK ve POMTK yöntemleri kullanılarak kestirilmiş olup bu iki yöntem karşılaştırılmıştır. Genel amaç doğrultusunda araştırmanın modeli, betimsel araştırma modelidir. Araştırmanın deseni ise korelasyonel araştırmadır. Korelasyonel araştırma, iki ya da ikiden fazla değişken arasındaki ilişkinin belirlenmesi ve neden-sonuç ilişkileriyle ilgili bilgi edinilmesi amaçlandığında kullanılan bir araştırma desendir (Kadıoğlu-Ateş ve Mazı, 2017).

3.2.Çalışma Grubu

Araştırma kapsamında 592 katılımcıya ulaşılmıştır. Katılımcılar, 2022-2023 ve 2023-2024 eğitim-öğretim yılında Bolu Abant İzzet Baysal Üniversitesinde Sınıf Öğretmenliği, Matematik Öğretmenliği, Rehberlik ve Psikolojik Danışmanlık ve Türkçe Öğretmenliği bölümlerinde öğrenim gören 3.sınıf öğrencileridir.

Araştırma örnekleme uygun örnekleme yöntemi ile belirlenmiştir. Uygun örnekleme yöntemi, veri toplama sürecini hızlandıran ve bireylerin kolay erişilebilir olduğu bir yöntem olmakla beraber diğer örneklem yöntemlerinin bazı kısıtlılıklar sebebiyle kullanılamadığı durumlarda kullanılabilir (Kılıç, 2013).

3.3.Verİ Toplama Aracı

Araştırmada veri toplama amacıyla öğretmen yapımı bir test olan “Ölçme ve Değerlendirme” testi sonuçları kullanılmıştır. Prof. Dr. İbrahim Alper Köse tarafından geliştirilen çoktan seçmeli test, 40 maddeden oluşmaktadır. Test 1-0 (doğru-yanlış) olarak puanlanmaktadır. Testin amacı, Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi öğrencilerinin, ölçme ve değerlendirme dersi kazanımlarının değerlendirilmesidir.

3.4.Verilerin Toplanması ve Analizi

Veriler, tez danışmanı tarafından araştırmacıya Excel yoluyla isimsiz bir şekilde aktarılmıştır. Veri toplama sürecinde, Eğitimde Ölçme ve Değerlendirme sınavı öncesi Bolu Abant İzzet Baysal Üniversitesi öğrencileriyle derste bir araya gelinmiş ve katılımcılar, gönüllük ve gizlilik ilkeleri hususlarında yazılı ve sözlü olarak bilgilendirilmiştir. Ayrıca katılımcılara araştırmanın amacı ve özeti hakkında bilgilendirmeler yapılmıştır. Katılımcılardan araştırmaya gönüllü katılım sağladıklarına dair onay alınmıştır.

Veri analizi gerçekleştirilmeden önce katılımcıların cevaplamış olduğu testte uç değerlerin yer alıp almadığı incelenmiştir. Analiz öncesinde puanlar Z puanına dönüştürülmüştür. Z puanı $\pm 3,29$ aralığında kalanlar kabul edilmiştir (Field, 2009). Bu aralık, verilerin 99.9 oranında veri setinde tutulması amaçlanarak seçilmiştir. Araştırmada uç değerlerin tespit edilmesi durumunda başvurulacak üç yöntem olmakla beraber bunlardan bir tanesi ilgili katılımcının araştırmadan çıkarılmasıdır (Field, 2009). Katılımcıların Z puanları incelendiğinde, bu puanların $\pm 3,29$ aralığında kaldığı görülmüş dolayısıyla katılımcıların hiçbirisi araştırmadan çıkarılmamıştır.

Madde istatistikleri kestirilmeden önce maddelerin varyansları incelenmiştir. Tek boyutluluk varsayımını ihlal eden ve varyansları sıfır olan 21 madde araştırma kapsamı dışında bırakılmıştır. Kalan 19 madde ile analizler gerçekleştirilmiştir.

KTİK'ye dayalı olarak kestirilen madde ayırt edicilik indeksleri “çift serili korelasyon” katsayısı ile belirlenmiştir. Değişkenlerden birinin sürekli, diğerinin yapay ikili (1-0) olduğu durumlarda çift serili korelasyon katsayısı madde ayırt edicilik indeksinin hesaplanmasında kullanılabilmektedir (Tarkun-Tavşancıl, 1996). Başarı testinden elde edilen toplam puan sürekli değişkeni, testte yer alan her bir maddeye verilen doğru-yanlış (1-0) cevaplar ise yapay sürekli değişkeni temsil etmektedir. Araştırma kapsamında kullanılan çoktan seçmeli testin değerlendirilmesinde maddeler, yapay ikili değişkendir. Eğitimde ölçme ve değerlendirme testinden elde edilen toplam puan ise sürekli değişkendir. Çift serili korelasyon katsayısı için normallik varsayımı çarpıklık-basıklık katsayılarından faydalanarak test edilmiştir. Çarpıklık-basıklık katsayılarının ± 1 aralığında kalması, normallik varsayımının karşılanmasında kritik değer olarak alınmıştır (Çokluk ve diğerleri, 2021).

KTK'ye dayalı olarak yapılan güvenilirlik incelemelerinde ise Cronbach Alfa katsayısı ve iki yarı güvenilirliği katsayılarından faydalanılmıştır. Cronbach Alfa katsayısı yorumlanırken; 0.5'ten küçük değerler “kabul edilemez”, 0.50-0.60 aralığındaki değerler “zayıf”, 0.60-0.70 aralığındaki değerler “şüpheli”, 0.70-0.80 aralığındaki değerler “kabul edilebilir”, 0.80-0.90 aralığındaki değerler “iyi”, 0.90 ve 0.90'dan büyük değerler “harika” olarak yorumlanmaktadır (George ve Mallery, 2016).

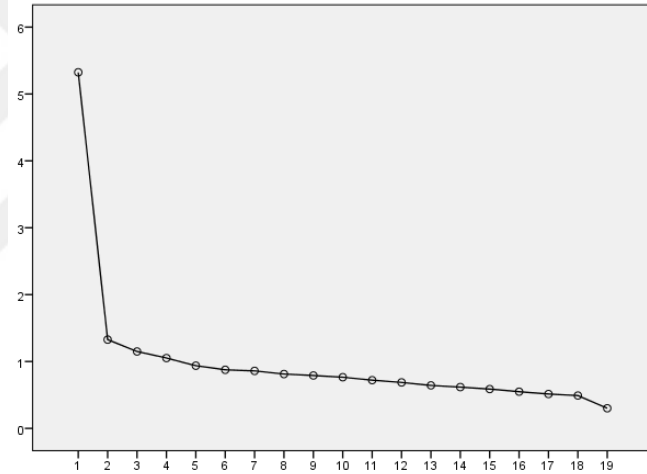
POMTK analizlerine geçilmeden önce veri setinin modele uygunluğu değerlendirilmiştir. POMTK model-veri uyumu değerlendirilirken maddeler arası kovaryansların pozitif olup olmadığı H_{ij} katsayılarından faydalanarak incelenmiştir. Veri setinin madde ve ölçek düzeyinde ölçeklenebilirliği ise H_i ve H katsayılarından faydalanarak incelenmiştir. H_i ve H katsayısı yorumlanırken; 0.30-0.40 aralığındaki değerler zayıf, 0.40-0.50 aralığındaki değerler orta ve 0.50-1.00 aralığında bir değerler güçlü olarak yorumlanmaktadır (Sijtsma ve Molenaar, 2002).

POMTK yöntemine dayanarak kestirilen madde ayırt edicilik indeksi ve güvenilirlik katsayıları belirlenmeden önce ise verilerin hangi POMTK modeline uygun olduğunun belirlenmesi amacıyla “tek boyutluluk”, “yerel bağımsızlık”, “gizil monoton artma” ve “değişmez madde sıralaması” varsayımları incelenmiştir. POMTK varsayımlarının incelenmesinde kullanılan yöntemlere “Parametrik Olmayan Madde Tepki Kuramı Varsayımlarının İncelenmesi” başlığı altında yer verilmiştir. Parametrik olmayan madde tepki kuramı kapsamında madde ayırt edicilik ve güvenilirlik katsayıları ise “Monoton Homojenlik Modeli” ile belirlenmiştir. Güvenirlik, LCRC ve MS katsayılarından; madde ayırt edicilik indeksleri H_i katsayısından faydalanarak incelenmiştir.

KTK ve POMTK madde ayırt edicilik parametresi ve güvenilirliği arasındaki ilişki ise Spearman Brown Korelasyon Katsayısı ile belirlenmiştir. Roscoe (1975), korelasyon katsayısı yorumlanırken 0.30'dan küçük değerlerin düşük düzeyde, 0.30-0.70 aralığında kalan değerlerin orta düzeyde, 0.70'den büyük değerlerin yüksek düzeyde ilişkiyi ifade ettiğini belirtmiştir (Akt. Büyüköztürk ve diğerleri, 2017).

Tek boyutluluk varsayımı OMSİ (Otomatik madde seçim işlemi) yöntemiyle test edilmiştir. Maddeler arası kovaryanların pozitif olup olmadığı değerlendirilmiştir. Ayrıca H ölçeklenebilirlik katsayısından faydalanarak incelemeler yapılmıştır. H katsayısı için önerilen kritik değer 0.3'tür. Fakat bazı araştırmacılar 0.3 değerini katı bulmakta ve farklı değerlerin denenmesini önermektedir (Akt. Mor-Dirlik, 2017; Sijtsma ve Meijer, 2007; Molenaar, 2001; Wismeijer ve diğerleri, 2008). Bu araştırma kapsamında 0.3 ve 0.2 değerleri kritik değer olarak alınmıştır. OMSİ yöntemini kullanmak için R paket programı içerisinde yer alan “mokken” paketinden faydalanılmıştır.

Yerel bağımsızlık varsayımı, AFA yamaç grafiği ile test edilmiştir. Yamaç grafiğine şekil.3.1'de yer verilmiştir. Ayrıca tek boyutluluk varsayımının sağlanması, yerel bağımsızlığın sağlanmasında kriter olarak alınmıştır.



Şekil.3.1 Yamaç grafiği

Faktör analizinden elde edilen yamaç grafiği (şekil 3.1) incelendiğinde, birinci faktörden ikinci faktöre doğru hızlı bir düşüş olduğu, ikinci faktörden itibaren grafiğin yatay bir şekilde devam ettiği ve önemli bir düşüşün olmadığı görülmüştür. Dolayısıyla testin tek bir boyutu ölçtüğü söylenebilir. Yamaç grafiğinde, öz değeri yüksek ve düşük faktörler yer almaktadır. Dolayısıyla grafikte yer alan eğri, keskin bir inişin ardından azalan bir şekle sahiptir. Cattell (1966), faktör seçerken eğrinin bükülme noktasının (eğimin önemli ölçüde değiştiği yer) referans alınmasını gerektiğini savunmaktadır (Akt. Field, 2009).

Gizil monoton artma varsayımı sınanırken, “mokken” paketinde yer alan “madde kalan puanı yöntemi” kullanılmış; varsayım ihlal sayısı ve bu ihlallerin istatistiksel olarak anlamlı olup olmadığı incelenerek monotonluk test edilmiştir. Varsayım ihlallerinin istatistiksel olarak anlamlı olması durumunda bu varsayımın karşılanmadığı söylenebilir.

Boyutluluğun test edilmesi sırasında H_j (maddelerin her biri için hesaplanan ölçeklenebilirlik katsayısı) için seçilen alt sınır (c) araştırmacının isteğine bağlı olarak seçilebilmekle beraber varsayılan değer 0.3’tür. C değeri tek başına varsayımın karşılanması için bir ölçüt olmamakla beraber 0.3 değeri altında olan maddeler ölçeğin dışında tutulmaktadır. Bu sebeple gizil monoton artma varsayımı, veri setinin POMTK modeline uyumluluğunun incelenmesi sırasında araştırılmaktadır (Wismeijer ve diğerleri, 2008).

Tek boyutluluk, yerel bağımsızlık ve monotonluk varsayımlarının test edilmesi, bireylerin sıralandığı ölçmeler ve birçok POMTK uygulaması için yeterlidir (Sijtsma ve Molenaar, 2002). Tek boyutluluk, yerel bağımsızlık ve monotonluk varsayımlarının karşılanması veri setinin “Monoton Homojenlik Modeli” ne uygun olduğunu göstermektedir.

Değişmez madde sıralaması varsayımı test edilirken “mokken” paketinde yer alan bir yöntem olan madde kalan puanı yöntemi kullanılmış, varsayım ihlal sayısı ve bu ihlallerin istatistiksel olarak anlamlı olup olmadığı incelenerek değişmez madde sıralaması varsayımı test edilmiştir. Madde kalan puanı yöntemi, her bir madde çiftinin madde tepki fonksiyonlarını, madde kalan puanı fonksiyonlarıyla karşılaştıran bir yöntemdir.

Kalan grup için hesaplanan yanıtlanma oranının, grubun tamamı için belirlenen yanıtlanma oranıyla uyumlu olmadığı durumlar, ihlal olarak raporlanmaktadır (Meijer ve diğerleri, 2015). Önemli ihlallerin bazen düşük etkiye sahip olması sebebiyle Molenaar ve Sijtsma (2000) ihlallerin etki büyüklüğünü değerlendirmek amacıyla Crit değerlerinin incelenmesini önermektedir. Crit değerinin 40’tan küçük olması ciddi ihlallerin olmadığını, 40-80 aralığındaki değerlerin küçük ihlaller olduğunu ve 80’den büyük değerlerin ciddi ihlaller olduğunu göstermektedir (Akt. Meijer ve diğerleri, 2015).

4. BULGULAR

Bu bölümde araştırmanın genel amacı ve alt amaçları doğrultusunda elde edilen bulgulara yer verilmiştir. KTK'ye dayalı madde ayırt edicilik indekslerinin sunulmasının ardından POMTK ölçeklenebilirlik katsayıları ve varsayım incelemeleri sunulmuştur. Son olarak KTK ve POMTK'ye dayalı olarak kestirilen madde ayırt edicilik parametreleri arasındaki ilişki ve güvenirlik incelemelerine yer verilmiştir.

4.1.Klasik Test Kuramında Dayalı Olarak Kestirilen Madde Ayırt Edicilik İndeksleri

KTK'ye dayalı madde istatistikleri “çift serili korelasyon” katsayısı ile kestirilmeden önce verilerin normal dağılım gösterip göstermediği çarpıklık-basıklık değerleriyle incelenmiştir. Tablo.4.1’de çarpıklık ve basıklık değerlerine ilişkin bulgular yer almaktadır:

Tablo 4.1. Eğitimde Ölçme Değerlendirme Testi Toplam Puanına İlişkin Çarpıklık-Basıklık Değerleri

Puan	N	Çarpıklık Değeri	Basıklık Değeri
Eğitimde Ölçme ve Değerlendirme Testi Toplam Puan	592	-0,673	-0,294

Tablo.4.1’de yer alan çarpıklık-basıklık değerleri incelendiğinde, çarpıklık ve basıklık değerinin ± 1 aralığında olduğu görülmektedir. Diğer bir deyişle veriler normal dağılım göstermektedir.

Normallik varsayımının sağlanmasının ardından çift serili korelasyon katsayısı ile madde ayırt edicilik indeksleri belirlenmiştir.

Tablo.4.2’de çift serili korelasyon katsayısından elde edilen madde ayırt edicilik indekslerine yer verilmiştir:

Tablo 4.2. KTK'ye Dayalı Olarak Kestirilen Madde Ayırt Edicilik İndeksleri

Maddeler	Madde Ayırt Edicilik İndeksi
M2	0.26
M5	0.80
M8	0.67
M11	0.65
M13	0.80
M15	0.37
M18	0.65
M19	0.62
M20	0.90
M23	0.39
M25	0.59
M28	0.50
M30	0.59
M32	0.61
M33	0.45
M34	0.68
M35	0.55
M36	0.55
M37	0.68

Tablo.4.2. incelendiğinde M2'nin ayırt ediciliğinin (0.26) zayıf düzeyde olduğu ve maddenin revize edilmesi gerektiği görülmektedir. M15 ve M23'ün ise iyi ayırt edici (0.37 ve 0.39) bir madde olmasına karşın bazı düzenlemelerle daha ayırt edici bir madde olacağı düşünülmektedir. Diğer 16 maddenin ise iyi ayırt edici maddeler olduğu (≥ 0.40) görülmektedir.

4.2.POMTK Ölçeklenebilirlik Katsayılarının Hesaplanması ve Modelin Belirlenmesi

POMTK analizi gerçekleştirilmeden önce, veri setinin Mokken ölçekleme için uygun olup olmadığı değerlendirilmiştir. Mokken ölçeklemeye uygunluğun değerlendirilmesi amacıyla ilk olarak maddeler arası kovaryansların pozitif olup olmadığı H_{ij} katsayılarından faydalanarak incelenmiştir.

Madde çiftleri için hesaplanan ölçeklenebilirlik katsayıları (H_{ij}) incelendiğinde, tüm maddeler arası kovaryansların pozitif olduğu yani POMTK modelinde ölçeklemeye uygun olduğu görülmüştür. Diğer bir deyişle maddeler aynı özelliği ölçmektedir ve model-veri uyumu sağlanmıştır.

Veri setinin madde ve ölçek düzeyinde ölçeklenebilirliği ise H_i ve H katsayılarından faydalanarak incelenmiştir. Madde ve ölçek düzeyinde hesaplanan ölçeklenebilirlik katsayıları Tablo.4.3'te sunulmuştur.

Tablo 4.3. Madde ve Ölçek Düzeyinde Hesaplanan Ölçeklenebilirlik Katsayıları

Maddeler	H_i	Sh	Z
M2	0.14	0.03	10.49
M5	0.40	0.02	32.65
M8	0.46	0.04	22.30
M11	0.38	0.02	27.33
M13	0.65	0.06	18.73
M15	0.27	0.03	15.65
M18	0.33	0.03	26.12
M19	0.35	0.03	26.30
M20	0.45	0.02	35.26
M23	0.20	0.03	15.80
M25	0.36	0.03	24.97
M28	0.25	0.03	20.52
M30	0.31	0.03	24.66
M32	0.33	0.03	23.54
M33	0.23	0.03	18.70
M34	0.36	0.03	26.16
M35	0.28	0.03	22.77
M36	0.36	0.03	23.28
M37	0.37	0.02	28.30
Ölçek için H	0.32	0.02	71.74

Tablo.4.3'te yer alan H_i katsayıları incelendiğinde, kritik değer olan 0.3'ün altında olan altı madde olduğu (M2, M15, M23, M28, M33 ve M35) görülmüştür. Bu maddelerin Z değerleri incelendiğinde ise bu değerlerin tümünün sıfırdan farklı olduğu yani manidar olarak farklı olduğu saptanmıştır. Kritik değerın altında olmaları ve Z değerlerinin sıfırdan manidar olarak farklı olması sebebiyle M2, M15, M23, M28, M33 ve M35'in düşük ayırt edici maddeler olduğu, POMTK modellerinde ölçeklenemeyeceği söylenebilir. Diğer bir deyişle, bu maddeler bireylerin sıralanması için gereken bilgiye düşük oranda katkı sağlamaları sebebiyle kullanışlı maddeler olmamakla beraber testten çıkarılması gereken maddelerdir (Koğar, 2022).

Diğer yandan ölçek için hesaplanan H değeri incelendiğinde, H değerinin 0.323 olduğu görülmüştür. Dolayısıyla veri setinin zayıf düzeyde Mokken ölçeklemeye uyum sağladığı söylenebilir. Ölçeğin genel olarak Mokken modeline ölçeklemeye uygun olması, maddeler arası kovaryansların pozitif olması ve model-veri uyumunun sağlanması sebebiyle analize madde çıkartılmadan devam edilmiştir.

Tek boyutluluk varsayımının test edilmesinde OMSİ yöntemi kullanılmış olup Tablo.4.4'te veri setinin OMSİ yöntemiyle tek boyutluluğunun incelenmesinden elde edilen bulgulara yer verilmiştir:

Tablo 4.4. Veri Setinin Otomatik Madde Seçim İşlemi Yöntemi İle Tek Boyutluluğunun İncelenmesi

Maddeler	$H \geq 0.3$	$H \geq 0.2$
M2	0*	0
M5	1**	1
M8	1	1
M11	1	1
M13	1	1
M15	0	1
M18	1	1
M19	1	1
M20	1	1
M23	0	1
M25	1	1
M28	0	1
M30	1	1
M32	1	1
M33	0	1
M34	1	1
M35	0	1
M36	1	1
M37	1	1
Uyumsuz Madde Sayısı	6	1

*0 sayısı maddenin kritik değerin (0.3 ve 0.2) altında olduğunu göstermektedir.

**1 sayısı maddenin kritik değere eşit ya da kritik değerin üstünde olduğunu göstermektedir.

Tablo.4.4'te OMSİ için belirlenen iki kritik değere de yer verilmiştir (0.3 ve 0.2). Tablo.4.4 incelendiğinde 0.3 kritik değerine uyumsuz 6 maddenin (M2, M15, M23, M28, M33 ve M35) olduğu görülmektedir.

Diğer bir deyişle kritik değer 0.3 olarak belirlendiğinde 6 maddenin tek boyutluluk varsayımını karşılamadığı söylenebilir. Diğer kritik değer olan 0.2 değerine uyumsuz madde sayısının ise bir (M2) olduğu görülmektedir. M2'nin tek boyutluluk varsayımının sağlanması açısından uygun bir madde olmadığı söylenebilir.

Monoton artma varsayımının test edilmesinde madde kalan puanı yöntemi kullanılmış olup aşağıda yer alan Tablo.4.5'te monotonluğa ilişkin hesaplanan değerler sunulmuştur:

Tablo 4.5. Monoton Artma Varsayımının İncelenmesinde Faydalanan Değerler

Maddeler	H_i	İhlal	Z^*
M2	0.14	3	0
M5	0.40	1	0
M8	0.46	0	0
M11	0.38	0	0
M13	0.65	0	0
M15	0.27	1	0
M18	0.33	0	0
M19	0.35	0	0
M20	0.45	1	0
M23	0.20	2	0
M25	0.36	0	0
M28	0.25	1	0
M30	0.31	0	0
M32	0.33	0	0
M33	0.23	0	0
M34	0.36	0	0
M35	0.28	0	0
M36	0.36	0	0
M37	0.37	1	0

* Varsayım ihlallerinin istatiksel olarak anlamlı olup olmadığını ifade etmektedir.

Tablo.4.5'te yer alan ihlal sütunu incelendiğinde M2, M5, M15, M20, M23, M28 ve M37'de varsayım ihlalleri olduğu görülmektedir. İhlallerin olduğu maddelere ait Z değerleri incelendiğinde, bu değerlerin sıfıra eşit olduğu yani ihlallerin istatiksel olarak anlamlı olmadığı görülmektedir.

Gizil monoton artma varsayımına dair bazı ihlaller bulunmakla beraber bu ihlaller istatiksel olarak anlamlı değildir. Dolayısıyla gizil monoton artma varsayımının tüm maddeler için karşılandığı söylenebilir.

Bir diğer varsayım olan değişmez madde sıralaması varsayımı madde kalan puanı yöntemiyle incelenmiştir. Tablo.4.6'da değişmez madde sıralaması varsayımının incelenmesinde faydalanan değerlere yer verilmiştir.

Tablo 4.6. Madde Kalan Puanı Grubu Yöntemiyle Yapılan Değişmez Madde Sıralaması İncelemesi

Maddeler	H _i	İhlal	Z*	Crit
M2	0.14	28	16	189
M5	0.40	19	10	140
M8	0.46	0	0	0
M11	0.38	5	2	56
M13	0.65	0	0	0
M15	0.27	3	0	29
M18	0.33	18	8	132
M19	0.35	6	1	50
M20	0.45	22	16	176
M23	0.20	20	9	129
M25	0.36	4	0	25
M28	0.25	13	5	118
M30	0.31	12	4	91
M32	0.33	9	4	89
M33	0.23	22	10	157
M34	0.36	9	3	80
M35	0.28	18	6	108
M36	0.36	1	0	16
M37	0.37	7	2	60

*Varsayım ihlallerinin istatistiksel olarak anlamlı olup olmadığını ifade etmektedir.

Tablo.4.6’da yer alan değerler incelendiğinde, iki madde hariç (M13 ve M8) kalan 17 maddede ihlaller olduğu görülmektedir. M2’de 28, M5’te 19, M11’de 5, M15’de 3, M18’te 18, M19’da 6, M20’de 22, M23’te 20, M25’te 4, M28’de 13, M30’da 12, M32’de 9, M33’te 22, M34’te 9, M35’te 18, M36’da 1 ve M37’de 7 ihlal gerçekleştiği görülmektedir. Toplam ihlal sayısı 216’dır.

Bu ihlallerin istatistiksel olarak anlamlı olup olmadığı incelendiğinde ise 14 maddede (M2, M5, M11, M18, M19, M20, M23, M28, M30, M32, M33, M34, M35 ve M37) istatistiksel olarak anlamlı ihlaller olduğu ($Z > 0$) görülmektedir.

Crit değerleri incelendiğinde ise M11, M19 ve M37’deki ihlallerin küçük ihlaller olduğu (0.40-0.80) fakat diğer 11 maddede ciddi ihlaller olduğu (> 0.80) görülmektedir.

Özetle; M8, M13, M15, M25 ve M36’da değişmez madde sıralaması varsayımının karşılandığı söylenebilir. Fakat kalan 14 maddede istatistiksel olarak anlamlı ihlaller söz konusudur. Toplam 216 ihlalin, 96’sı istatistiksel olarak anlamlıdır ve bu ihlallerin 78’inde ciddi ihlaller olduğu görülmüştür.

Dolayısıyla değişmez madde sıralaması varsayımına dair önemli ihlallerin olduğu ve bu varsayımın veri setinin genelinde karşılanamadığı söylenebilir.

Sonuç olarak değişmez madde sıralaması varsayımının sağlanamaması sebebiyle POMTK modellerinden MHM modelinin veri setine uygun olduğu görülmüştür.

4.3.POMTK ve KTK Kuramına Dayalı Olarak Kestirilen Madde Ayırt Edicilik İndeksleri Arasındaki İlişkinin Belirlenmesi

POMTK ve KTK kuramına dayalı olarak kestirilen madde ayırt edicilik indeksleri arasındaki ilişki incelemeden önce Tablo.4.7’de POMTK ve KTK yöntemleriyle kestirilen madde ayırt edicilik indeksleri sunulmuştur:

Tablo 4.7. POMTK ve KTK Madde Ayırt Edicilik İndeksleri

Kullanılan Kuramsal Yaklaşım	POMTK	KTK
Maddeler	H_i	r
M2	0.14	0.26
M5	0.40	0.80
M8	0.46	0.67
M11	0.38	0.65
M13	0.65	0.80
M15	0.27	0.37
M18	0.33	0.65
M19	0.35	0.62
M20	0.45	0.90
M23	0.20	0.39
M25	0.36	0.59
M28	0.25	0.50
M30	0.31	0.59
M32	0.33	0.61
M33	0.23	0.45
M34	0.36	0.68
M35	0.28	0.55
M36	0.36	0.55
M37	0.37	0.68

Tablo.4.7 incelendiğinde, iki yaklaşımda da en düşük ayırt ediciliğe sahip maddenin M2 olduğu görülmüştür.

KTK'ye dayalı olarak kestirilen değerler incelendiğinde, en yüksek ayırt ediciliğe sahip maddelerden M13 ve M20'nin, POMTK'ye dayalı olarak kestirilen değerlerde de en yüksek ayırt ediciliğe sahip maddeler olduğu görülmektedir.

POMTK yaklaşımında kritik değer olan 0.30'un altında ayırt edicilik değerine sahip maddeler KTK yaklaşımına göre incelendiğinde, bu maddelerden M2'nin KTK'de da düşük ayırt edici bir madde olduğu; M15 ve M23'ün iyi bir madde olmasına karşın bazı değişiklikler yapılması gerektiği; M28, M33 ve M35'ün ise iyi ayırt edici bir madde olduğu görülmüştür.

Tablo.4.7'de benzer değerler de farklı değerler de yer almaktadır. KTK ve POMTK'ye dayalı olarak kestirilen madde ayırt edicilik değerleri arasındaki benzerliklerin daha sağlıklı değerlendirilmesi amacıyla Spearman Brown Sıra Farkları Korelasyon Katsayısı ile hesaplama yapılmıştır. Tablo.4.8'de Spearman Brown Sıra Farkları Korelasyon sonuçları sunulmuştur:

Tablo 4.8. KTK ve POMTK Madde Ayırt Ediciliği Arasındaki Benzerliklerin Spearman Brown Sıra Farkları Korelasyon Katsayısından Faydalananarak İncelenmesi

	POMTK	KTK
POMTK	1,000	0,893**
KTK	0,893	1,000

**p<0,05

Tablo.4.8 incelendiğinde KTK ve POMTK yaklaşımlarından elde edilen madde ayırt edicilik değerlerinin pozitif yönde, yüksek ve manidar bir ilişkisi olduğu görülmektedir. İki yaklaşımdan elde edilen sonuçlar, benzerlik göstermektedir.

Son olarak iki yaklaşıma dayalı olarak güvenirlik incelemeleri yapılmıştır. POMTK'ye dayalı olarak güvenirlik incelemeleri; MS ve LCRC katsayılarından, KTK'ye dayalı olarak Cronbach Alfa ve iki yarı güvenirliği katsayılarından faydalananarak yapılmıştır. Tablo.4.9'da KTK ve POMTK yaklaşımlarına göre yapılan güvenirlik incelemelerine yer verilmiştir.

Tablo 4.9. Güvenirlik İncelemeleri

POMTK		KTK	
MS Katsayısı	0.85	Cronbach Alfa Katsayısı	0.85
LCRC Katsayısı	0.88	İki Yarı Güvenirliği	0.76

Tablo.4.9 incelendiğinde, tüm güvenirlik hesaplamalarından elde edilen değerlerin 0.75'ten büyük olduğu görülmektedir. Dolayısıyla testin güvenirliği, her iki kurama göre hesaplanan güvenirlik katsayılarına göre yüksek ve benzerdir.



5. TARTIŞMA

Bu araştırmada KTK'ye dayalı olarak geliştirilen, ikili olarak puanlanan ve başarı testi olan Eğitimde Ölçme ve Değerlendirme testindeki maddelere ait madde ayırt edicilik parametresi KTK ve POMTK yöntemleri kullanarak kestirilmiş ve iki yöntemden elde edilen değerler karşılaştırılmıştır. Ayrıca iki kurama dayalı olarak güvenirlik incelemeleri yapılmış ve farklı güvenirlik katsayıları karşılaştırılmıştır.

Araştırma sonucunda ulaşılan sonuçlar aşağıda verilmiştir;

- i) Araştırma kapsamında Eğitimde Ölçme ve Değerlendirme testindeki maddelerden 21 tanesi varyanslarının sıfır olması ve tek boyutluluk varsayımını ihlal eden sonuçlara yol açması sebebiyle araştırmanın dışında bırakılmıştır. POMTK ve KTK kapsamında yapılan madde analizlerinde M2, M15, M23, M28, M33 ve M35'in POMTK modellerinde ölçeklenemeyeceği ve testten çıkarılması gerektiği; M2, M15 ve M23'ün ise KTK kapsamında değerlendirildiğinde revize edilmesi gerekten maddeler olduğu görülmüştür. Dolayısıyla iki yaklaşıma göre de revize edilmesi gereken maddelerin benzer maddeler olduğu görülmüştür.
- ii) POMTK ve KTK yaklaşımlarına dayalı olarak kestirilen madde geçerlik indeksleri, iki yaklaşıma göre de en düşük ayırt ediciliğe sahip maddenin aynı madde olduğunu göstermektedir.
- iii) KTK'ye ve POMTK'ye dayalı olarak kestirilen madde geçerlik indeksleri, en yüksek ayırt ediciliğe sahip maddelerin iki yaklaşıma göre de aynı maddeler olduğunu göstermektedir.
- iv) KTK ve POMTK yaklaşımlarından elde edilen madde ayırt edicilik indeksleri arasındaki ilişkinin pozitif yönde, yüksek ve manidar bir ilişki olduğu saptanmıştır. Dolayısıyla iki yaklaşımdan elde edilen sonuçların benzerlik gösterdiği bulunmuştur.
- v) POMTK ve KTK'ye dayalı olarak yapılan güvenirlik incelemelerinde iki yaklaşımdan elde edilen değerlerin 0.75'ten büyük olduğu görülmüştür. Dolayısıyla testin güvenirliği, her iki kurama göre hesaplanan güvenirlik katsayılarına göre yüksek ve benzerdir.

Araştırmadan elde edilen bulgular alanyazınla benzerlik göstermektedir. Farklı yaklaşımlara dayalı olarak kestirilen madde istatistikleri arasında manidar ilişkiler olduğu daha önce yapılan araştırmalarda saptanmıştır (Fan, 1998; Courville, 2005; Progar ve diğerleri, 2008; McDonald ve Paunonen, 2002; Bıkmaz-Bilgen ve Doğan, 2017; Mor-Dirlik ve Koç, 2019).

Alanyazında yer alan POMTK ve KTK madde ayırt ediciliklerini karşılaştıran araştırmanın bulguları da araştırmanın bulgularıyla paralellik göstermektedir (Mor-Dirlik, 2021). Her iki araştırmada da POMTK ve KTK yaklaşımlarına dayalı olarak kestirilen madde ayırt edicilik indeksleri arasında yüksek ve manidar bir ilişki bulunmuştur. Güvenirlik incelemelerinden elde edilen bulgular da Mor-Dirlik'in (2021) araştırmasıyla paralellik göstermektedir. İki araştırmada da güvenirliliğin incelenmesinde faydalanan katsayıların, iki yaklaşımda benzer sonuçlar verdiği görülmüştür.

Bu araştırmada KTK kuramına göre geliştirilen bir ölçme aracı kullanılmıştır. Araştırma kapsamında kullanılan testin geliştirilme aşamasında büyük madde havuzları kullanılmamıştır. Maddelere ait betimsel istatistikler, bazı maddelerin varyanslarının sıfır olduğunu göstermiştir. Bu maddeler de analiz dışında bırakılmıştır. Dolayısıyla ilgili maddeler iki yaklaşıma göre karşılaştırılamamıştır. Bu durum araştırmanın zayıf yönünü oluşturmaktadır.

Diğer yandan araştırma kapsamında “POMTK yaklaşımı, KTK yaklaşımına alternatif olarak kullanılabilir mi?” sorusuna yanıt aranmıştır. Elde edilen bulgular, iki yaklaşımdan benzer sonuçlar elde edildiğini göstermiştir. Dolayısıyla sıralama amaçlı ölçmelerde, kısa testlerde, ölçek geliştirme çalışmalarında, küçük örneklemelerde POMTK yaklaşımı kullanılabilir. POMTK yaklaşımının, KTK yaklaşımına göre daha veri odaklı ve keşfedici olması, gruba ve testte yer alan maddelere bağımlı olmaması gibi avantajları göz önüne alındığında verileri daha derinden incelemek amacıyla POMTK, KTK'den daha çok tercih edilen bir yaklaşım olabilir.

Sijtsma ve Molenaar (2002), POMTK yaklaşımlarının çoğunluk tarafından uygulanabilir olduğunu; Molenaar (2001), madde sayısı az olduğunda POMTK yöntemlerinin kullanılabileceğini ifade etmiştir. Bu araştırma, POMTK yaklaşımının kolaylıkla uygulanabileceğini göstermiştir. Aynı zamanda bu çalışmada, az sayıda madde kullanılarak analizler gerçekleştirilmiştir. POMTK modellerinden MHM'nin, az sayıda maddeden oluşan testlerde model-veri uyumunun sağlandığı bulgusu, alanyazınla paralellik göstermektedir (Avşar, 2018; Bıkmaz-Bilgen ve Doğan, 2017; Koğar, 2015).

Eğitimde ölçme doğrudan yapılamadığından, ölçme araçlarının geçerli ve güvenilir olması büyük önem taşımaktadır. Okullarda kullanılan öğretmen yapımı testler ile öğrencilerin başarıları değerlendirilmektedir. Okul sınavlarından alınan notlar, öğrencilerin gelecek eğitim hayatını da etkilemektedir. Dolayısıyla okullarda kullanılan ölçme araçlarının daha derinden incelenmesi ve daha nitelikli ölçme araçlarının geliştirilmesinin eğitime katkısı olacağı düşünülmektedir.

6. SONUÇ VE ÖNERİLER

Araştırmadan elde edilen bulgular ve ilgili alanyazından elde edilen bilgiler doğrultusunda araştırmacılara uygulamaya ve araştırmaya yönelik bazı önerilerde bulunulmuştur:

- Madde sayısı az olduğu durumlarda POMTK yaklaşımı kullanılabilir.
- POMTK yaklaşımına dayalı olarak geliştirilen bilişsel bir test kullanarak iki yaklaşıma göre kestirilen madde ayırt edicilik parametresi karşılaştırılabilir.
- KTK ve POMTK yaklaşımlarına dayalı olarak kestirilen madde geçerlik parametresi, geniş ölçekli bir testten elde edilen verilerle karşılaştırılabilir.
- Araştırma farklı öğretmen yapımı testlerle tekrarlanabilir.
- KTK ve POMTK yaklaşımlarından elde edilen madde ayırt edicilik indeksleri, farklı test uzunlukları ve farklı örneklem büyüklükleri değişken olarak belirlenerek karşılaştırılabilir.
- KTK ve POMTK yaklaşımlarına göre test geliştirme süreçleri karşılaştırılabilir.
- KTK, POMTK ve MTK yaklaşımlarından elde edilen geçerlik indeksleri ve güvenirlik katsayıları karşılaştırılabilir.

7. KAYNAKLAR

Bu tez çalışmasında APA 7th Edition atıf sistemi kullanılmıştır.

- Atalay, K., Gök, B., Kelecioğlu, H., ve Arsan, N. (2012). Değişen madde fonksiyonunun belirlenmesinde kullanılan farklı yöntemlerin karşılaştırılması: Bir simülasyon çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 43(43), 270-281.
- Avşar, A. Ş. (2018). Kategori sayısının psikometrik özellikler üzerine etkisinin mokken homojenlik modeline göre incelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 9(1), 49-63. <https://doi.org/10.21031/epod.357160>
- Baker, F. B. (2001). *The basics of Item Response Theory* (2.baskı). United States of America: ERIC Clearinghouse on Assessment and Evaluation.
- Baykul, Y. (2021). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması* (5.baskı). Ankara: Pegem Akademi Yayıncılık.
- Bıkmaz- Bilgen, Ö., ve Doğan, N. (2017). Çok kategorili parametrik ve parametrik olmayan madde tepki kuramı modellerinin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(4), 354-372. <https://doi.org/10.21031/epod.346650>
- Bulut, G. (2018). Açık ve uzaktan öğrenmede şans başarısı: Klasik Test Kuramı (KTK) ve Madde Tepki Kurama (MTK) temelinde karşılaştırmalı bir analiz. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 4(1), 78-93.
- Büyükoztürk, Ş., Çokluk-Bökeoğlu, Ö. ve Köklü, N. (2017). *Sosyal bilimler için istatistik* (19.baskı). Ankara: Pegem Akademi Yayıncılık.
- Courville, T. G. (2005). *An empirical comparison of item response theory and classical test theory item /person statistics* [Doktora Tezi]. Texas Üniversitesi. <https://core.ac.uk/download/pdf/147123147.pdf>
- Crocker, L., ve Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Harcourt Brace Jovanovich.
- Çelen, Ü., ve Aybek, E. C. (2013). Öğrenci başarısının öğretmen yapımı bir testle klasik test kuramı ve madde tepki kuramı yöntemleriyle elde edilen puanlara göre karşılaştırılması. *Journal of Measurement and Evaluation in Education and Psychology*, 4(2), 64-75. <https://doi.org/10.21031/epod.77213>
- Çelikkaleli, Ö. (2014). Bilim, eğitim psikolojisi ve gelişimle ilgili kavramlar. A.R. Çeçen Eroğul ve F. Yurtal (Ed.), *Eğitim psikolojisi el kitabı* (s. 3-44) içinde. Mentis.
- Çokluk, Ö., Şekercioğlu, G. ve Büyükoztürk, Ş. (2021). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve Lisrel uygulamaları* (3.baskı). Ankara: Pegem Akademi Yayıncılık.
- DeMars, C. (2010). *Item response theory*. Oxford University Press.
- Delavar, A., Moghaddamzadeh, A., ve Tayebbeh Motiee Langroudi, S. (2006). A Comparison of Classical Test Theory and Nonparametric Item Response Theory with Respect to the Effects of Testees' Characteristics on Item Characteristics and vice versa. *Educational Innovations*, 5(4), 41-56.
- Doğan, N., ve Tezbaşaran, A. A. (2003). Klasik test kuramı ve örtük özellikler kuramının örneklem bağlamında karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 25(25).
- Dyehouse, M. A. (2009). A comparison of model-data fit for parametric and nonparametric item response theory models using ordinal level ratings [Doktora Tezi]. Purdue Üniversitesi.
- Embretson, S.E. ve Reise, S.P. (2000). *Item Response Theory For Psychologists*. Lawrence Erlbaum Associate, Inc.
- Erkuş, A. (2003). *Psikometri üzerine yazılar* (1.baskı). Ankara: Türk Psikologlar Derneği Yayınları.
- Fan, X. (1998). Item response theory and classical theory: an empirical comparison of their item-person statistic. *Educational and Psychological Measurement*, 58(3), 357-381. <https://doi.org/10.1177/0013164498058003001>
- Field, A. (2009). *Discovering Statistics Using SPSS for Windows*. London – Thousand Oaks – New Delhi: Sage Publications.
- George, D., ve Mallery, P. (2019). *IBM SPSS statistics 26 step by step: A simple guide and reference* (4.baskı). New York: Routledge.
- Ghiselli, E. E., Campbell, J. P., ve Zedeck, S. (1981). *Measurement Theory For The Behavioral Sciences*. San Francisco: W.H. Freeman.
- Hambleton, R. K., Swaminathan, H. ve Rogers, H. (1991). *Fundamentals Of Item Response Theory*. Newbury Park CA: Sage.

- Jones, L. V., ve Thissen, D. (2006). A History and Overview of psychometrics. *Psychometrics*, 26, 1–27. doi:10.1016/s0169-7161(06)26001-2
- Junker, B. W., ve Sijtsma, K. (2001). Nonparametric item response theory in action: An overview of the special issue. *Applied Psychological Measurement*, 25(3), 211-220. Doi: [10.1177/0146621012203202](https://doi.org/10.1177/0146621012203202)
- Kadioğlu Ateş, H., & Mazi, M. G. (2017). Kitap Özeti: Bilimsel araştırma yöntemleri (Şener Büyüköztürk, Ebru Kılıç Çakmak, Özcan Erkan Akgün, Şirin Karadeniz, Funda Demirel). *Avrasya Eğitim ve Literatür Dergisi*, 7, 56-94. <https://hdl.handle.net/20.500.12436/2837>
- Keeves, J. P. (Ed.). (1998). *Educational research, methodology and measuremet: An international Hand-book*. Oxford: Pergamon Pres.
- Kelecioğlu, H. (2001). Örtük özellikler teorisindeki b ve a parametreleri ile klasik test teorisindeki p ve r istatistikleri arasındaki ilişki. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 20, 104-110.
- Kıbrıslıoğlu Uysal, N. (2020). Parametrik ve parametrik olmayan madde tepki modellerinin kestirim hatalarının karşılaştırılması. (Doktora tezi). Hacettepe Üniversitesi.
- Kılıç, S. (2013). Örneklem yöntemleri. *Journal of Mood Disorders*, 3(1), 44-46.
- Koğar, H. (2014). *Madde tepki kuramının farklı uygulamalarından elde edilen parametrelerin ve model uyumlarının örneklem büyüklüğü ve test uzunluğu açısından karşılaştırılması* [Yayımlanmamış doktora tezi]. Hacettepe Üniversitesi.
- Koğar, H. (2015). Madde tepki kuramına ait parametrelerin ve model uyumlarının karşılaştırılması: Bir Monte Carlo çalışması. *Journal of Measurement and Evaluation in Education and Psychology*, 6(1), 142-157. <https://doi.org/10.21031/epod.02072>
- Koğar, H. (2019). Belirti Tarama Listesi'nin (SCL-90) geçerlik ve güvenirlik çalışması: Mokken ölçekleme analizleri. *Turkish Psychological Counseling and Guidance Journal*, 9(54), 689-705.
- Koğar, H. (2022). *R ile geçerlik ve güvenirlik analizleri* (3.baskı). Ankara: Pegem Akademi.
- Köse, İ. A. (2010). *Madde tepki kuramına dayalı tek boyutlu ve çok boyutlu modellerin test uzunluğu ve örneklem büyüklüğü açısından karşılaştırılması* [Yayımlanmamış doktora tezi]. Ankara Üniversitesi.
- Kujipers, R. E., Van der Ark, L. A. , Croon, M. A. ve Sijtsma, K. (2016). Bias in point estimates and standard errors of Mokken's scalability coefficients. *Applied Psychological Measurement*, 40 (5), 331-345. Doi: [10.1177/0146621616638500](https://doi.org/10.1177/0146621616638500)
- Lee, S. H. (2007). *Multiimentional Item Response Theory: A SAS MDIRT MACRO and Emprical Study of PIAT MATH Test* [Yayımlanmamış doktora tezi]. Oklahoma Üniversitesi.
- Lord, F. M. ve Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. USA: Information Age Publishing.
- McDonald, R.P. ve Paunonen, S. (2002). A Monte Carlo Comparison of Item and Person Statistics based on Item Response Theory versus Classical Test Theory. *Journal of Educational and Psychological Measurement*, 62 (6), 921-943. Doi: [10.1177/0013164402238082](https://doi.org/10.1177/0013164402238082)
- McDonald, R.P. (1982). Linear Versus Models in Item Response Theory Applied *Psychological Measurement*, 6 (4), 379-396. <https://doi.org/10.1177/014662168200600402>
- Meijer, R.R., ve Baneke, J.J. (2004). Analyzing psychopathology items: a case for nonparametric item response theory modeling. *Psychological Methods*, 9 (3), 354- 368. <https://doi.org/10.1037/1082-989X.9.3.354>
- Meijer, R. R., Sijtsma, K., ve Smid, N. G. (1990). Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. *Applied Psychological Measurement*, 14(3), 283-298. <https://doi.org/10.1177/014662169001400306>
- Meijer, R. R., Tendeiro, J. N., ve Wanders, R. B. K. (2015). The use of nonparametric item response theory to explore data quality. In S. P. Reise ve D. A. Revicki (Ed.), *Handbook of item response theory modeling: Applications to typical performance assessment* (s. 85–110) içinde. Routledge/Taylor & Francis Group.
- Molenaar, I. W. (2001). Thirty years of nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 295-299. <https://doi.org/10.1177/01466210122032091>
- Mor-Dirlik, E. (2021). Farklı test kuramlarından hesaplanan madde ayırt edicilik parametrelerinin karşılaştırılması. *Trakya Eğitim Dergisi*, 2 (11), 732-744. <https://doi.org/10.24315/tred.700445>
- Mor-Dirlik, E. M., ve Koç, N. (2019). Parametrik ve parametrik olmayan madde tepki kuramı modellerinden çeşitli faktörlere göre elde edilen madde ve yetenek kestirimlerinin karşılaştırılması. *İlköğretim Online*, 18(3). Doi: 10.17051/ilkonline.2019.610162

- Mor-Dirlik, E. (2017). *Parametrik ve parametrik olmayan madde tepki kuramı modellerinden çeşitli faktörlere göre elde edilen madde ve yetenek kestirimlerinin karşılaştırılması* [Yayımlanmamış doktora tezi]. Ankara Üniversitesi.
- Özgülven, İ. E. (2020). *Psikolojik testler* (16. baskı). Ankara: Nobel Akademik Yayıncılık.
- Progar, S., Socan, G., ve Peč, M. (2008). An empirical comparison of item response theory and classical test theory. *Horizons of Psychology*, 17(3), 5-24.
- R Development Core Team (2022). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org/>.
- Reise, S. P., Ainsworth, A. T., ve Haviland, M. G. (2005). Item response theory: Fundamentals, applications, and promise in psychological research. *Current Directions In Psychological Science*, 14(2), 95-101. <https://psycnet.apa.org/doi/10.1111/j.0963-7214.2005.00342.x>
- Reyhanlıoğlu, Ç., ve Doğan, N. (2020). Parametrik ve parametrik olmayan madde tepki kuramında farklı örneklem büyüklüklerine ve boyutluluklarına göre parametre değişmezliğinin incelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 11(2), 98-112. Doi: [10.21031/epod.584977](https://doi.org/10.21031/epod.584977)
- Sayın, A., ve Koğar, H. (2019). Grup başarısına göre madde ve kişi parametreleri arasındaki ilişkinin incelenmesi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 39(1), 249-270.
- Sijtsma, K. (2005). Nonparametric item response theory models. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (s. 875-882) içinde. Elsevier.
- Sijtsma, K., ve Junker, B. W. (2006). Item response theory: Past performance, present developments, and future expectations. *Behaviormetrika*, 33(1), 75-102.
- Sijtsma, K., ve Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Newbury Park. CA: Sage.
- Sijtsma, K. ve Meijer, R. R. (2007). Nonparametric item response theory. C. R. Rao. ve S. Sinharay (Ed.) *Handbook of statistics 26: psychometrics*. (s.719-746) içinde. Amsterdam: Elsevier
- Sijtsma, K., Meijer, R. R., ve Van der Ark, L. A. (2011). Mokken scale analysis as time goes by: An update for scaling practitioners. *Personality and Individual Differences*, 50(1), 31-37. <https://doi.org/10.1016/j.paid.2010.08.016>
- Spencer, G.S. (2004). *The Strength of Multidimensional Item Response Theory in Exploring Construct Space That is Multidimensional and Correlated* [Yayımlanmamış doktora tezi]. Brigham Young Üniversitesi.
- Stout, W., Habing, B., Douglas, J., Kim, H. R., Roussos, L., ve Zhang, J. (1996). Conditional covariance-based nonparametric multidimensionality assessment. *Applied Psychological Measurement*, 20 (4), 331-354. <https://doi.org/10.1177/014662169602000403>
- Strout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55(2), 293-325. <https://doi.org/10.1007/BF02295289>
- Tarkun-Tavşancıl, E. (1996). Alternatif korelasyon teknikleri. *Marmara Üniversitesi Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi*, 8(8), 237-249.
- Thorndike, R. B. ve Thorndike-Christ, T. (2017). Ölçme aracında aranan özellikler: Güvenirlik. M. Otrar (Çev. Ed.) ve E. B. Tunç (Çev.). *Psikolojide ve eğitimde ölçme ve değerlendirme* (1.baskı, 118-153) içinde. Nobel Akademik Yayıncılık. (Orijinal eserin yayın tarihi 2010, 8.baskı).
- Turgut, F. ve Baykul, Y. (2019). *Eğitimde ölçme ve değerlendirme* (8.baskı). Ankara: Pegem Akademi.
- Van Schuur, W. H. (2003). Mokken scale analysis: between the guttmann scale and parametric item response theory. *Political Analysis*, 11(2), 139-163. Doi: [10.7910/dvn/8vwe6a](https://doi.org/10.7910/dvn/8vwe6a)
- Volainiuk, N., ve Bokovets, O. (2017). Psychological structure of innovative potential of a personality. *Science Education*, 12(1), 9-15. Doi: 10.24195/2414-4665-2017-1-2
- Yıldırım, S. (2011). Self-efficacy, intrinsic motivation, anxiety and mathematics achievement: Findings from Turkey, Japan and Finland. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, 5(1), 277-291.
- Zhao, Y. (2008). *Approaches for addressing the fit of item response theory models to educational test data* [Doktora tezi]. Massachusetts Amherst Üniversitesi. <https://core.ac.uk/download/pdf/334980854.pdf>
- Wismeijer, A. A. J., Sijtsma, K., Van Assen, M.A.L.M. ve Vingerhoets, J.J.M. (2008). A comparative study of the dimensionality of the self concealment scale using principal component analysis and Mokken scale analysis. *Journal of Personality Assessment*, 90 (4), 323-334.

8. EKLER

EK.1 Açıklama

Ek.1’de KTK’ya dayalı madde ayırt edicilik indekslerinin kestirilmesi için yazılan R kodları yer almaktadır.

install.packages("CTT")

#Analizde kullanılacak CTT paketinin kurulması için yazılan kod.

library (CTT)

#Analizde kullanılacak CTT paketinin yüklenmesi için yazılan kod.

madde_analizi <- itemAnalysis(veri_son, itemReport = TRUE, NA.Delete = TRUE, rBisML = FALSE, .20, .80, .20, .20)

#Madde analizi yapabilmek için yazılan kod. “veri_son” araştırma kapsamında kullanılan veri setini ifade etmektedir. Araştırmacıların bu kısma kendi veri setlerinin isimlerini yazması gerekmektedir.

madde_analizi\$itemReport

Madde istatistiklerinin sunulduğu tabloda yer alan “itemMean” madde güçlük indeksi değerlerini, “pBis” nokta çift serili korelasyon katsayılarını, “bis” çift serili korelasyon katsayılarını ifade etmektedir.

EK.2 Açıklama

Ek.2’de POMTK’ya dayalı madde ayırt edicilik indekslerinin kestirilmesi için yazılan R kodları yer almaktadır. Araştırma kapsamında Monoton Homojenlik Modeli kullanılmıştır.

install.packages("mokken")

#Analizde kullanılacak mokken paketinin kurulması için yazılan kod.

library(mokken)

#Analizde kullanılacak mokken paketinin yüklenmesi için yazılan kod.

monoton <- check.monotonicity(veri_son) ve summary(monoton)

#Gizil monoton artma varsayımının test edilmesi ve bulguların özetlenmesi için yazılan kod.

plot(monoton)

#Gizil monoton artma varsayımına dair grafiğin çizilmesi için yazılan kod.

dms <- check.iio(veri_son)

#Değişmez madde sıralaması varsayımının test edilmesi için yazılan kod.

summary(dms)

Değişmez madde sıralaması varsayımına ait bulguların özetlenmesi için yazılan kod.

coefH(veri_son)

#H katsayılarının elde edilmesi için yazılan kod.

coefZ(veri_son)

#H katsayılarına ait Z değerlerinin elde edilmesi için yazılan kod.

check.reliability(veri_son, LCRC = TRUE, MS= TRUE)

#Güvenirlilik katsayılarının elde edilmesi için yazılan kod.

Ek.3 Açıklama

Ek.3'te KTK'ya dayalı olarak kestirilen güvenilirlik katsayılarının elde edilmesi için yazılan kodlar yer almaktadır.

install.packages("psych")

#Analizde kullanılacak psych paketinin kurulması için yazılan kod.

library (psych)

#Analizde kullanılacak psych paketinin yüklenmesi için yazılan kod.

alpha(veri_son, n.iter=100, use="pairwise", n.obs = 592)

*#Cronbach alfa güvenilirlik katsayısının hesaplanması için yazılan kod. "n.obs"
=veri setindeki katılımcı sayısı yazılmalıdır.*

guttman (veri_son)

#Lambda katsayıları, beta katsayısı ve tenBerge güvenilirlik sınırlarının hesaplanması için yazılan kod.yöntem