



HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

MADDE TAKIMI TABANLI TESTLERDE ÖLÇEK DÖNÜŞTÜRME HATALARININ İNCELENMESİ

Ertunç UKŞUL

Doktora Tezi

Ankara, 2024

Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

Daha ileriye ... En İyiyeye ...



Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

MADDE TAKIMI TABANLI TESTLERDE ÖLÇEK DÖNÜŞTÜRME HATALARININ
İNCELENMESİ

INVESTIGATION OF SCALE CONVERSION ERRORS IN TESTLET-BASED TESTS

Ertunç UKŞUL

Doktora Tezi

Ankara, 2024

Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

Ertunç UKŞUL'un hazırladıđı "Madde Takımı Tabanlı Testlerde Ölçek Dönüştürme Hatalarının İncelenmesi" başlıklı bu çalışma jürimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde Ölçme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiştir.

J¼ri Başkanı	Prof. Dr. Selahattin GELBAL
J¼ri Üyesi (Danışman)	Prof. Dr. Hülya KELECİOđLU
J¼ri Üyesi	Prof. Dr. Cem Oktay GÜZELLER
J¼ri Üyesi	Doç. Dr. Neşe ÖZT¼RK G¼BEŞ
J¼ri Üyesi	Doç. Dr. Sevda ÇETİN

Bu tez Hacettepe Üniversitesi Lisansüstü Eđitim, Öğretim ve Sınav Yönetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından 02 / 02 / 2024 tarihinde uygun gör¼lm¼ş ve Enstitü Yönetim Kurulunca / / tarihi itibarıyla kabul edilmiştir.

Prof. Dr. İsmail Hakkı MİRİCİ
Eđitim Bilimleri Enstitüsü M¼d¼r¼

Öz

Bu çalışmada, testlerde yer alan madde takımlarının uygulanan modellere göre ölçek dönüştürme hatasına olan etkisi araştırılmıştır. Bu amaçla, madde takımı içeren testlere madde tepki kuramı modellerinden tek boyutlu madde tepki kuramı, iki faktör ve madde takımı tepki kuramı modelleri uygulanarak madde ve yetenek parametreleri kestirilmiştir. Testlerin eşitlenebilmesi için elde edilen parametreler, eşdeğer olmayan gruplar ortak madde deseni altında ortalama-ortalama, ortalama-sigma, Stocking-Lord ve Haebara ölçek dönüştürme yöntemleri kullanılarak ortak ölçeğe yerleştirilmiştir. Daha sonra modellerin ölçek dönüştürme yöntemine bağlı olarak değişen hataları hesaplanmış ve karşılaştırılmıştır. Değerlendirme ölçütü olarak hata karelerinin ortalamasının karekökü (RMSE) tercih edilmiştir. Araştırmada, veri toplama aracı olarak Trends in International Mathematics and Science Study (TIMSS) projesinin 2019 yılında uygulanan fen bilimleri testi kullanılmıştır. Modeller madde takımı sayısı, bağımsız madde sayısı ve örneklem büyüklüğü koşullarına göre incelenmiştir. Araştırma sonucunda bağımsız madde sayısı ve örneklem büyüklüğü arttıkça genel olarak hatanın azaldığı görülmüştür. Madde takım sayısı arttıkça iki faktör ve madde takımı tepki kuramı modellerinde hatanın azaldığı, tek boyutlu madde tepki kuramı modelinde ise hatanın arttığı bulunmuştur. Bununla birlikte Haebera ve Stocking-Lord ölçek dönüştürme yöntemlerinden elde edilen elde edilen eşitleme hatalarının ortalama-ortalama ve ortalama-sigma yöntemlerine göre daha düşük olduğu sonucuna ulaşılmıştır.

Anahtar Sözcükler: test eşitleme, ölçek dönüştürme, madde tepki kuramı, madde takımı tepki kuramı modeli, iki faktör model

Abstract

This study aims to examine the effects of testlets on scale conversion. For this purpose unidimensional item response theory, bifactor and testlet response theory models were applied to the testlet-based tests for the estimation of item and ability parameters. In order to equate the tests, the parameters were placed on the common scale using mean-mean, mean-sigma and Stocking-Lord and Haebara scale conversion methods under the common-item non-equivalent groups design. Then, the errors of the models depending on the scale conversion method were calculated and compared. Equating errors were compared with Root Mean Squared Error (RMSE). In the study, the science test of the Trends in International Mathematics and Science Study (TIMSS) project administered in 2019 was used as the data collection tool. Models were examined according to the conditions as the the number of testlets, the number of independent items and the sample size. As a result of the study, it was observed that the error decreased in all models as the number of independent items and sample size in general. It was found that as the number of testlet increased, the error decreased in bifactor model and testlet response theory model, while the error increased in unidimensional item response theory model. In addition, it was concluded that the equating errors obtained from the Haebera and Stocking-Lord scale transformation methods were lower than the mean-mean and mean-sigma methods.

Keywords: test equating, scale conversion, item response theory, testlet response theory model, bifactor model

Teşekkür

Öncelikle doktora eğitimim boyunca bana rehber olan, hoşgörüsü ve desteğiyle eğitimime ve tezime büyük katkılar sağlayan değerli danışman hocam Prof. Dr. Hülya KELEÇİOĞLU'na,

Tezimin gelişiminde anlayış ve sabırla destek veren değerli tez izleme komitesi üyelerim Prof. Dr. Selahattin GELBAL'a ve Prof. Dr. Cem Oktay GÜZELLER'e,

Tez sürecinde destek ve katkılarıyla yanımda olan Doç. Dr. Sevda ÇETİN'e, Doç. Dr. Neşe ÖZTÜRK GÜBEŞ'e, Araş. Gör. Dr. Yusuf AYDIN'a, Dr. Yaşar Mehmet ZOR'a, Dr. Süleyman YILDIRIM'a, Uzm. Cabir Alparslan DEMİR'e, Dr. Hüseyin İnan ŞAYRAN'a, Murat BULUT'a, Uğur YEŞİLGÖZ'e, Sevinç BAŞARAN'a ve değerli aileme teşekkür ederim.

İçindekiler

Kabul ve Onay.....	ii
Öz.....	iii
Abstract.....	iv
Teşekkür.....	v
Tablolar Dizini.....	viii
Şekiller Dizini.....	ix
Simgeler ve Kısaltmalar Dizini.....	x
Bölüm 1 Giriş.....	1
Problem Durumu.....	3
Araştırmanın Amacı ve Önemi.....	4
Araştırma Problemi.....	5
Sayıtlılar.....	6
Sınırlılıklar.....	6
Tanımlar.....	7
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	8
Madde Tepki Kuramı Modeli.....	8
İki Faktör Modeli.....	21
Madde Takımı Tepki Kuramı Modeli.....	24
Test Eşitleme.....	26
İlgili Araştırmalar.....	31
Bölüm 3 Yöntem.....	38
Araştırmanın Türü.....	38
Çalışma Grubu.....	38
Veri Toplama Aracı.....	39
Verilerin Analizi.....	41
Değerlendirme Kriteri.....	43

Bölüm 4 Bulgular, Yorumlar ve Tartışma.....	45
Alt Problemlere İlişkin Bulgular ve Yorumlar	45
Bölüm 5 Sonuç ve Öneriler.....	68
Sonuçlar	68
Öneriler	72
Kaynaklar	76
EK-A: Araştırmada Yer Alan Koşullara İlişkin Elde Edilen RMSE Değerleri	85
EK-B: Araştırma Etik Komisyon İzin Muafiyeti Formu/ Araştırma Etik Komisyonu Onay Bildirimi	86
EK-C: Etik Beyanı.....	87
EK-Ç: /Doktora Tez Çalışması Orijinallik Raporu	88
EK-D: Thesis/Dissertation Originality Report.....	89
EK-E: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	90

Tablolar Dizini

Tablo 1 <i>Veri Setlerindeki Madde Takımı ve Bağımsız Madde Sayıları</i>	40
Tablo 2 <i>Madde Takımlarındaki Maddelerin LD X^2 ve Q_3 değerleri</i>	42
Tablo 3 <i>Çalışmada İncelenen Değişken ve Koşullar</i>	43
Tablo 4 <i>Ortalama-Ortalama Yönteminden Elde Edilen RMSE Değerleri</i>	46
Tablo 5 <i>Ortalama-Sigma Yönteminden Elde Edilen RMSE Değerleri</i>	50
Tablo 6 <i>Stocking-Lord Yönteminden Elde Edilen RMSE Değerleri</i>	54
Tablo 7 <i>Haebara Yönteminden Elde Edilen RMSE Değerleri</i>	58
Tablo 8 <i>Modellere Göre Maddelerin a Parametresinden Kestirilen Ortalama Hata Değerlerinin Karşılaştırılması</i>	62
Tablo 9 <i>Modellere Göre Maddelerin a Parametresinden Kestirilen Ortalama Hata Değerlerine İlişkin İkili Karşılaştırma Testi Sonuçları</i>	63
Tablo 10 <i>Modellere Göre Maddelerin c Parametresinden Kestirilen Ortalama Hata Değerlerinin Karşılaştırılması</i>	64
Tablo 11 <i>Modellere Göre Maddelerin c Parametresinden Kestirilen Ortalama Hata Değerlerine İlişkin İkili Karşılaştırma Testi Sonuçları</i>	65
Tablo 12 <i>Modellere Göre Theta Parametresinden Kestirilen Ortalama Hata Değerlerinin Karşılaştırılması</i>	66
Tablo 13 <i>Modellere Göre theta Parametresinden Kestirilen Ortalama Hata Değerlerine İlişkin İkili Karşılaştırma Testi Sonuçları</i>	67

Şekiller Dizini

Şekil 1 <i>1PLM için Madde Karakteristik Eğrisi</i>	11
Şekil 2 <i>2PLM için Madde Karakteristik Eğrisi</i>	13
Şekil 3 <i>3PLM için Madde Karakteristik Eğrisi</i>	14
Şekil 4 <i>Madde Karakteristik Eğrisi</i>	15
Şekil 5 <i>Test Bilgi Fonksiyonu (TBF)</i>	17
Şekil 6 <i>İki Faktör Modelin Yapısı</i>	22
Şekil 7 <i>İki Faktör Modelin Örüntüsü</i>	23
Şekil 8 <i>Madde Takımı Tepki Kuramı Modelinin Yapısı</i>	24
Şekil 9 <i>Madde Takımı Tepki Kuramı Modelinin Örüntüsü</i>	25
Şekil 10 <i>Eşdeğer olmayan gruplar ortak madde deseni</i>	27
Şekil 11 <i>Ortalama-Ortalama Yönteminden Elde Edilen RMSE Değerlerinin Değişimi</i>	48
Şekil 12 <i>Ortalama-Sigma Yönteminden Elde Edilen RMSE Değerlerinin Değişimi</i>	52
Şekil 13 <i>Stocking Lord Yönteminden Elde Edilen RMSE Değerlerinin Değişimi</i>	56
Şekil 14 <i>Haebara Yönteminden Elde Edilen RMSE Değerlerinin Değişimi</i>	60

Simgeler ve Kısaltmalar Dizini

TIMMS: Trends in International Mathematics and Science Study

MTK: Madde Tepki Kuramı

TBMTK: Tek Boyutlu Madde Tepki Kuramı

2FM: İki faktör Model

MTTK: Madde Takımı Tepki Kuramı

MKE: Madde Karakteristik Eğrisi

TBF: Test Bilgi Fonksiyonu

MBF: Madde Bilgi Fonksiyonu

Bölüm 1

Giriş

Eğitim sistemi, bireylerin ve toplumların gelişimine katkı sağlayan önemli bir yapıdır. Eğitim sistemine yön veren temel öğelerden biri de öğrencilerin bilgi, beceri ve yeteneklerini değerlendirmek amacıyla uygulanan sınavlar ve ölçme araçlarıdır. Bu değerlendirmelerin geçerli ve güvenilir bir biçimde yapılması, eğitim sistemine doğru yön vermek için büyük bir önem taşımaktadır.

Uluslararası alanda ülkelerin eğitim sistemlerinin değerlendirilmesinde The Trends in International Mathematics and Science Study (TIMSS), Programme for International Student Assessment (PISA) gibi büyük ölçekli sınavlar kullanılmaktadır. Bu sınavlardan toplanan veriler, ülkelerin öğrencilerin başarılarını uluslararası ve ulusal düzeyde, öğrenci ve okul özelliklerine göre anlamalarına, karşılaştırmalarına ve analiz etmelerine olanak tanır (Martin ve diğerleri, 2008). Böylece, sonuçlar hem eğitimciler hem de politika yapıcılar için eğitim reformları yapmak üzere kanıta dayalı kararlar alınırken güvenilir bir kaynak oluşturur (Gonzales ve diğerleri, 2009).

Büyük ölçekli testler, boylamsal ve kestisel karşılaştırmalar yapabilmek için farklı zamanlarda, farklı gruplara uygulanmak durumundadır. Bu testlerde kullanılan soru ve soru kitapçıkları da çoğu zaman kendi içerisinde farklılık gösterebilmektedir. Bu noktada, aynı amaca yönelik farklı zamanlarda veya farklı ölçme araçları kullanılarak elde edilen öğrenci performans verilerinin karşılaştırılabilir olması gerekmektedir. Çünkü uygulanan testlerdeki farklılıklar, testlerin ve öğrencilerin değerlendirilmesinde ve karşılaştırılmasında bir takım hatalara yol açabilmektedir. Bu hataları azaltmak, testlerden daha geçerli ve güvenilir sonuçlar elde etmek ve elde edilen sonuçları daha doğru bir şekilde yorumlamak için testlerin eşitlenmesi gerekebilmektedir (Cook & Eignor, 1991). Bu nedenle, birden fazla test formu kullanılan sınavlar için testlerin eşitlenmesi ölçme ve değerlendirme açısından dikkat edilmesi gereken bir süreçtir (Gübeş & Kelecioğlu, 2015) .

Cook ve Eignor'a (1991) göre standartlaştırılmış bir testin ana amaçlarından biri, adil ve karşılaştırılabilir bir psikolojik veya eğitimsel değerlendirme sağlayabilmesidir. Test eşitleme uygulaması da, öğrenci performansının zaman içinde izlenmesine ve farklı öğrenci grupları arasındaki başarı farklılıklarının standart biçimde analiz edilmesine dolayısıyla öğrencilerin adil bir şekilde değerlendirilmesine ve karşılaştırılmasına yardımcı olur.

Test eşitlemede kullanılan birçok model ve yöntem vardır. Bunlardan bazıları Madde Tepki Kuramı (MTK) modellerini kullanırken diğerleri Klasik Test Kuramı'na (KTK) dayalı modeller kullanmaktadır. MTK, madde ve test istatistiklerini gruptan, yetenek düzeylerini ise testten bağımsız kestirilmesini sağlayan KTK'ya göre matematiksel altyapısı güçlü modellerdir (Hambleton & Swaminathan, 1985). Babcock'un (2009) da belirttiği gibi MTK, KTK'nın sınırlılıklarını aşarak daha güçlü varsayımlar geliştirir. MTK, bu sınırlılıkları aşarken güçlü bir matematiksel temele dayanır ve bilgisayarlar tarafından işlenebilen karmaşık algoritmalar içerir. Bu nedenle MTK'nın güvenilir sonuçlar verebilmesi için gerekli varsayımların sağlanması ve doğru matematiksel modellerin seçilmesi oldukça önemlidir. Uygulanan modelin model-veri uyumu, maddeleri kalibre etmek için kullanılan algoritmayı ve madde-yetenek parametre kestirimini dolayısıyla eşitleme sonuçlarının doğruluğunu doğrudan etkiler (Hambleton & Swaminathan, 1985; Zhao, 2008).

MTK ile eşitleme yapılabilmesi için karşılanması gereken bazı varsayımlar vardır. Bu varsayımların en önemlilerinden biri ise yerel bağımsızlıktır. Yerel bağımsızlık, yetenek faktörünün sabit olduğu durumlarda, her iki soruya verilen yanıtların istatistiksel olarak birbirlerinden bağımsız olduğunu ifade eder (Lord, 1980). Bir başka deyişle, maddelerin doğru cevaplanma olasılıklarının, belirli bir yetenek düzeyinde o bireyin yeteneği dışında başka bir etkene bağlı olmamasıdır (Hambleton & Swaminathan, 1985). Eğer bir sorunun içeriği başka bir soruya verilecek yanıtı etkileyebilecek ipuçları ya da yönlendirmeler içeriyorsa ve bu durum o sorunun doğru cevaplanmasını etkiliyorsa, bu, testin yerel bağımsızlık ilkesinin ihlal edildiğini gösterir (Embretson & Reise, 2000). Yerel madde bağımlılığı (local item dependence- LID)

ihlalinin birçok nedeni ve türü vardır. Bunlardan en önemlisi birden fazla test maddesinin bir paragrafa bağlı olmasıdır.

Problem Durumu

Büyük ölçekli testlerin birçoğu, karmaşık ve birbiriyle ilişkili sorulara olanak tanıyan, testi zaman açısından daha verimli kılan aynı paragrafa bağlı birden fazla maddenin oluşturduğu madde takımı adı verilen soru tipini içermektedir (DeMars, 2006). Büyük ölçekli testlerin bu içeriğinden dolayı, yerel bağımlılık üzerine yapılan çalışmalarda en çok paragraf bağımlılığına odaklanılmış, madde takımlarının kendi içerisinde birbiriyle olan ilişkilerinin ve doğru cevabı bulmaya etkilerinin yerel bağımlılığa neden olduğu belirlenmiştir (Koğar & Kelecioğlu, 2017).

Madde takımı içindeki maddelerin ortak bir uyarıcıyı paylaşmaları maddelere verilen yanıtların yerel olarak bağımsız olmayacağı anlamına gelir. Bu nedenle, koşullu bağımlılıkların tespit edilmesi önem arz etmektedir. Testlet içindeki yerel bağımlılığı modellemek veya tespit etmek üzere birçok çalışma gerçekleştirilmiştir. Drasgow vd. (1985) standart tek boyutlu madde tepki modellerinden sapmaların tanımlanması için parametrik yaklaşımları açıklamış, Stout (1987) tek boyutluluğun ihlal edilip edilmediğini belirlemek için parametrik olmayan yaklaşımlar geliştirmiştir. Başlangıçta, geleneksel MTK modeli, her bir madde takımı içindeki maddelere verilen yanıtlar arasındaki koşullu bağımlılıkları göz ardı ederek madde takımı tabanlı testleri kalibre etmek için kullanılmıştır. Bu nedenle, mevcut yerel bağımlılıklar, geleneksel MTK modelinin koşullu bağımsızlık varsayımını ihlal ettiği için sorunlar ortaya çıkmıştır. Bu nedenle, bağımlılığı tespit edebilen bazı yeni modeller veya mevcut modeller, madde takımı tabanlı testlere uyacak şekilde geliştirilmiştir. Bu probleme yanıt bulabilmek için bu çalışmada test eşitleme sürecinde yerel madde bağımlılığına neden olan madde takımları üzerinde durulmuştur.

Çalışmada madde takımı tabanlı testlerden tek boyutlu madde tepki kuramı (TBMTK), iki faktör model (2FM) ve madde takımı tepki kuramı (MTTK) modelleri kullanılarak elde edilen

madde ve yetenek parametrelerine ölçek dönüştürme işlemleri uygulandığında parametrelere ilişkin eşitleme hatalarının nasıl değiştiği sorusuna yanıt aranmıştır.

Araştırmanın Amacı ve Önemi

Test eşitleme, farklı test formlarının puan ölçeklerinin uygun yöntemlerle ayarlanması ile ölçme aracından kaynaklanan göreceli zorluktaki farklılıkları telafi etmek ve böylece puanları eşdeğer ve karşılaştırılabilir hale getirmek amacıyla yapılır. Yapılan bu işlem sayesinde, aynı amaca yönelik farklı test formlarından elde edilen puanların birbirinin yerine kullanılması sağlanır (Angoff, 1984; Hambleton & Swaminathan, 1985). Test eşitlemenin sağladığı bu imkân, test formlarının güçlük seviyeleri arasındaki farklılıkların neden olduğu avantaj veya dezavantajların önüne geçer ve öğrencilerin daha kolay veya daha zor bir test formu almasından kaynaklanan olumsuz etkileri azaltır (Tan, 2005).

Madde takımı, bir test veya sınavda belirli bir konu veya konu alanına odaklanan, öğrencilerin belirli bir konudaki bilgi ve becerilerini ölçmeye yönelik daha özel ve verimli bir değerlendirme yapılmasına olanak tanıyan, genellikle birkaç sorudan oluşan bir alt birimdir. Bu özelliği sayesinde öğrenci performansını daha ayrıntılı bir şekilde değerlendirmek ve öğrenme sürecini daha etkili bir şekilde yönlendirmek amacıyla test geliştiren eğitimciler ve ölçme uzmanları tarafından son zamanlarda sıklıkla kullanılmaktadır (DeMars, 2006; Sireci ve diğerleri, 1991).

Madde takımlarının testlerde kullanımı daha öncesine dayansa da madde takımı (testlet) tanımı ilk kez 1987 yılında Wainer ve Kiely tarafından literatüre kazandırılmıştır (Wainer ve diğerleri, 2007). Bu tarihten itibaren madde takımlarının test üzerindeki etkileri eğitim araştırmacıları tarafından dikkate değer bulunmuş ve bu konuda çeşitli çalışmalar yapılmıştır (Ackerman, 1987; Bishop & Omar, 2002; Bradlow, Wainer & Wang, 1999; Wainer & Wang, 2000; Wainer ve diğerleri, 2007; Yen, 1993; Li, Bolt & Fu, 2006). Bu çalışmalar, madde takımları içeren testlere madde takımı etkisini göz ardı eden geleneksel MTK modellerinin uygulanmasının, MTK'nın yerel bağımsızlık varsayımını ihlal ettiğini ve ölçekleme

hatalarına yol açtığını göstermiştir. Bu nedenle MTK'ya alternatif olarak, Madde Takımı Tepki Kuramı (MTTK) adıyla alternatif bir kuram oluşturulmuştur (Lee ve diğerleri, 2001).

Bu araştırmada, madde takımı içeren testlerin eşitlenmesinde, TBMTK, 2FM ve MTTK modellerinden elde edilen ölçek dönüştürme hatalarının (RMSE) eşdeğer olmayan gruplar ortak madde deseni altında çeşitli koşullar altında karşılaştırılması amaçlanmıştır. Literatürde eşitleme çalışmalarında, bu modellerin gerçek veri üzerinde kullanıldığı ve madde takımlarının çeşitli koşullar altında etkisinin incelendiği çalışmalar oldukça kısıtlıdır. Bu araştırmada madde takımı tabanlı testlerin eşitlenmesinde kullanılan modeller gerçek veriler kullanılarak madde takım sayısı, bağımsız madde sayısı, örneklem büyüklüğü gibi koşullarda karşılaştırıldığı için araştırmanın sonuçlarının ilgili alan için önemli olacağı düşünülmektedir.

Araştırma Problemi

Madde takımı tabanlı testlerde TBMTK, 2FM ve MTTK modelleri kullanılarak kestirilen madde ve yetenek parametrelerine eşdeğer olmayan gruplar ortak madde deseni altında uygulanan ölçek dönüştürme yöntemlerinin hatası madde takımı sayısına (2/4), bağımsız madde sayısına (11/22) ve örneklem büyüklüğüne (1000/7946) göre nasıl değişmektedir?

Alt Problemler

1. Madde takımı tabanlı testlerde TBMTK, 2FM ve MTTK modelleri kullanılarak kestirilen madde ve yetenek parametrelerine eşdeğer olmayan gruplar ortak madde deseni altında ortalama-ortalama ölçek dönüştürme yöntemi uygulandığında elde edilen hatalar madde takımı sayısına (2/4), bağımsız madde sayısına (11/22) ve örneklem büyüklüğüne (1000/7946) göre nasıl değişmektedir?
2. Madde takımı tabanlı testlerde TBMTK, 2FM ve MTTK modelleri kullanılarak kestirilen madde ve yetenek parametrelerine eşdeğer olmayan gruplar ortak madde deseni altında ortalama-sigma ölçek dönüştürme yöntemi uygulandığında elde edilen hatalar madde takımı sayısına (2/4), bağımsız madde sayısına (11/22) ve örneklem büyüklüğüne (1000/7946) göre nasıl değişmektedir?

3. Madde takımı tabanlı testlerde TBMTK, 2FM ve MTTK modelleri kullanılarak kestirilen madde ve yetenek parametrelerine eşdeğer olmayan gruplar ortak madde deseni altında Stocking-Lord ölçek dönüştürme yöntemi uygulandığında elde edilen hatalar madde takımı sayısına (2/4), bağımsız madde sayısına (11/22) ve örneklem büyüklüğüne (1000/7946) göre nasıl değişmektedir?
4. Madde takımı tabanlı testlerde TBMTK, 2FM ve MTTK modelleri kullanılarak kestirilen madde ve yetenek parametrelerine eşdeğer olmayan gruplar ortak madde deseni altında Haebara ölçek dönüştürme yöntemi uygulandığında elde edilen hatalar madde takımı sayısına (2/4), bağımsız madde sayısına (11/22) ve örneklem büyüklüğüne (1000/7946) göre nasıl değişmektedir?
5. TBMTK, 2FM ve MTTK modelleri testin tümüne uygulandığında elde edilen ölçmek dönüştürme hatalarının koşullar arası etkileşimi ve anlamlılığı
 - a) a parametresine
 - b) c parametresine
 - c) theta parametresine göre nasıl değişmektedir?

Sayıtlılar

Bu araştırma TIMSS 2019 uygulamasının Fen Bilimleri testinden elde edilen gerçek verilere dayandığı için sınava giren öğrencilerin test maddelerine samimi yanıtlar verdiği varsayılmıştır.

Sınırlılıklar

Araştırmada, TIMSS uygulamasındaki fen bilimleri alanına ait bilişsel testlerden 7. ve 8. kitapçıklardan elde edilen veriler kullanılmıştır. Bu bakımdan çalışma bu kitapçıklarda yer alan maddeler ve bu testi alan öğrencilerin verdikleri yanıtlar ile sınırlıdır.

Tanımlar

Madde Takımı: Aynı içeriğe baęlı soru maddelerinin oluřturduęu soru kmeleridir.

Baęımsız Madde: Aynı içeriğe baęlı olmayan, tek bařına kullanılan soru maddesidir.

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Madde Tepki Kuramı Modeli

Test eşitleme için birbirinden farklı test formlarından elde edilen test puanlarının ortak bir ölçeğe yerleştirilmesi gerekmektedir. Bu amaçla uygulanan, yeni test formunun puan ve parametre dönüşümlerini sağlayan birçok eşitleme süreci vardır. Bunlardan bazıları Madde Tepki Kuramı (MTK) modellerini kullanırken diğerleri ortalama, doğrusal ve eşit yüzdelikli yöntemler gibi Klasik Test Kuramı'na (KTK) dayalı modeller kullanmaktadır.

KTK, karmaşık kuramsal modeller kurmadan daha çok test düzeyinde bilgi sağlayan, madde parametrelerini gruba bağlı kestiren matematiksel altyapısı zayıf modellerdir. Buna karşın MTK, bireyin performansı ile maddelere verdiği yanıtlar arasındaki ilişkiyi "madde karakteristik eğrisi" adı verilen bir fonksiyonla tanımlayan, madde ve test istatistiklerini gruptan, yetenek düzeylerini ise testten bağımsız kestirilmesini sağlayan matematiksel altyapısı güçlü modellerdir (Hambleton & Swaminathan, 1985). MTK ile kestirilen parametreler testin uygulandığı gruba bağlı olmadığı için değişmezlik özelliğine sahiptir ve bu yöntemler uygulanarak elde edilen sonuçlara bakılarak farklı grupların birbiriyle karşılaştırılması mümkün olmaktadır. MTK, güçlü ve etkili matematiksel altyapıya sahip olduğu için geniş ölçekli testlerle yapılan çalışmalarda sıklıkla kullanılmaktadır (Orlando & Thissen, 2000).

Lord (1952) ve Birnbaum (1968) tarafından geliştirilen MTK, KTK'ya göre birçok önemli avantaj sunmaktadır. Embretson ve Reise (2000), KTK yerine MTK kullanmanın faydalarını aşağıdaki gibi sıralamıştır.

Kurallar	Eski Kurallar (KTK)	Yeni Kurallar (MTK)
Kural 1	Bir ölçümün standart hatası, belirli bir evrende yer alan tüm puanlar için geçerlidir.	Ölçümün standart hatası puanlar (veya yanıt kalıpları) arasında farklılık gösterir, ancak evrene genellenir.
Kural 2	Uzun testler, kısa testlerden daha güvenilirdir.	Kısa testler, uzun testlerden daha güvenilir olabilir.
Kural 3	Çoklu formdaki test puanlarını karşılaştırmak, formlar paralel olduğunda uygundur.	Test puanlarının birden fazla formda karşılaştırılması, test zorluk seviyeleri kişiler arasında farklılık gösterdiğinde optimaldir.
Kural 4	Madde özelliklerinin yansız tahminleri, temsili örneklere sahip olmaya bağlıdır.	Madde özelliklerine ilişkin yansız tahminler, temsili olmayan örneklemlerden elde edilebilir.
Kural 5	Test puanları, norm grubundaki konumlarına göre anlam kazanır.	Test puanları, maddelere olan uzaklık açısından karşılaştırıldıklarında anlam kazanır.
Kural 6	Aralıklı ölçek özellikleri, normal puan dağılımları elde edilerek elde edilir.	Aralıklı ölçek özellikleri, gerekçelendirilebilir ölçme modelleri uygulanarak elde edilir.
Kural 7	Karma madde formatları test toplam puanları üzerinde dengesiz etkiye yol açar.	Karma madde formatları en uygun test puanlarını verebilir.
Kural 8	Başlangıç puan düzeyleri farklı olduğunda değişim puanları anlamlı bir şekilde karşılaştırılmaz.	Başlangıç puan düzeyleri farklı olduğunda değişim puanları anlamlı bir şekilde karşılaştırılabilir.
Kural 9	İkili maddeler üzerinde yapılan faktör analizi, gerçek faktörler yerine yapay faktörler üretir.	Ham madde verisi üzerinde yapılan faktör analizi, tam bilgi faktör analizi sağlar.
Kural 10	Madde uyarıcı özellikleri psikometrik özelliklere kıyasla önemsizdir.	Madde uyarıcı özellikleri, psikometrik özelliklerle doğrudan ilişkili olabilir.

MTK, bir testteki kişi parametresi ile madde parametreleri arasındaki ilişkiye odaklanır.

Lord (1980) MTK'yı şöyle tanımlamıştır:

“Maddeleri madde parametreleriyle ve bireyleri birey parametreleriyle öyle bir şekilde tanımlamamız gerekiyor ki, herhangi bir bireyin herhangi bir maddeye olan tepkisini olasılıksal olarak tahmin edebilelim” (s11).

Cook ve Eignor (1991) MTK'nın amacını, bireylerin yetenek düzeyleri ile belirli bir soruyu doğru cevaplama ihtimalleri arasındaki ilişkiyi matematiksel bir modelle açıklama olarak tanımlamıştır. Crocker ve Algina'nın (1986) da belirttiği gibi MTK'nın hedefi, yetenek düzeyleri farklı bireylerin bir maddeye verecekleri tepkinin matematiksel olarak modellenmesidir.

Bir parametrelili normal ogive modelinin (Lord, 1952) matematiksel formu şu şekildedir:

$$P_i(\theta) = \int_{-\infty}^{\theta-b_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

$P_i(\theta)$, i maddesinin θ yetenek düzeyinde doğru yanıtlanma olasılığını, b_i teorik olarak $-\infty$ ve $+\infty$ arasında, pratik olarak -3 ile +3 arasında yer alan madde güçlük parametresini, z ise yanıtlayıcının i maddesine ilişkin standart puanını ($z = \frac{X - \mu}{\sigma}$) temsil eder (van der Linden & Hambleton, 1997).

Yanıtlayıcının MTK'daki yeteneği kişi parametresi, kişinin yeteneği, theta (θ) parametresi veya kişinin gizli özelliği olarak adlandırılır (Hambleton & van der Linden, 1982). MTK'ya dayalı psikometrik kalibrasyon en az yedi raporlama endeksi sağlar: θ -parametresi, a-parametresi, b-parametresi, c-parametresi, madde karakteristik eğrisi (MKE), madde bilgi fonksiyonu (MBF) ve test bilgi fonksiyonu (TBF). θ -parametresi, yanıt deseninin olasılık fonksiyonu kullanılarak tahmin edilir. θ ve b parametreleri aynı ölçekte olup, sırasıyla ortalama ve standart sapma için 0 ve 1 olacak şekilde ölçeklenebilir.

Parametre Sayısına Göre Madde Tepki Kuramı Modelleri

MTK'da iki kategorili (1-0) puanlanan verilerin modellenmesi için kullanılan üç model vardır. Bunlar bir parametrelili lojistik model (1PLM), iki parametrelili lojistik model (2PLM) ve üç parametrelili lojistik model (3PLM)'dir.

Bir Parametrelili Lojistik Model (1PLM). Adayların yetenek düzeyini sadece madde güçlük parametresini (b) kullanılarak modellerdir.

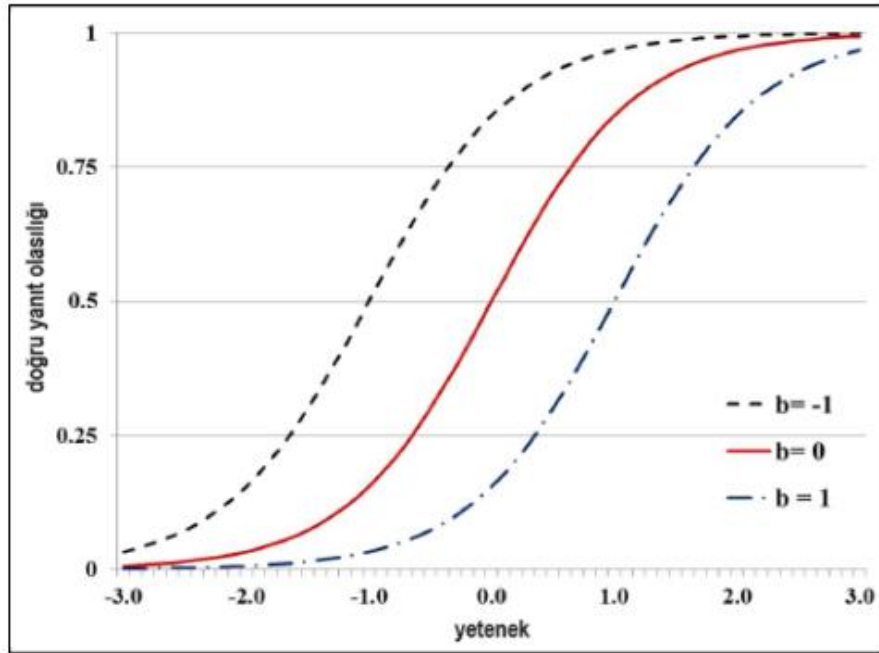
b parametresi -3 ile +3 arasında, ayırt edicilik parametresi (a) her madde için eşit, şans parametresi ise 0 olarak modellenir (DeMars, 2016). 1PLM'nin matematiksel gösterimi eşitlik 1'de verilmiştir.

$$P(Y_{ij} | \theta) = \frac{1}{1 + e^{-D(\theta_i - b_j)}} \quad (1)$$

θ_i yetenek düzeyini, b_j güçlük parametresini, e 2,718 sabitini, D ölçekleme faktörünü ($D=1.7$) temsil eder. Güçlük parametresi, Şekil 1'de gösterildiği gibi yanıtlayıcının gizil yetenek ölçeği üzerinde bir soruyu 0.5 olasılıkla doğru yanıtladığı değere denk gelir.

Şekil 1

1PLM için Madde Karakteristik Eğrisi



Şekil 1, bireyin yetenek düzeylerine göre maddeyi doğru yanıtlama olasılığını gösteren üç farklı madde karakteristlik eğrisini göstermektedir (b=-1 kolay, b=0 orta, b=1 zor madde). 1PLM'de çizilen madde karakteristlik eğrileri sadece b_j'nin farklı yer parametrelerine göre yeri değişen birbirinin aynısı eğrilerden oluşur.

İki Parametrelili Lojistik Model (2PLM). 1PLM'ye madde karakteristik eğrisinin eğimine karşılık gelen değerin (ayırt edicilik parametresi) eklenmiş halidir. Bu parametre MKE'nin dönüm noktasındaki eğimiyle orantılıdır (De Ayala, 2009).

b parametresi -3 ile +3 arasında, ayırt edicilik parametresi (a) 0 ile 2 arasında, şans parametresi ise 0 olarak modellenir (DeMars, 2016).

2PLM'nin matematiksel gösterimi eşitlik 2'de verilmiştir.

$$P(Y_{ij} | \theta) = \frac{1}{1 + e^{-D a_j (\theta_i - b_j)}} \quad (2)$$

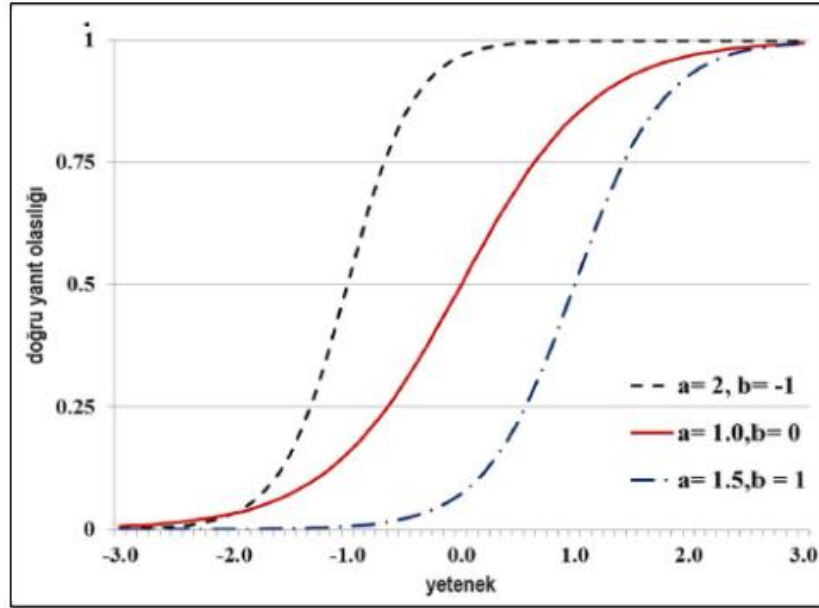
θ_i yetenek düzeyini, a_j ayırt edicilik parametresini, b_j güçlük parametresini, e 2,718 sabitini, D ölçekleme faktörünü (D=1.7) temsil eder.

Eşitlik 2 incelendiğinde bir parametrelili modelin, ayırt edicilik parametresinin eşit olarak sabit kabul edildiği iki parametrelili modelin özel bir durumu olduğu görülebilir. Ayırt edicilik parametresi maddenin verdiği bilgi ile orantılıdır. Madde karakteristik eğrisinde eğim arttıkça a parametresinin değeri artar.

Yüksek a parametresi değerine sahip maddeler belirli yetenek düzeyindeki daha düşük a parametresine sahip maddelerden daha fazla bilgi verir. 2PLM'nin madde karakteristik eğrisi Şekil 2'de gösterilmiştir.

Şekil 2

2PLM için Madde Karakteristik Eğrisi



Şekil 2, yetenek düzeylerine göre farklı ayırt edicilik parametresine sahip üç maddenin doğru yanıt olasılığını belirten üç farklı madde karakteristik eğrisini göstermektedir (1. madde ($a=2, b=-1$), 2. madde ($a=1, b=0$), 3. madde ($a=1.5, b=1$)). Bu maddelerden 3. madde yetenek düzeyi 1 olan yanıtlayıcı için madde bilgi fonksiyonunda diğer maddelerden daha fazla bilgi vermektedir.

Üç Parametrelili Lojistik Model (3PLM). 2PLM'ye maddeye şans ile doğru yanıt verme olasılığını gösteren şans başarısı parametresinin eklenmiş halidir. b parametresi -3 ile $+3$ arasında, a parametresi 0 ile 2 arasında, şans parametresi (c) ise 0 ile 1 arasında modellenir (DeMars, 2016).

Şans parametresinin teorik aralığı $0 \leq c \leq 1$ 'dir, ancak $0,35$ 'in üzerindeki değerler genellikle kabul edilebilir değerlerin üzerindedir (Baker, 2001). Bu model tahmin yoluyla doğru yanıt verme olasılığının olduğu çoktan seçmeli testlerde kullanılabilir. Dört olası cevap seçeneği olan çoktan seçmeli bir soruda tahmin yolu ile teorik olarak %25 veya 4'te 1 doğru cevap verme şansı vardır. Bu durumda, teorik olarak c 'nin değeri $0,25$ olur.

3PLM, tahmini genel başarı olasılığına katkıda bulunan bir faktör olarak değerlendirir, bu nedenle tahmin şansının olduğu testlerde MTK modeline şans parametresini de ekler. 3PLM'nin matematiksel gösterimi eşitlik 3'te verilmiştir:

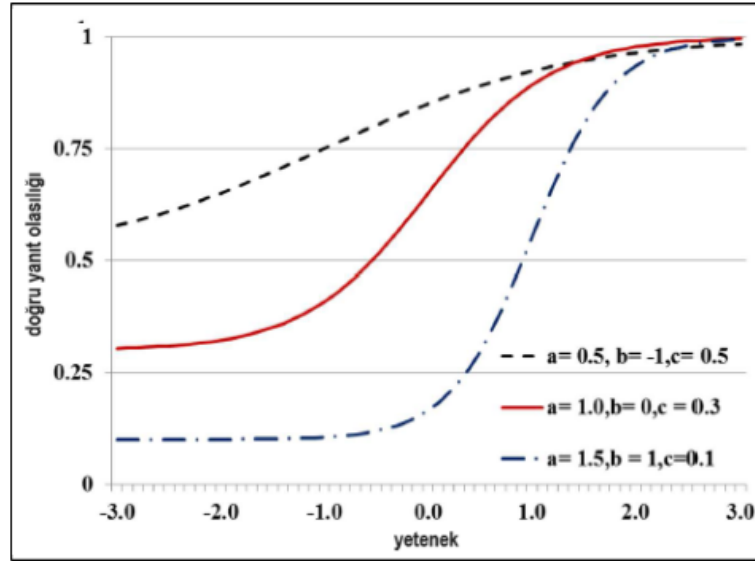
$$P(Y_{ij} | \theta) = c_j + (1 - c_j) \frac{1}{1 + e^{-D a_j (\theta_i - b_j)}} \quad (3)$$

θ_i yetenek düzeyini, a_j ayırdedicilik parametresini, b_j güçlük parametresini, c_j şans parametresini, e 2,718 sabitini, D ölçekleme faktörünü ($D=1.7$) temsil eder.

Şans parametresi, en düşük yetenek düzeyinde maddenin doğru yanıtlanma olasılığıdır. 3PLM'nin madde karakteristik eğrisi Şekil 3'te gösterilmiştir.

Şekil 3

3PLM için Madde Karakteristik Eğrisi

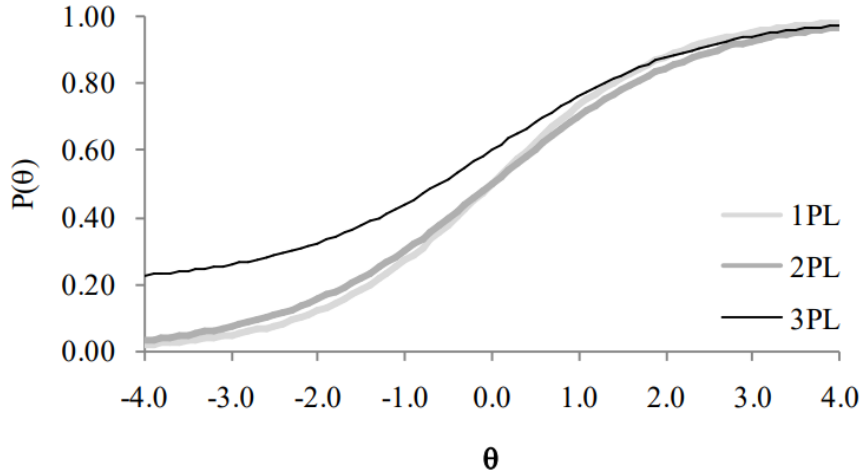


Şekil 3, yetenek düzeylerine göre farklı ayırt edicilik ve şans parametresine sahip üç maddenin doğru yanıtlanma olasılığını belirten üç farklı madde karakteristik eğrisini göstermektedir (1. madde ($a=0.5, b=-1, c=0.5$), 2. madde ($a=1, b=0, c=0.3$), 3. madde ($a=1.5, b=1, c=0.1$)). Şans parametresi en yüksek olan 1. maddeye bakıldığında en düşük yetenek seviyesinde maddenin doğru yanıtlanma olasılığının %50 olduğu görülmektedir.

Şekil 4'te 1PLM, 2PLM ve 3PLM'nin bir arada gösterildiği madde karakteristik eğrisi verilmiştir.

Şekil 4

Madde Karakteristik Eğrisi



Şekil 4'te 1-, 2- ve 3-PL modellerinin karşılaştırılması gösterilmiştir. Bu örnek maddede ayırt edicilik parametresi 1,00, zorluk parametresi 0,00 ve tahmin parametresi 0,20'ye eşittir. Şekilde görüldüğü gibi, 1PLM ve 2PLM'nin alt asimptotları 0,00'e eşittir. Ayrıca, ortalama yeteneğe sahip bir adayın ($\theta = 0,00$) bu öğeye doğru yanıt verme olasılığı 0,50'dir. 3-PL modelinin alt asimptotu, tahmini ayarlamak için c birim yukarı kaydırılmıştır. Bu nedenle, yetenek seviyesi 0,00'e eşit olan bir adayın bu öğeye başarılı bir şekilde yanıt verme olasılığı 0,60'tır ($0,50 + c/2$). 2PL ve 3-PL modellerinde a parametrenin bulunması, bu örnekteki maddenin aşırı yetenek seviyelerinden ziyade orta yetenek aralığında (-2,00 ile 2,00 arasında) adaylar arasında daha iyi ayırım yapmasına neden olur.

Genel olarak, modeldeki parametre sayısı 3'ten 1'e düştükçe, bir maddedeki başarı olasılığını hesaplamak için gereken bilgi miktarı da azalır. 3PLM üç modelin en ayrıntılı olduğu için, bu modelin şartlar uygun olduğunda üç model arasında en doğru model olabileceği söylenebilir. Ancak, maddede çok az tahmin şansı varsa, 3PLM'deki c parametresine gerek olmayabilir ve 2-PL modeli daha iyi bir seçim olabilir. Benzer şekilde, 1PL modeli de uygun

şartlar olduğunda 2PLM veya 3PLM kadar iyi bir başarı tahmini sağlayabilir. Bunun yanında 1PLM'nin basitliği, bazı durumlarda 2PLM veya 3PLM kullanırken sağlanan ekstra hassasiyetten daha önemli olabilir.

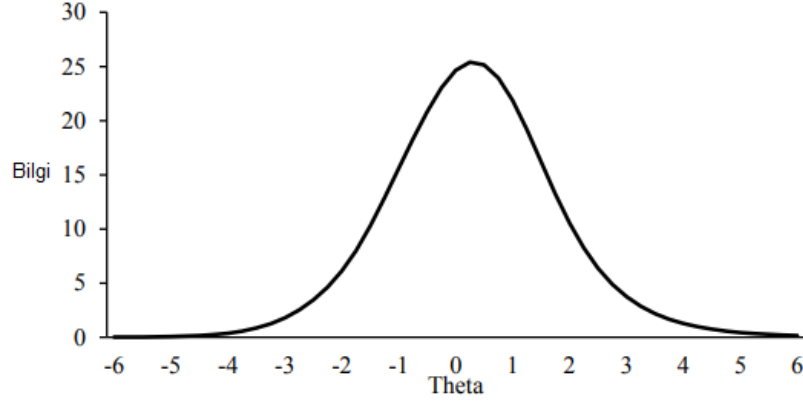
MKE, tek bir maddenin özelliklerini grafiksel olarak ifade etmek için kullanılırken, test karakteristik eğrileri (TKE), bir madde grubunun veya bir testin tamamının özelliklerini görüntülemek için kullanılır. TKE, θ 'ye bağlı olarak testteki her maddenin MKE'lerinin toplanmasıyla oluşturulur. MKE, maddenin θ ölçeği boyunca nasıl işlediğini göstermek için kullanılırken TKE, testin θ ölçeği boyunca nasıl işlediğini göstermek için kullanılır.

MTK yetenek tahminleri, testteki tüm maddelerin MKE'nin matematiksel ifadesi olan madde karakteristik fonksiyonları (MKF) üzerinden toplanarak test karakteristik fonksiyonu (TKF) aracılığıyla toplam puanlara dönüştürülür. Bu olasılıkların toplamı, adayın doğru cevaplama beklendiği testteki madde sayısını gösterir ve aynı zamanda bir adayın gerçek puanının MTK tanımı olarak da bilinir. Bu nedenle, gerçek puanlar ile bunlara karşılık gelen θ 'ler arasındaki tahmini ilişki, TKE aracılığıyla grafiksel olarak gösterilir.

Her θ seviyesiyle ilişkili puan yorumlamasını iyileştirmenin bir başka yolu da test bilgi fonksiyonudur (TBF). Madde bilgi fonksiyonu (MBF) her madde için ölçmenin standart hatasının tersi olarak hesaplanır ve θ 'ye bağlıdır. TBF ise madde bilgi fonksiyonlarının toplamı olarak alınır. Hata varyansı ne kadar büyük olursa, bir adayın yeteneğini tahmin etmek için o kadar az bilgi bulunur ve dolayısıyla puan yorumlamasında o kadar fazla hata olur. Şekil 5'te örnek bir TBF grafiği gösterilmiştir.

Şekil 5

Test Bilgi Fonksiyonu (TBF)



Şekil 5'te görülen TBF eğrisinin en tepe noktası, testin yeteneği tahmin etme konusunda θ ölçeğinde en başarılı olduğu yeri gösterir. TBF'ler genellikle, test oluşturma sırasında belirli bir kesme puanının yakınında veya belirli bir puan aralığında en yüksek bilgiyi sağlayan testleri geliştirmek için kullanılır.

MTK'nın Varsayımları

MTK, KTK'nın sınırlılıklarını aşarak KTK'dan daha güçlü tahminler geliştirir. MTK, bu sınırlılıkları aşarken güçlü bir matematiksel temele dayanır ve bilgisayarlar tarafından işlenebilen karmaşık algoritmalar içerir. Bu nedenle MTK'nın güvenilir sonuçlar verebilmesi için gerekli varsayımların sağlanması ve doğru matematiksel modellerin seçilmesi oldukça önemlidir. Uygulanan modelin model-veri uyumu, maddeleri kalibre etmek için kullanılan algoritmayı ve madde-yetenek parametre kestirimini dolayısıyla eşitleme sonuçlarının doğruluğunu doğrudan etkiler (Hambleton & Swaminathan, 1985; Zhao, 2008). MTK modellerinin avantajları, modelin verilen test verilerine uyduğu durumlarda elde edilebilir (Hambleton ve diğerleri, 1991).

MTK'nın temel varsayımları şunlardır:

Boyutluluk. Boyutluluk varsayımı, tek boyutlu veya çok boyutlu olabilen ve test performansının altında yatan gizli özellik sayısı ile ilgilidir. Tek boyutluluk, testteki tüm

maddelerin tek bir yeteneği ölçtüğü veya gizil bir özelliği ölçme anlamında homojen olduğu anlamına gelir. Çok boyutluluk ise katılımcının performansının altında yatan birden fazla potansiyel özelliği belirtir (Hambleton & Traub, 1973; Lord, 1968; Raykov & Marcoulides, 2018).

Monoton Artış. Bu varsayım, maddenin ölçtüğü yeteneğe göre maddenin başarı olasılığını ilişkilendiren ve doğrusal olmayan bir regresyon fonksiyonu olan Madde Karakteristik Eğrisi'ne (MKE) dayanmaktadır. Uygulamada, bir MKE'yi tanımlamak için gereken parametre sayısı belirli gizil özellik modeline bağlıdır ve bu modeller yaygın olarak bir, iki ve üç parametrelidir. Madde karakteristik eğrisi, katılımcıların özellik seviyeleri, madde özellikleri ve maddeyi doğru yanıtlama olasılığı arasındaki ilişkiyi modelleyen Madde Tepki Fonksiyonunun (MTF) grafiksel bir gösterimidir (Baker & Kim, 2004). Katılımcının yetenek seviyesi arttıkça o madde için doğru yanıt verme olasılığının da artması beklenmektedir. Düşük yetenek seviyesindeki katılımcının bir maddeye doğru yanıt verme olasılığını sıfıra yaklaştıkça, yüksek yetenekli katılımcının doğru yanıt verme olasılığının 1'e yaklaşacağı anlamına gelir.

Yerel Bağımsızlık. MTK'nın yerel bağımsızlık varsayımı, aynı yetenek düzeyindeki bireylerin bir test maddesine verdiği yanıtın diğerler yanıtlarından istatistiksel olarak bağımsız olduğunu belirtir. Bir maddeye verilen yanıt testteki diğer maddelere verilen yanıtı etkilememelidir. Bu varsayımın karşılanması için herhangi bir katılımcının yanıt örüntüsünün (1'ler ve 0'lar) olasılığı aynı yetenek düzeyindeki bireylerin her madde için elde edilen doğru cevaplama olasılıklarının çarpımıyla elde edilir; bu, test maddelerinin aynı yetenek düzeyindeki katılımcılar için ilişkisiz olduğu anlamına gelir (Lim, 2020; Lord & Novick, 1968).

MTK ile eşitleme yapılabilmesi için karşılanması gereken varsayımların en önemlilerinden biri yerel bağımsızlıktır. Yerel bağımsızlık, bir testteki yetenek değişkeninin sabit tutulduğu durumlarda, herhangi iki soruya verilen yanıtların istatistiksel olarak birbirlerinden bağımsız olmasını ifade eder (Lord, 1980). Bir başka deyişle, maddelerin doğru cevaplanma olasılıklarının, belirli bir yetenek düzeyinde o bireyin yeteneği dışında başka bir etkene bağlı olmamasıdır (Hambleton & Swaminathan, 1985).

Madde takımı içindeki maddelerin aynı paragrafa bağlı olması, maddelere verilen yanıtların madde takımındaki diğer maddelerden etkilenmesine ve dolayısıyla doğru cevaplanma olasılıklarının değişmesine yol açabilir. Bu durum, maddelerin doğru cevaplanma olasılıklarının, o bireyin yeteneği dışında başka bir etkene (madde takımı etkisi) bağlı olmasına ve MTK'nın yerel bağımsızlık varsayımının ihlaline neden olabilir.

Bishop ve Omar (2002), Lee vd. (2001), Marais ve Andrich (2008) ve Li vd. (2006) çalışmalarında, madde takımı etkisini göz ardı etmenin, eşitleme ve ölçekleme çalışmalarında hatalara ve madde uyumsuzluğuna yol açabileceğini vurgulamışlardır. Bu sebeple, madde çiftleri arasındaki yerel bağımlılığı göz önünde bulunduran modellerin tercih edilmesi, parametre kestirimlerinde, ölçek dönüştürme ve eşitleme çalışmalarında daha doğru sonuçlar elde etmeyi sağlar. Brennan (2001) tarafından belirtildiği üzere, madde takımları, madde tepki kuramının temel varsayımlarından biri olan yerel bağımsızlık varsayımını ihlal ettiği için önemli bir araştırma konusu haline gelmiştir.

Yerel bağımsızlık varsayımında madde çiftleri arasındaki kovaryans değerinin sıfır veya sıfıra yakın olduğunu kabul edilmektedir (Dresher, 2002). Madde takımları gibi aynı soru köküne bağlı, maddeler arasında yerel bağımlılık bulunan madde kümelerinde bu durum bozulabilmektedir.

Yerel madde bağımlılığının birçok nedeni ve türü vardır. Bunlardan en önemlilerinden biri paragraf bağımlılığıdır. Büyük ölçekli testlerin birçoğu, karmaşık ve birbiriyle ilişkili sorulara olanak tanıyan, testi zaman açısından daha verimli kılan aynı paragrafa bağlı birden fazla soru maddesinin oluşturduğu madde takımı adı verilen soru tipini içermektedir (DeMars, 2006). Büyük ölçekli testlerin bu içeriğinden dolayı, yerel bağımlılık üzerine yapılan çalışmalarda en çok paragraf bağımlılığına odaklanılmış, madde takımlarının kendi içerisinde birbiriyle olan ilişkilerinin ve doğru cevabı bulmaya etkilerinin yerel bağımlılığa neden olduğu belirlenmiştir (Koğar & Kelecioğlu, 2017). Dolayısıyla bu çalışma, test eşitleme sürecinde yerel madde bağımlılığına neden olan madde takımları üzerinde durmaktadır.

Madde takımları ikincil bir faktör olarak ele alındığı için yerel bağımlılığın etkisini inceleyen modeller çok boyutlu madde tepki kuramına göre oluşturulur.

Çok Boyutlu Madde Tepki Kuramı

MTK modelleri ve uygulamaları, genellikle tek boyutluluk varsayımına dayanır. Ancak, çoğu ölçüm, özellikle de çoğu psikolojik ölçüm, içerik açısından heterojen göstergelere sahip oldukları için kavramsal olarak geniştir. Örneğin, bir Türkçe testi hem dilbilgisi hem okuduğunu anlama yeteneği için maddeler içerebilir. Bir ölçüm ne kadar fazla içerik çeşitliliği göstergesi içerirse, çok boyutlu modeller için o kadar uygun hale gelir. Çoğu standart MTK modelinde tek boyutluluk varsayımı nedeniyle, içerik çeşitliliği MTK model seçimi ve standart MTK modellerinin uygun olup olmadığı konusunda doğrudan etkilere sahiptir (Reise, Morizot, & Hays, 2007).

Kişinin ve maddenin iki veya daha fazla boyut veya gizil özellikleri yansıttığı çok boyutlu MTK modellerine izin veren yöntemler geliştirilmiştir (Embretson & Reise, 2000; Reckase, 2009). Bu yöntemler, iki (veya daha fazla) farklı boyutun modellendiği yerlerde uygulanır. Ancak, birçok psikolojik ölçüm, göstergelerin hiyerarşik bir yapıyı yansıttığı durumlar içerir. Örneğin, Türkçe yeteneği ölçümü hem dilbilgisi hem de okuduğunu anlama için maddeler içerse de, bu boyutlar dil yeteneğinin daha geniş boyutundan bağımsız değildir. Bunlar, alt boyutlar olarak kabul edilebilir.

Çok boyutlu kavramları modellemek için genellikle hiyerarşik faktör modelleri kullanılmaktadır (Brown, 2006). Bu tür modellerde, faktörlerin katmanları daha geniş içerik alanının heterojenliğini temsil eder. Modeldeki gözlemlenen değişkenler, bu faktörlerden bir veya daha fazlasına yüklenir. Bu hiyerarşik modellerden biri iki faktör modelidir. İki faktör modelinde, maddelerin bir genel boyuta veya faktöre ve iki veya daha fazla alt boyuta veya grup faktörlerine ilişkin orantılılık kısıtlamaları vardır. Her madde, genel faktöre ve grup faktörlerinden birine yüklenmek üzere kısıtlanır. Bu, iki faktör modeli içinde yer alan ikinci dereceden modele zıt bir durum oluşturur.

İkinci dereceden modelde, birinci dereceden faktörler (veya grup faktörleri) belirli ölçüm alanlarını veya alt boyutları temsil ederken, bu birinci dereceden faktörlerin üzerinde yer alan ikinci dereceden faktör, bu alt boyutların genel bir özelliğini veya daha geniş bir kavramı temsil eder. Örneğin, bir eğitim araştırmasında öğrencilerin matematik, fen ve okuma becerileri gibi birinci dereceden faktörler, öğrencilerin akademik başarısının farklı yönlerini temsil edebilir. Bu birinci dereceden faktörler üzerinde yer alan ikinci dereceden faktör ise, genel akademik yeteneği temsil edebilir.

İki faktör modelde ise, her bir madde hem genel bir faktöre (örneğin, genel akademik yetenek) hem de bir veya daha fazla grup faktörüne (örneğin, matematik, fen ve okuma becerileri) yüklenir. Bu yapı, maddelerin genel bir özelliği ne ölçüde ve aynı zamanda belirli alt boyutları ne ölçüde yansıttığını ayrıntılı olarak inceleme imkanı sunar. İki faktör model, her bir madde için genel faktör ile grup faktörleri arasındaki ilişkiyi ayrı bir şekilde modelleyerek, ölçümün altında yatan çok boyutlu yapıyı daha net bir şekilde ortaya çıkarır. Bu sayede araştırmacılara veri setlerindeki hem genel hem de özgü boyutları aynı anda değerlendirme esnekliği sağlar. Bu, özellikle psikoloji, eğitim ve sosyal bilimler gibi alanlarda, bireylerin davranışlarını ve yeteneklerini çok boyutlu olarak değerlendirirken önemlidir. Bu modeller, ölçüm araçlarının geçerliliğini ve güvenilirliğini artırmaya, aynı zamanda ölçüm değişmezliği ve farklı gruplar arası karşılaştırmalar gibi önemli analizler yapmaya olanak tanır.

İki Faktör Modeli

Madde takımlarındaki, test maddeleri arasındaki bağımlılığı modellemek için kullanılan çözümlerden biri iki faktör modelidir (Rijmen, 2010). Hiyerarşik model (Markon, 2019) veya iç içe faktör modeli (Brunner ve diğerleri, 2012) olarak da bilinen iki faktör model (2FM), maddelerin bir yandan genel bir faktöre yük verdiği ve diğer yandan özel faktörlerin faktör yapısını gösterdiği bir gizil yapıdır. Genel faktör, maddeler arasındaki toplam varyansı açıklayarak maddelerin ortak yönlerini gösterir, özel faktör ise diğer sıradan faktör analitik yöntemleri gibi yorumlanır (Reise ve diğerleri, 2010).

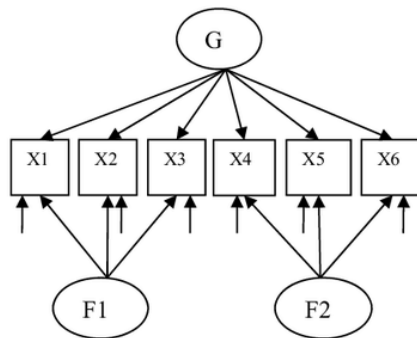
İki faktör modelin özellikleri, özellikle genel faktörde sıfırdan farklı yüklerin olması ve yalnızca bir grup faktöründe kısıtlamaların olması, onu hesaplama açısından cazip kılar (Gibbons & Hedeker, 1992). Hesaplama avantajlarının yanı sıra, Chen vd. (2006), ikinci dereceden modele göre iki faktör modelin avantajlarını şöyle açıklamaktadır. İlk olarak, iki faktör modellemesi, genel faktörün üzerinde grup faktörlerinin tahmin edici değeri hakkında çıkarımlar yapılmasına izin verir. İkinci olarak, iki faktör modeli, maddeler ile grup faktörleri arasındaki ilişkinin gücünün doğrudan incelenmesine olanak tanır. Üçüncü olarak, genel ve grup faktörleri için farklı gruplar arasında (örn. erkekler vs. kadınlar) ölçüm değişmezliği doğrudan test edilebilir.

Gibbons ve Hedeker (1992) tarafından çok boyutlu madde tepki kuramı modelinin özel bir hali olarak geliştirilen bu modelde, her maddenin yük değerine sahip olduğu bir genel faktör (birincil faktör) ve madde takımlarının yük değerine sahip olduğu ikincil bir faktör olarak grup-özellik faktörü (özel faktörler) vardır.

2FM'nin görselleştirilmiş ifadesi Şekil 6'da verilmiştir.

Şekil 6

İki Faktör Modelin Yapısı



Şekil 6 incelendiğinde her madde genel özelliğin (G) yanı sıra ikincil özelliğe de (F1 ve F2) yük vermektedir. Model her maddeyi genel faktöre birincil yük verecek şekilde ve özel faktörlerden ise sadece birine ikincil yük verecek şekilde sınırlandırır. Genel ve özel faktörlerin

tümünde madde faktör yüklerinin ölçeği serbest bırakılır. İki faktörlü ölçme yapısını gösteren örnek faktör örüntüsü Şekil 7’de verilmiştir.

Şekil 7

İki Faktör Modelin Örüntüsü

$$\begin{pmatrix} a_{10} & a_{11} & 0 \\ a_{20} & a_{11} & 0 \\ a_{30} & 0 & a_{32} \\ a_{40} & 0 & a_{42} \\ a_{50} & 0 & a_{52} \end{pmatrix}$$

Şekil 7’de görüldüğü gibi her madde genel boyuta yük vermekte, özel boyutlardan ise sadece birine yük vermektedir ayrıca genel ve özel faktörlerin tümünde madde faktör yüklerinin ölçeği serbest bırakılır. İki-faktör modelin matematiksel ifadesi eşitlik 4’te gösterilmiştir.

$$t_{ij} = a_{1i}\theta_j - \delta_i + a_{2i}\gamma_{d(i),j} \quad (4)$$

Eşitlikte yer alan yetenek θ_j ve rastgele madde takımı etkisi $\gamma_{d(i)}$ standart normal dağılıma sahiptir ve birbirinden bağımsızdır. a_{1i} ve a_{2i} sırasıyla j maddesi ile yetenek ve madde takımı etkisi arasındaki ilişkiyi belirleyen j maddesinin ayırt etme gücü parametreleri, δ_j ise güçlkle ilgili parametredir. Bu modelde, madde takımı etkisinin büyüklüğü MTTK’dan farklı olarak yorumlanır; burada büyüklük, testlet i parametresinin varyansının büyüklüğü $\gamma_{d(i)}$ yerine ayırt edicilik gücü (a_{2i}) ile ifade edilir. Başka bir deyişle, bir madde takımı içindeki maddeler arasında a_{2i} büyükse, madde takımı için olan ikincil boyut daha belirgindir.

Ölçme pratiği açısından bakıldığında, ek parametrelerin (örneğin, her madde için iki ayırt edicilik parametresi) dahil olması nedeniyle 2FM’de, tek boyutlu MTK modellerinden çok daha büyük bir örneklem boyutu gereklidir (de la Torre & Hong, 2010).

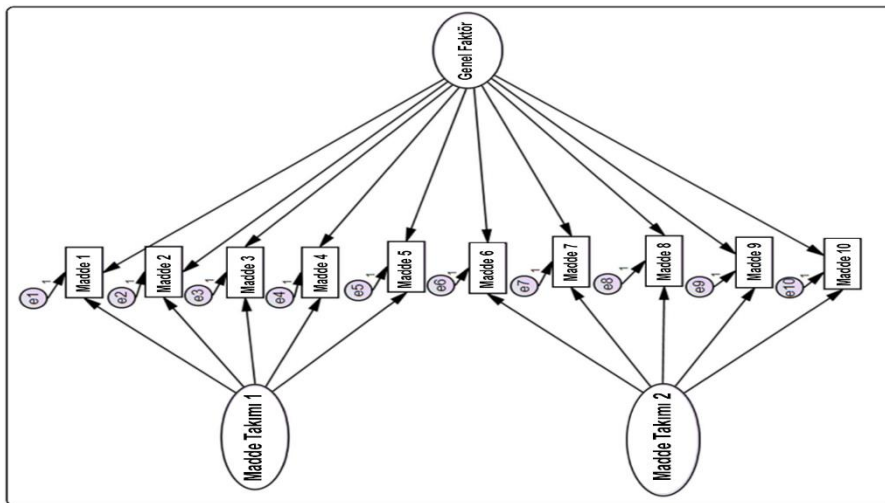
Madde Takımı Tepki Kuramı Modeli

MTK modellerinde yerel madde bağımsızlığı ortak bir varsayımdır. Bu varsayım, öğrenci yeteneklerinin, performansı açıklamak için gereken tüm bilgileri sağladığını ve performansı etkileyen faktörlerin rastgele olduğunu kabul eder. Ancak, madde takımı içeren testlerde MTK modelleri uygulandığında bu varsayım ihlal edilir çünkü belirtilen yetenek düzeyinde, madde çiftlerine verilen yanıtlar arasında bir korelasyon oluşur (Wainer ve diğerleri, 2000).

Madde takımı tepki kuramı (MTTK) yerel bağımlılık sorunuyla karşılaştığında, test maddeleri arasındaki bağımlılığı modellemek için bir çözüm üretir (Baldwin, 2007). Wainer vd. (2000), Birnbaum'un (1968) üç parametrelili lojistik (3PL) MTK modelini genişleterek, modele madde takımı etkisini eklemiş ve 3PL MTTK modelini geliştirmiştir. MTTK, MTK modellerinden iki faktör modeli (2FM) temel alır ve yerel bağımlılığı oluşturan madde takımlarını ölçülmek istenen birincil boyutun dışında ikincil bir boyut olarak ele alır. Bu amaçla geleneksel MTK parametrelerine madde takımındaki madde çiftlerinin paylaştığı varyansın etkisini gösteren rastgele etki parametresi eklenir (Bao, 2007; Rijmen, 2009; Wainer ve diğerleri, 2007). Şekil 8'de örnek bir varsayımsal madde takımı modeli için bir yol diyagramı gösterilmiştir.

Şekil 8

Madde Takımı Tepki Kuramı Modelinin Yapısı



Şekil 8'de görüldüğü üzere MTTK'da da 2FM'de olduğu gibi her madde hem özel faktöre hem genel faktöre yük verir. İki model arasındaki tek fark MTTK'da özel faktördeki maddelerin faktör yük ölçeklerinin genel faktöre göre sabitlenmesidir. MTTK aslında özel faktörlerdeki faktör yüklerinin, her madde takımı içindeki genel faktöre olan yüklerle orantılı olacak şekilde kısıtlandığı doğrulayıcı bir 2FM modelidir (Li ve diğerleri, 2006). MTTK ölçme yapısını gösteren örnek faktör örüntüsü Şekil 9'da verilmiştir.

Şekil 9

Madde Takımı Tepki Kuramı Modelinin Örüntüsü

$$\begin{pmatrix} a_{10} & a_{10} & 0 \\ a_{20} & a_{20} & 0 \\ a_{30} & 0 & a_{30} \\ a_{40} & 0 & a_{40} \\ a_{50} & 0 & a_{50} \end{pmatrix}$$

Şekil 9'da görüldüğü gibi her madde genel boyuta yük vermekte, özel boyutlardan ise sadece birine yük vermektedir. Ayrıca özel faktörlerin tümünde madde faktör yüklerinin ölçeği genel boyut yükü ile sınırlandırılır.

Geleneksel 2PL MTK denklemi Eşitlik (5)'te, 2PL MTK'ya madde takımı etkisinin $\gamma_{d(i)}$ eklendiği 2PL MTTK denklemi Eşitlik (6)'da gösterilmiştir.

$$P_i(\theta) = \frac{e^{(a_i (\theta - b_i))}}{1 + e^{(a_i (\theta - b_i))}} \quad (5)$$

$$P_i(\theta) = \frac{e^{(a_i (\theta - b_i - \gamma_{d(i)}))}}{1 + e^{(a_i (\theta - b_i - \gamma_{d(i)}))}} \quad (6)$$

Eşitliklerde P, j kişinin i maddesine doğru yanıt verme olasılığını; a_i , ayırt edicilik parametresini; θ , j kişinin yetenek düzeyini; b_i , madde güçlüğü ve $\gamma_{d(i)}$, i maddesinin ait

olduđu $d_{(i)}$ madde takımının j kişisi için rastgele madde takımı etkisini göstermektedir. $\gamma_{d(i)}$ için varyans arttıkça d madde takımı için daha büyük bir etki gözlemlenir (Bradlow ve diđerleri, 1999).

Test Eşitleme

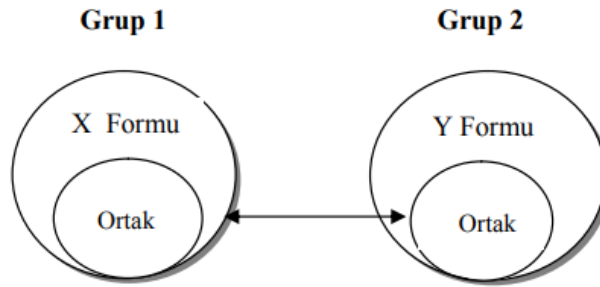
Geniş ölçekli testlerde kitapçıklar benzer yapılara sahip olsa da, kitapçıklardan elde edilen test puanları aynı ölçekte olmayabilir. Farklı kitapçıklardan elde edilen ve farklı ölçeklerde yer alan bu puanların karşılaştırılabilmesi için puanların aynı ölçeğe yerleştirilmesi, bir başka ifadeyle eşitlenmesi gerekir (Gübeş & Keleciođlu, 2015)

Test eşitleme, eğitim ve psikolojik deđerlendirmelerde adil ve dođru karşılaştırmalar yapılmasını sağlar. Farklı test formlarının kullanılması durumunda, eşitleme işlemi, puanların zaman içinde veya farklı koşullar altında karşılaştırılabilirliğini artırır. Bu, özellikle uzun süreli eğitim programlarının deđerlendirilmesi, sertifika sınavlarının standardizasyonu ve geniş çaplı akademik araştırmalar için kritik öneme sahiptir.

Kolen ve Brennan'a (2014) göre eşitleme sürecinde ilk aşama veri toplamada veya veri üretiminde kullanılacak olan eşitleme desenini belirlemektir. Tek grup deseni, eşdeđer gruplar deseni ve eşdeđer olmayan gruplarda ortak madde deseni en yaygın kullanılan eşitleme desenleridir. Bu araştırma TIMSS 2019 verileri üzerinden gerçekleştirildiđi için gruplar arasındaki farklılığın, her kitapçıkta yer alan ortak maddelerle kontrol edildiđi ve test kitapçıklarının bu ortak maddeler aracılıđı ile birbirine eşitlendiđi eşdeđer olmayan gruplarda ortak madde deseni kullanılmıştır. Eşdeđer olmayan gruplar ortak madde deseni Şekil 10'da gösterilmiştir.

Şekil 10

Eşdeğer olmayan gruplar ortak madde deseni



Şekil 10'da görüldüğü gibi iki test formu (X ve Y) ortak maddeler üzerinden eşitlenir. Eşitlemede veriye uygun desen seçildikten sonra, eşitleme yöntemine karar verilir. Eşitleme yöntemi, test eşitlerken uygulanan süreçlere denir (Öztürk, 2010). Bu süreçte önce uygun modele göre madde parametreleri kestirilir. Modellere göre parametreler kestirildikten sonra ölçek dönüştürme ve test eşitleme işlemleri yapılır. Bu çalışmada parametre kestiriminde KTK'ya dayalı yöntemler madde bazında daha az bilgi ve daha fazla hata içerdiği için (Hambleton, 1985), KTK yerine MTK modellerinden TBMTK, 2FM ve MTTK'ya dayalı modeller kullanılmış ve test eşitleme sürecinde ölçek dönüştürme işlemlerine odaklanılmıştır.

Ölçek Dönüştürme Yöntemleri

Kolen ve Brennan'a (2014) göre madde ve yetenek parametreleri uygun modellerle kestirildikten sonra eğer eşdeğer olmayan gruplarda ortak madde deseni kullanılacaksa kestirilen parametrelerin karşılaştırılabilir olması için önce ortak bir ölçeğe yerleştirilmesi gerekmektedir. Parametreler farklı örneklemlerden elde edildiği için farklı ölçekler üzerindedir. Kestirilen parametrelerin doğrusal bir dönüştürme işlemi yapılarak aynı ölçek üzerine yerleştirildiği bu sürece "kalibrasyon" adı verilir. Kalibrasyon sürecinde, ayrı kalibrasyon ve eş zamanlı kalibrasyon olarak iki farklı kalibrasyon yöntemi bulunmaktadır. Bu çalışmada çok boyutlu verilerde ayrı kalibrasyon yöntemi daha doğru ve güvenilir sonuçlar verdiği için (Kolen ve Brennan, 2014) ayrı kalibrasyon yöntemi kullanılmıştır.

Ayrı kalibrasyon yönteminde, kestirilen madde parametreleri test eşitleme yapılmadan önce aynı ölçeğe yerleştirilmelidir. Ölçek dönüşümünde yeni formdan kestirilen madde ve yetenek parametreleri, ortak maddelerin eski formdaki parametreleri aracılığıyla, eski formun ölçeğine dönüştürülür.

Dönüşüm için öncelikle eğim (A) ve kesişim (B) sabitleri elde edilir (Eşitlik 7).

$$a_{Ei} = \frac{a_{Yi}}{A}$$

$$b_{Ei} = Ab_{Yi} + B \quad (7)$$

a_{Ei} : i maddesinin E testinden (eski formdan) elde edilen ayırt edicilik parametresi

b_{Ei} : i maddesinin E testinden (eski formdan) elde edilen güçlük parametresi

a_{Yi} : i maddesinin Y testinden (yeni formdan) elde edilen yeniden ölçeklenmiş ayırt edicilik parametresi

b_{Yi} : i maddesinin Y testinden (yeni formdan) elde edilen yeniden ölçeklenmiş güçlük parametresi

Eşitlik (7)'den elde edilen A ve B sabitleri kullanılarak Y testindeki j kişisine ait yetenek değerinin E testindeki karşılığı Eşitlik (8)'deki gibi hesaplanır.

$$\theta_{Ej} = A\theta_{Yj} + B \quad (8)$$

Ayrı kalibrasyon yapılırken ölçek dönüşümü için kullanılan yöntemler moment yöntemleri (ortalama-ortalama (OO), ortalama-standart sapma (OS)) ve karakteristik eğri yöntemleri (Stocking-Lord (SL), Haebara (HA)) diye ikiye ayrılır. Bu yöntemlerin birbirinden farklı ölçek dönüştürme için gerekli olan A ve B katsayılarını elde etme yöntemlerinin farklılığıdır (Embretson & Reise, 2000).

Ortalama Ortalama Yöntemi (OO). Loyd ve Hoover'ın (1980) tanıttığı bir moment ölçek dönüştürme yöntemidir. Bu yöntem madde ayırt ediciliklerini ve madde güçlüklerini dikkate alır ve A ve B ölçekleme sabitlerini tahmin etmek için Eşitlik (9)'u kullanır:

$$A_{MM} = \frac{M(a_N)}{M(a_O)}$$

$$B_{MM} = M(b_O) - A_{MM}M(b_N) \quad (9)$$

$M(\cdot)$: aritmetik ortalama

a_N, b_N : eski ölçekteki ayırt edicilik ve güçlük parametreleri

a_O, b_O : ayırt edicilik ve güçlük parametrelerinin yeni ölçekteki karşılıkları

Ortalama Sigma yöntemi (MS). Marco (1977) tarafından tanıtılmış bir ölçek dönüştürme yöntemidir. Bu yöntem madde güçlüklerini dikkate alır ve A ve B ölçekleme sabitlerini tahmin etmek için Eşitlik (10)'u kullanır:

$$A_{MS} = \frac{SD(b_O)}{SD(b_N)}$$

$$B_{MS} = M(b_O) - A_{MS}M(b_N) \quad (10)$$

$SD(\cdot)$: standart sapma

b_N : eski ölçekteki güçlük parametresi

b_O : güçlük parametresinin yeni ölçekteki karşılığı

Stocking & Lord yöntemi (SL). Stocking ve Lord (1983) tarafından tanıtılmış madde karakteristik eğrisine dayalı bir ölçek dönüştürme yöntemidir. Bu yöntem belirli bir θ düzeyinde her bir madde için madde karakteristik eğrileri arasındaki farkların toplamının karesini alarak karakteristik eğriler arasındaki farkı hesaplar. Belirli bir θ_i için ortak maddeler ($j:V$) üzerindeki

karesel farkların toplamı, Kolen ve Brennan (2004) tarafından belirtilen Eşitlik (11) kullanılarak gösterilebilir.

$$SLdiff(\theta_i) = \left[\sum_{j \in V} p'_{ij}(\theta_{ji}, a_{jj}, b_{jj}, c_{jj}) - p_{ij}\left(\theta_{ji}, \frac{a_{ij}}{A}, Ab_{ij} + B, c_{ij}\right) \right]^2$$

$$SL_{critt} = \sum_i SLdiff(\theta_i) \quad (11)$$

A : Eğim sabiti

B : Kesişim sabiti

P_{ij} (-): Madde karakteristik fonksiyonu

P'_{ij} (-): Dönüştürülmüş madde karakteristik fonksiyonu

Haebara yöntemi (HA). Habera (1980) tarafından tanıtılmış madde karakteristik eğrisine dayalı bir ölçek dönüştürme yöntemidir. Bu yöntemde belirli bir θ düzeyinde her madde karakteristik eğrisi için madde karakteristik fonksiyonu ile dönüştürülmüş madde karakteristik fonksiyonu arasındaki farkın karesi toplanarak karakteristik eğriler arasındaki fark hesaplanır (Eşitlik 12).

$$Hdiff(\theta_i) = \sum_{j \in V} \left[p'_{ij}(\theta_{ji}, a_{jj}, b_{jj}, c_{jj}) - p_{ij}\left(\theta_{ji}, \frac{a_{ij}}{A}, Ab_{ij} + B, c_{ij}\right) \right]^2$$

$$HL_{critt} = \sum_i Hdiff(\theta_i) \quad (12)$$

A : Eğim sabiti

B : Kesişim sabiti

P_{ij} (-): Madde karakteristik fonksiyonu

P'_{ij} (-): Dönüştürülmüş madde karakteristik fonksiyonu

İlgili Araştırmalar

Bu bölümde literatürde MTTK, 2FM ve MTTK modelleri kullanılarak yapılan ölçek dönüştürme/eşitleme çalışmaları incelenecektir.

Davey vd. (1996) simülasyon verileriyle yaptığı çalışmada çok boyutlu verilerde çok boyutlu madde tepki kuramına dayalı modelin, tek boyutlu madde takımı modelinden daha iyi parametre tahmini yaptığı bulgusuna ulaşmışlardır. Araştırmacılar tek boyutlu madde tepki kuramı çalışmalarında adayın performansının altında yatan ve onu etkileyen tek bir özelliğin olduğu varsayımını basit bir kolaylık olarak gördüklerini belirtmiş ve bunun madde yanıtlarının aslında birden fazla özelliğin fonksiyonu olduğu gerçeğiyle çeliştiğini vurgulamıştır. Ayrıca uygun koşullarda pratik olduğu için tek boyutlu modellerin kullanılabilirliğini ama uygun olmayan koşullarda daha gerçekçi modeller mevcut olmadığı veya uygulanması zahmetli olduğu için tek boyutlu modellerin kullanılmasının uygun olmayacağını belirtmişlerdir.

Bolt (1999) çalışmasında MTK gerçek puan eşitleme yönteminin performansını, testin çok boyutluluğu koşulu altında simülasyon verileriyle incelemiştir. Çalışmada, madde yanıt matrisinin tek boyutlu olduğunu açıkça varsayan MTK gerçek puan eşitleme yönteminin, çok boyutluluğun varlığında iki geleneksel eşitleme yöntemlerine göre (doğrusal ve eşit yüzdellikli eşitleme) olumsuz etkilenip etkilenmediği araştırılmıştır. İki simülasyon çalışmasından elde edilen sonuçlara göre, MTK yönteminin, boyutlar arası korelasyon yüksek olduğunda (≥ 0.7) geleneksel yöntemler kadar iyi performans gösterdiğini, korelasyon orta düzeyde veya düşük olduğunda ise (≤ 0.5) eşit yüzdellikli yöntemle kıyasla daha kötü performans gösterdiği bulunmuştur.

Lee vd. (2001), üç farklı alanda (Okuma, Harita ve Matematik) gerçek veriler kullanarak, nominal model, aşamalı tepki modeli (ATM) ve 3PL MTK modeli kullanarak madde takımı tabanlı testlerin eşitleme sonuçlarını karşılaştırmıştır. Sonuçlar üç temel eşitleme yöntemi (ortalama, doğrusal ve eşit yüzdellikli yöntemler) kullanılarak karşılaştırılmış, çok kategorili MTK modellerinin eşitlemede iki kategorili modelden daha iyi sonuçlar verdiğini bulmuşlardır.

Bu sonuç, MTK varsayımlarının daha az ihlal edildiği Matematik testine kıyasla Okuma ve Harita testlerinde daha belirgin görülmüştür.

Zhang (2007) madde takımı tabanlı testlerin MTK eşitlemesinde iki kategorili ve çok kategorili MTK modellerini karşılaştırmıştır. Zhang (2007), hafif ya da orta düzeyde çok boyutluluğun söz konusu olduğu madde takımı tabanlı testlerin eşitlenmesinde çok kategorili MTK modelinin iki kategorili 3PL MTK modelinden daha iyi performans gösterdiğini bulmuştur. Bununla birlikte, Lee vd.'nin (2001) sonuçlarının aksine, Zhang, madde takımı tabanlı testler yüksek düzeyde çok boyutluluk sergilediğinde, 3PL MTK modelinin MTK eşitlemesinde geliştirilmiş kısmi kredi modelinden daha iyi performans gösterdiğini bulmuştur. Sonuçlardaki bu tutarsızlık, iki çalışma için kullanılan testlerdeki madde takımlarının özelliklerine daha yakından bakıldığında sorunlu görünmemektedir. Zhang (2007), madde takımı tabanlı testlerin MTK eşitlemesinde 3PL MTK modelini geliştirilmiş kısmi kredi modeliyle karşılaştırmakla kalmamış, aynı zamanda iki modeli kullanarak eşitleme sonuçlarındaki farklılıklara neden olan koşulları da (boyutluluk özellikleri ve yerel madde bağımlılığı örüntüleri gibi) da incelemiştir. Zhang (2007), DIMTEST sonuçlarına ve yerel madde bağımlılığı (LID) endeks değerlerine dayanarak, hem çok boyutluluğun hem de LID'nin işaretlerinin ve büyüklüklerinin madde takımlarında eşitleme sonuçlarını etkileyebileceğini bulmuştur. Çalışmada kullanılan çok boyutlu veriler için 3PL MTK eşitleme sonuçları eşit yüzdelikli eşitleme sonuçlarına daha yakın bulunsa da, madde takımları arası LID'nin işaretleri ve nispeten yüksek büyüklükleri ve madde takımları içi LID'nin düşük büyüklükleri madde takımlarının testteki boyutsal yapıyı yeterince yansıtmadığını göstermiştir. Çok kategorili kısmi kredi modeli, görece yüksek madde takımları arası LID'yi dikkate alamadığı için 3PL MTK modeli daha az hatalı sonuçlar vermiş olabilir.

Huiqin vd. (2008) yaptığı araştırmada MTK tabanlı eşitleme sonuçları üzerinde tutarsız b-parametrelerine sahip uç değer ortak maddelerle daha az hatalı sonuçlar elde etmek amacıyla, uç değerler göz ardı edildiğinde veya dikkate alındığında dört MTK tabanlı eşitleme yönteminin (eşzamanlı kalibrasyon, test karakteristik eğrisi [TCC] ve ortalama/sigma [M/S])

dönüşümleri ile ayrı kalibrasyon ve sabit ortak madde parametreleri [FCIP] ile kalibrasyon) hatalarını incelemiştir. Eşdeğer olmayan gruplar ortak madde deseni için simüle edilen veriler, grup yetenek farkları, uç değerlerin sayısı/değeri ve uç değer türleri açısından incelenmiştir. Ortak maddelerde uç değerler bulunmadığında, TCC ve M/S dönüşümlerinin en iyi performansı gösterdiği, uç değerler bulunduğunda, genel olarak, uç değerleri dikkate alan yöntemlerin (ağırlıklandırılmış uç değer M/S dönüşümü hariç) onları göz ardı eden yöntemlere kıyasla daha iyi sonuçlar verdiği bulunmuştur.

Zhang (2010) yaptığı çalışmada ortak madde seti ile toplam test arasındaki farklılıkların, diğer koşullar eşit tutulduğunda, eşitleme sonuçlarını önemli ölçüde etkilemediğini belirtmiştir. Bunun dışında hem küçük örneklem büyüklüğü hem de ortak madde sayısının azlığı eşitleme hatalarının biraz daha fazla olmasına yol açtığını raporlamıştır. Daha sonra, karakteristik eğri yöntemlerinin ve eşzamanlı kalibrasyon yöntemlerinin, farklı koşullar altında, ikili puanlanan madde takımı tabanlı maddelerden oluşan alternatif test formlarındaki madde parametrelerini ilişkilendirmedeki etkililiği karşılaştırılmıştır. Yeni geliştirilen madde karakteristik eğrisi yöntemi ve madde takımı karakteristik eğrisi yönteminin Stocking-Lord test karakteristik eğrisi yöntemi ve eşzamanlı kalibrasyon yöntemine benzer ve hatta daha iyi performans gösterdiği görülmüştür. Model kalibrasyonunda yerel bağımlılığın göz ardı edilmesi eşitleme hatalarını önemli ölçüde artırmıştır ve ortak testler için daha büyük test varyansları daha büyük eşitleme hatalarına yol açmıştır. Ortak madde takımları için daha küçük madde takımı varyansları daha doğru eşitleme sonuçları vermiştir.

Tao vd. (2011), 3PL MTK modeli, aşamalı tepki modeli (ATM) ve MTTK modelinin madde takımı tabanlı testlerin gözlenen ve gerçek puan eşitlemesindeki performansını karşılaştırmak için simülasyon çalışmaları yürütmüştür. Çalışmanın yazarları, MTK gözlenen ve gerçek puan eşitleme yöntemlerini MTTK modeline genişletmiş ve madde takımı etki faktörünün standart sapmasını kontrol ederek üç LID seviyesini (sıfır, düşük-orta ve yüksek) incelemiştir. Çalışmada MTTK modelinde doğru yanıt olasılığının hem birincil özellik hem de madde takımına özgü özellik tarafından etkilendiği belirtilmiştir. Çalışmada ayrıca MTTK

modeli uygulanırken MTK gerçek puan eşitlemesini gerçekleştirmek için iki formun yalnızca birincil özellik aracılığıyla eşitlenmesini ve madde takımına özgü özelliğin dışarıda bırakılması önerilmiştir.

Zhang (2012) simülasyon verilerine dayanarak yaptığı yaptığı çalışmada tek boyutlu ve çok boyutlu modelleri marjinal maksimum olabilirlik yöntemini kullanarak karşılaştırmıştır. Çalışmada ortak maksimum olabilirlik yöntemi kullanıldığında tek boyutlu ve çok boyutlu modellerinin parametre tahmininde eşdeğer sonuçlar verdiğini, marjinal maksimum olabilirlik yöntemi uygulandığında ise bu iki modelin farklılaştığını belirtmektedir. Simülasyon sonuçlarına göre çok boyutlu modelde madde sayısı az olduğunda madde parametrelerinin daha doğru tahmin edildiği, buna karşın tek boyutlu modelde her alt testteki madde sayısı yeterince büyük olduğunda madde parametrelerinin daha doğru tahmin edildiği bulgusu elde edilmiştir. Araştırmanın sonucuna göre testin basit bir yapıya sahip olmadığı ve kalibrasyonda basit yapılı olarak belirlendiği durumlarda, modellerin yanlış tahmin edileceğini ve yetenekler arasındaki korelasyon katsayılarının olduğundan yüksek tahmin edileceğini belirtmiştir.

Brossman ve Lee (2013) yaptıkları çalışmada gözlenen puan ve gerçek puan eşitleme yöntemlerini çok boyutlu madde tepki kuramı (ÇBMTK) çerçevesinde incelemiştir. Çalışmada ikisi gözlenen puan, biri gerçek puan olmak üzere üç eşitleme yöntemi oluşturulmuştur. TBMTK gözlenen ve gerçek puan eşitlemesi, ÇBMTK gözlenen puan ve gerçek puan eşitlemesi ve eşit yüzdelli eşitleme yapmak için üç ayrı test kullanılmıştır. Çalışmanın sonucunda, ÇBMTK kullanılarak uygulanan eşitleme yöntemlerinin, TBMTK kullanılarak uygulanan eşitleme yöntemlerine kıyasla eşit yüzdelli eşitleme yöntemine daha benzer şekilde performans gösterdiği bulunmuştur. Araştırmacılar bu durumun TBMTK eşitleme yöntemlerinin MTK'nın tek boyutluluk varsayımını ihlal etmesinden kaynaklandığını belirtmişlerdir.

Chen (2014) 3PL MTK modeli, aşamalı tepki modeli (ATM), MTTK modeli ve 2FM olmak üzere dört farklı MTK modelinin, madde takımı tabanlı testlerin MTK gerçek ve gözlenen puan eşitlemesindeki performansını madde takımı uzunluğu ve LID seviyeleri koşullarına göre

karşılaştırmıştır. Çalışmada, hem madde takımı uzunluğunun hem de LID seviyesinin, madde takımı tabanlı testlerin MTK gerçek ve gözlenen puan eşitlemesinde incelenen modellerin performansını etkilediği bulunmuştur. Uzun madde takımlarına sahip testler için geleneksel 3PL MTK modeli kullanıldığında, daha yüksek yerel madde bağımlılığı seviyeleri, temel yöntemden elde edilenlerden daha fazla eşitleme hatasına yol açmıştır. Ancak, yerel bağımlılığının MTK eşitleme sonuçları üzerindeki etkisi kısa madde takımlarından oluşan testler için belirgin bulunmamıştır. Ayrıca, uzun madde takımlarından oluşan (10 veya daha fazla) ve çok düşük düzeyde yerel madde bağımlılığına sahip testler ve kısa madde takımlarından oluşan (5 civarında) testler için, incelenen dört MTK modeli de MTK gerçek ve gözlenen puan eşitlemesinde iyi çalışmıştır. Uzun madde takımlarına ve nispeten yüksek düzeyde yerel madde bağımlılığına sahip testler için GRM, 2FM ve MTTK modelleri, madde takımı tabanlı testlerin MTK gerçek ve gözlenen eşitlemesinde geleneksel 3PL MTK modelinden daha iyi performans göstermiştir.

Cao vd. (2014), ayrı kalibrasyon yöntemleri kullanarak rastgele gruplar deseni altında gerçek puan eşitleme konusunda üç MTK modelinin (2PL MTK modeli, GRM ve 2PL MTTK modeli) performansını karşılaştırmak için bir simülasyon çalışması yürütmüştür. Sonuçlar, LID'nin varlığında, 2PL MTK modelinden ve 2PL MTTK modelinden elde edilen sonuçların, GRM'den elde edilen sonuçlara kıyasla, temel eşit yüzdelikli yöntemden elde edilen sonuçlara daha yakın olduğunu göstermektedir. Buna ek olarak, 2PL MTK modeli yerel madde bağımsızlığı varsayımının ihlaline karşı oldukça dayanıklı bulunmuştur.

Xu (2016) araştırmasında, yerel bağımlılığın var olduğu ortak madde test deseni altında test eşitlemede eşzamanlı kalibrasyon modelleri ile çok düzeyli modellerin etkinliğini araştırmıştır. Çok düzeyli modellerin performansı, iki simülasyon çalışması aracılığıyla geleneksel MTK modelleri ve madde takımı tepki modeli (MTTK) ile karşılaştırılmıştır. İlk çalışmada yerel madde bağımlılığı (LID), ikinci çalışmada ise hem LID hem de kişi bağımlılığı göz önünde bulundurulmuştur. İlk çalışmada, madde takımı tabanlı testlerin eşitlenmesinde dört eşzamanlı kalibrasyon yaklaşımının performansı karşılaştırılmıştır: Araştırmada madde

takımı puanlaması kullanılmış ve aşamalı yanıt modeli (GRM) uygulanmıştır. Sonuçlar, iki düzeyli HGLM ve Rasch yaklaşımlarının, beklenen puan iyileşmesi açısından yerel madde bağımsızlığı varsayımının ihlaline karşı sağlam olduğunu göstermiştir. Ayrıca, ilk üç yaklaşım GRM kullanılarak yapılan eş zamanlı kalibrasyondan daha iyi eşitleme sonuçları sağlamıştır. İkinci çalışmada, üç modelin (yani 3PL MTK modeli, 3PL MTTK modeli ve 3PL çok düzeyli MTTK modeli) eşzamanlı kalibrasyon modelleri kullanılarak, madde takımı tabanlı testlerin eşitlenmesindeki performansı karşılaştırılmıştır. Sonuçlar, LID'nin göz ardı edilmesinin madde parametresi geri kazanımını etkilediğini göstermiştir. Hem LID hem de kişi bağımlılığının varlığında, 3PL çok düzeyli MTTK modeli, özellikle yüksek derecede kişi bağımlılığı ile kişi parametreleri için en doğru tahmini sağlamıştır.

Gübeş ve Kelecioğlu (2016) karışık-formatlı testlerin eşitlemesinde çok boyutluluğun, ortak madde yapısının ve ölçek dönüştürme yöntemlerinin eşitlik özelliğinin korunmasına olan etkisini incelemiştir. Çalışma TIMSS 2011 8. sınıf matematik değerlendirmesinden elde edilen gerçek madde parametre tahminlerine dayanarak üretilen simülasyon verileriyle gerçekleştirilmiştir. Çalışmada MTK gerçek puan eşitlemesi (TSE) ve MTK gözlenen puan eşitlemesi (OSE) yöntemleri, eşdeğer olmayan gruplar ortak-madde deseni altında birinci derece eşitlik (FOE) ve ikinci derece eşitlik (SOE) özellikleri temelinde değerlendirilmiştir. Çalışmanın sonuçlarına göre FOE ve SOE özelliklerinin, tek boyutlu durum altında en iyi şekilde korunduğu, çok boyutluluğun derecesi ciddi olduğunda ise korunamadığı bulgusuna ulaşılmıştır. Ayrıca tek boyutlu ve çok boyutlu test yapısı altında, karakteristik eğri yöntemlerinin, FOE ve SOE özelliklerini koruma açısından moment yöntemlerinden önemli ölçüde daha iyi performans gösterdiğini belirtmişlerdir.

Atar ve Yeşiltaş (2017) çalışmalarında çok boyutlu veri kullanarak Stocking-Lord, ortalama/ortalama ve ortalama/sigma ölçek dönüştürme yöntemlerinin hatalarını örneklem büyüklüğü, yetenek dağılımı, boyutlar arası korelasyon ve testteki ortak madde oranı koşulları altında incelemiştir. Çalışmada madde parametrelerine Stocking-Lord yöntemiyle ölçek

dönüştürme işlemi uygulandığında çoğu koşulda diğer iki yönteme göre daha küçük RMSE ve yanlılık değerleri elde edildiği belirtilmiştir.

Bökeoğlu vd. (2022) yaptıkları çalışmada MTK'ya dayalı gerçek puan eşitlemede, ölçek dönüştürme yöntemlerinin eşitleme hatalarının karşılaştırmıştır Çalışmada eşdeğer olmayan gruplarda ortak madde desenine uygun üretilen simülasyon verileri kullanılmış, eşitleme hataları (RMSD) ölçütüne göre değerlendirilmiştir. Araştırmada, SL yönteminden elde edilen RMSD değerlerinin, tüm koşullarda diğer yöntemlere göre daha yüksek bulunduğu belirtilmiştir. Bunun dışında, 2PLM ve 3PLM'nin benzer sonuçlar verdiği, örneklem büyüklüğü ve test uzunluğu arttığında, ortak madde oranı %40 ve grupların yetenek dağılımının benzer olduğu durumlarda daha düşük hatalar elde edildiği raporlanmıştır.

Bölüm 3

Yöntem

Bu bölümde araştırmanın türü, evren ve örneklem, veri toplama aracı ve verilerin analizi başlıkları altında bilgiler verilmiştir.

Araştırmanın Türü

Bu çalışma, madde takımı tabanlı testlerde tek boyutlu madde tepki kuramı, iki faktör ve madde takımı tepki kuramı modellerini kullanarak ölçek dönüştürmeyi ve elde edilen ölçek dönüştürme hatalarını çeşitli koşullar altında incelemeyi amaçlamaktadır.

Araştırma, gerçek veriler üzerinde bir kuramı test etmeyi ve model karşılaştırmayı amaçladığı için betimsel araştırma niteliğindedir.

Çalışma Grubu

Çalışmada International Association for the Evaluation of Educational Achievement'ın (IEA), öğrencilerin matematik ve fen alanlarındaki bilgi ve becerilerinin değerlendirilmesine yönelik projesi olan TIMSS (Trends in International Mathematics and Science Study) verileri kullanılmıştır.

TIMSS matematik ve fen alanında dünya genelinde birçok ülkede yapılan ve dört yılda bir tekrarlanan bir uygulamadır. TIMSS, 1995 yılından bu yana her dört yılda bir 60'tan fazla ülke ve bölgeden öğrencilerin katılımıyla gerçekleştirilmektedir. Öğrencilerin matematik ve fen bilimleri alanlarındaki başarılarını ölçmenin yanı sıra, öğrencilerin eğitim ortamları, öğretmenler ve okullar hakkında da bilgi toplamaktadır. Bu bilgiler, eğitim politikalarının geliştirilmesi ve eğitim sistemlerinin iyileştirilmesi için kullanılmaktadır.

TIMSS değerlendirmesi, matematik ve fen bilimleri olmak üzere iki ana içerik alanına odaklanır. Her bir içerik alanı, çeşitli konu alanlarını kapsar. Matematik değerlendirmesi sayılar, cebir, geometri, veri analizi ve olasılık gibi konuları içerirken; fen bilimleri değerlendirmesi biyoloji, kimya, fizik ve dünya bilimlerini kapsar. Uygulama, çoktan seçmeli sorular, kısa ve uzun açık uçlu sorular içeren bir test formatında yapılır. Sorular, ortak maddeler içeren

bağlanabilir/eşitlenebilir çeşitli sayıda kitapçığa dağıtılarak kağıt veya bilgisayar yoluyla uygulanır. Böylece öğrencilerin bilgi düzeylerini, problem çözme becerilerini ve uygulamalı düşünme yeteneklerini adil ve güvenilir bir şekilde ölçmeyi amaçlar. Bu çalışmada da TIMSS'in bilgisayar ortamında uygulandığı e-TIMSS uygulmasının verileri kullanılmıştır.

TIMSS öğrenci evrenini 4. ve 8. sınıf öğrencileri oluşturmaktadır. TIMSS için örneklem seçimi, katılımcı eğitim sistemlerinin ilköğretim 4. ve 8. yıllarında eğitim gören öğrenciler arasından olasılık temelli örnekleme yöntemleri kullanılarak yapılır. Bu süreç, öğrencilerin eğitim sistemlerini adil ve temsilci bir şekilde yansıtacak şekilde seçilmesini sağlamak için tasarlanmıştır. Örneklem, iki aşamalı bir katmanlaştırılmış küme örnekleme tasarımı kullanılarak belirlenir: İlk aşamada, okullar sistemli bir olasılık-ölçeğine-oranla (PPS) teknikle seçilir; ikinci aşamada ise seçilen okullardaki sınıflar örneklenir. TIMSS'de, hedef sınıf seviyelerindeki tam sınıflar örneklenir ve mümkün olduğunca her okuldan iki sınıf seçilir (TIMSS, 2019).

Araştırmanın veri toplama aracı olan TIMSS 2019'a 39 ülkeden 8. sınıfa devam eden yaklaşık 250.000 öğrenci, 4. Sınıfa devam eden yaklaşık 310.000 öğrenci katılmıştır. TIMSS 2019'da 14 soru kitapçığı yer almakta olup, bunlar arasından 7. ve 8. kitapçıklar eşitlenmek üzere çalışmanın veri toplama aracı olarak kullanılmıştır. Kayıp veriler çıkarıldıktan sonra 7. kitapçığı alan 7988 öğrenci ve 8. kitapçığı alan 7946 öğrenci çalışma grubunu oluşturmuştur.

Veri Toplama Aracı

Araştırmada TIMSS 2019'da uygulanan 14 soru kitapçığının tümü incelenmiş olup araştırmanın amacına uygun olarak içlerinde en çok madde takımı içeren 7. ve 8. kitapçıkların Fen Bilimleri Testi eşitlenmek üzere çalışmanın veri toplama aracı olarak kullanılmıştır. Veriler seçilirken ülke ayrımı yapılmamış ve bu iki kitapçığı alan verileri uygun olan tüm yanıtlayıcılar araştırmaya dâhil edilmiştir. Çalışmada 7. kitapçığı alan 7988 kişinin, 8. kitapçığı alan 7946 kişinin verileri kullanılmıştır.

7. ve 8. kitapçıkta 22 bağımsız madde ve 20 maddeden oluşan 4 madde takımı olmak üzere toplam 42 madde analize alınmıştır. Kitapçıklarda yer alan 4 madde takımı sırasıyla 4, 8, 3 ve 5 maddeden oluşmaktadır. 7. ve 8. kitapçıkların, ilk 11 bağımsız maddesi ve ilk 3 madde takımı (4, 8 ve 3 maddeden oluşan madde takımları) olmak üzere ilk 26 maddesi ortak maddedir. Eşdeğer olmayan gruplar ortak madde deseninde ölçek dönüştürme işlemi için gerekli olan denklemler bu ortak maddeler üzerinden elde edilmektedir.

7. ve 8. kitapçıkta yer alan maddelerden madde takımı sayısı ve bağımsız madde sayısı koşulu için 4 veri seti oluşturulmuştur. Madde takımı sayısı (4/8) ve bağımsız madde sayısı (11/22) koşulu için oluşturulan 4 ayrı veri setine ait madde takımı ve bağımsız madde sayıları Tablo 1'de verilmiştir.

Tablo 1

Veri Setlerindeki Madde Takımı ve Bağımsız Madde Sayıları

		Set 1	Set 2	Set 3	Set 4
7. ve 8. Kitapçıklar	Madde Takımı Sayısı	2	2	4	4
	Bağımsız	11	22	11	22
	(Madde Takımlarındaki Madde Sayıları)	(4,8)	(4,8)	(4,8,3,5)	(4,8,3,5)
	Toplam	23	34	31	42
Ortak Maddeler	Madde Takımı Sayısı	2	2	3	3
	Bağımsız	0	11	0	11
	(Madde Takımlarındaki Madde Sayıları)	(4,8)	(4,8)	(4,8,3)	(4,8,3)
	Toplam	12	23	15	26

Tablo 1 incelendiğinde 7. ve 8. kitapçıklardan oluşturulan 1. veri setinin 2 madde takımı ve 11 bağımsız maddeden, 2. veri setinin 2 madde takımı ve 22 bağımsız maddeden, 3. veri setinin 4 madde takımı ve 11 bağımsız maddeden, 4. veri setinin ise 4 madde takımı ve 22 bağımsız maddeden oluştuğu görülmektedir. Veri setlerindeki ortak maddelere bakıldığında 1. veri setindeki ortak maddelerin 2 madde takımından, 2. veri setindeki ortak maddelerin 2 madde takımı ve 11 bağımsız maddeden, 3. veri setindeki ortak maddelerin 3 madde

takımından, 4. veri setindeki ortak maddelerin 3 madde takımı ve 11 bağımsız maddeden oluştuğu görülmektedir.

Araştırmada incelenen madde ve yetenek parametreleri TBMTK, 2FM ve MTTK'nın 2PL modelleri kullanılarak elde edilmiştir. Bu şekilde, test forumda yer alan madde takım sayısının, bağımsız madde sayısının ve örneklem büyüklüğünün kullanılan modellere ve yöntemlere göre eşitleme hatasına olan etkisinin belirlenmesi amaçlanmıştır.

Verilerin Analizi

Çalışmada, TIMSS 2019'daki 7 ve 8 ve numaralı kitapçıklarda yer alan fen bilimleri bilişsel testinin maddelerinden kestirilen madde ve yetenek parametreleri kullanılmıştır. Bu parametreler, IRTPRO 5.2 (Cai ve diğerleri, 2011) yazılımında TBMTK, 2FM ve MTTK modelleri kullanılarak elde edilmiştir. Veri setlerinde tüm maddeler için ayrı ayrı a (eğim), b (eğim-eşik), c (eğim-kesişim) ve theta (yetenek) parametreleri elde edilmiştir. Çok boyutlu modellerde b (eğim-eşik) parametresi tek boyutlu modellerdekine benzer yorumlanmadığı ve doğru genellemeler yapamadığı için onun yerine c ($b=-c/a$) parametresinin kullanılması daha çok tercih edilen bir yöntemdir (Cai ve diğerleri, 2011; Min & He, 2014). Bu nedenle bu çalışmada modeller arası karşılaştırma yapabilmek için çok boyutlu modelde c parametresi kullanılmıştır.

Parametre kestiriminde Bock Aitkin yöntemi (Bock & Aitkin, 1981) kullanılmıştır. Elde edilen parametrelerin aynı ölçek üzerinde gösterilebilmesi için 8. kitapçıkta yer alan maddelerin parametrelerine, 7. kitapçıkta yer alan ortak maddelerin parametreleri referans alınarak ölçek dönüştürme işlemi uygulanmıştır. Ölçek dönüşümü için IRTEQ (Han, 2009) ve LINKMIRT (Yao, 2004) yazılımları kullanılmış, ayrı kalibrasyon ölçek dönüştürme yöntemlerinden ortalama-ortalama, ortalama-sigma, Stocking Lord ve Haebara yöntemleri uygulanmıştır. Eşitleme hatalarını elde edebilmek için bu üç ölçek dönüştürme yöntemiyle ayrı ayrı eğim (A) ve kesişim (B) katsayıları hesaplanmış, bu katsayılarla daha sonra kestirilen madde parametrelerine ilişkin hata miktarını gösteren Root Mean Squared Error (RMSE) değeri bulunmuştur.

Formların faktör yapıları incelendiğinde genel faktörün yanında madde takımlarını içeren ikincil faktörler olduğu gözlenmiştir. Madde takımlarındaki maddelerin yerel bağımsızlık varsayımına karar verebilmek için için yerel bağımlılık X^2 (LD X^2) istatistiği ve Yen'in Q_3 indeksi kullanılmıştır. LD X^2 'nin 10'dan büyük değerleri, madde takımı etkisinin 1'den büyük değerleri maddeler arasında yüksek yerel bağımlılık, 5-10 arasındaki değerleri orta düzey yerel bağımlılık, 5'den küçük değerleri ise düşük yerel bağımlılık olduğuna işaret eder (Cai ve diğerleri, 2011). Q_3 değeri'nin 0.2'den büyük olması maddeler arasında yerel bağımlılığın olduğuna, 0.4'ten büyük olması yüksek yerel bağımlılık olduğuna işaret eder (Yen & Fitzpatrick, 2006). Araştırmada yer alan madde takımlarındaki maddelerin, madde takımı içindeki LD X^2 ranj ve Q_3 değerleri Tablo 2'de verilmiştir. Tablo 3'te yer alan 1,2 ve 3 ile kodlanmış madde takımları her iki kitapçıkta da yer almaktadır, yani ortak maddeler içerir. 4 ve 5 ile kodlanmış madde takımları ise iki kitapçıktan sadece birinde yer almaktadır, yani farklı maddeler içerir.

Tablo 2

Madde Takımlarındaki Maddelerin LD X^2 ve Q_3 değerleri

Kitapçık	Madde Takımı	LD X^2 Ranjı	Q_3
Kitapçık 7	Madde Takımı 1 (4 madde)	(31.9 – 330.7)	0.32
	Madde Takımı 2 (8 madde)	(-0.5 – 504.4)	0.13
	Madde Takımı 3 (3 madde)	(11.3 – 41.1)	0.09
	Madde Takımı 4 (5 madde)	(0.0 – 107.0)	0.53
Kitapçık 8	Madde Takımı 1 (4 madde)	(11.4 – 229.2)	0.33
	Madde Takımı 2 (8 madde)	(0.4 – 475.4)	0.12
	Madde Takımı 3 (3 madde)	(-0.4 – 29.3)	0.08
	Madde Takımı 5 (5 madde)	(1.8 – 323.3)	0.21

Tablo 2 incelendiğinde madde takımlarındaki madde çiftlerinin LD X^2 değerlerinin genel olarak 10'un üzerinde ve yüksek düzeyde olduğu görülebilir. Q_3 değerlerine bakıldığında madde takımı 1, madde takımı 4 ve madde takımı 5'in yüksek düzeyde yerel bağımlılığa sahip olduğu söylenebilir. Buradan yola çıkarak madde çiftleri arasında yerel

bağımlılığın olduğu, bu kitapçıklar için 2FM ve MTTK yöntemlerinin kullanımının uygun olacağı söylenebilir. Çalışmada incelenen değişken ve koşullar Tablo 3'te verilmiştir.

Tablo 3

Çalışmada İncelenen Değişken ve Koşullar

Değişkenler	Koşullar	Koşul Sayısı
Madde takımı sayısı	2-4	2
Bağımsız madde sayısı	11-22	2
Örneklem büyüklüğü	1000-7946	2

Araştırma kapsamında uygulanan ölçek dönüştürme işlemlerin sonucu bulunan eşitleme hataları madde takımı sayısı (2-4), bağımsız madde sayısı (11-22) ve örneklem büyüklüğü (1000-7986) açısından incelenmiştir.

Çalışmada MTK, 2FM ve MTTK modellerinden elde edilen madde ve yetenek parametrelerine ölçek dönüştürme yöntemlerinden ortalama-ortalama (OO), ortalama-sigma (OS), Stocking-Lord (SL) ve Haebara (HA) yöntemleri uygulanmış, ölçek dönüştürme yöntemlerinden elde edilen eşitleme hataları karşılaştırılmıştır.

Değerlendirme Kriteri

Ölçek dönüştürme işlemi sonrası madde ve yetenek parametrelerinden elde edilen hataları uygulanan model ve yöntemlere göre karşılaştırabilmek için RMSE (Root Mean Squared Error) değeri hesaplanmıştır. RMSE değerinin formülü Eşitlik (13)'te verilmiştir.

$$RMSE(\tau_j) = \sqrt{\frac{\sum_{r=1}^R (\tau_{jr} - \tau_j)^2}{R}} \quad (13)$$

τ_j : j parametresine ait gerçek değer

τ_{jr} : r . gözlem için j parametresinin tahmin edilen değeri

R : gözlem sayısı

Çalışma kapsamında kullanılan MTK modellerinden kestirilen madde ve yetenek parametrelerinin ölçek dönüştürme yöntemlerinden elde edilen hata ortalamaları arasında farkın istatistiksel olarak anlamlı olup olmadığı tek yönlü varyans analizi (ANOVA) yapılarak değerlendirilmiştir. Analizler SPSS 25 yazılımı kullanılarak yapılmıştır.

Bölüm 4

Bulgular, Yorumlar ve Tartışma

Çalışma kapsamında TIMSS 2019 testinin 7. ve 8. kitapçıklarına araştırma koşulları çerçevesinde sırasıyla TBMTK, 2FM ve MTTK modelleri kullanılarak kestirilen madde ve yetenek parametrelerine ölçek dönüştürme işlemleri uygulanmış, elde edilen bulgular yorumlanmıştır.

Alt Problemlere İlişkin Bulgular ve Yorumlar

“Madde takımı tabanlı testlerde TBMTK, 2FM ve MTTK modelleri kullanılarak kestirilen madde ve yetenek parametrelerine eşdeğer olmayan gruplar ortak madde deseni altında OO, OS, S-L ve Haebara ölçek dönüştürme yöntemleri uygulandığında elde edilen hatalar madde takımı sayısı (2-4), bağımsız madde sayısı (11-22) ve örneklem büyüklüğüne (1000-7946) göre nasıl değişmektedir?”

Alt problemlerin çözümüne yönelik, TIMSS 2019 testinin 7. ve 8. kitapçıklarından belirlenen madde takımı sayısı, bağımsız madde sayısı koşulları göz önünde bulundurularak veri setleri oluşturulmuştur. Daha sonra bu veri setlerine TBMTK, 2FM ve MTTK modelleri uygulanarak madde ve yetenek parametreleri kestirilmiş, kestirilen parametrelere ölçek dönüştürme işlemi yapılarak RMSE değerleri bulunmuştur. Bulunan RMSE değerleri her ölçek dönüştürme yöntemi için tablo ve şekillerle gösterilmiştir.

Ortalama-Ortalama Yönteminden Elde Edilen RMSE Değerleri

Bu bölümde TBMTK, 2FM ve MTTK modelleri kullanılarak elde edilen madde ve yetenek parametrelerine ortalama-ortalama ölçek dönüştürme yöntemi uygulanmış ve elde edilen RMSE değerleri çalışmada ele alınan koşullara göre tablo ve grafiklerle yorumlanmıştır. Ortalama-ortalama ölçek dönüştürme yöntemi uygulandığında elde edilen RMSE değerleri Tablo 4’te verilmiştir.

Tablo 4*Ortalama-Ortalama Yönteminden Elde Edilen RMSE Değerleri*

Veri Seti	Örneklem Büyüküğü	Model	RMSE		
			a	c	theta
Veri Seti 1 (2 madde takımı) (11 bağımsız madde)	1000	MTTK	0.16	0.76	0.94
		2FM	0.02	0.42	0.40
		TBMTK	0.02	0.07	0.09
	7946	MTTK	0.06	0.18	0.08
		2FM	0.14	0.25	0.27
		TBMTK	0.08	0.12	0.12
Veri Seti 2 (2 madde takımı) (22 bağımsız madde)	1000	MTTK	0.09	0.35	0.41
		2FM	0.10	0.41	0.44
		TBMTK	0.01	0.02	0.01
	7946	MTTK	0.05	0.15	0.08
		2FM	0.10	0.17	0.08
		TBMTK	0.07	0.11	0.07
Veri Seti 3 (4 madde takımı) (11 bağımsız madde)	1000	MTTK	0.08	0.52	0.62
		2FM	0.06	0.54	0.60
		TBMTK	0.3	1.09	1.39
	7946	MTTK	0.03	0.18	0.13
		2FM	0.10	0.17	0.12
		TBMTK	0.22	0.84	1.05
Veri Seti 4 (4 madde takımı) (22 bağımsız madde)	1000	MTTK	0.06	0.31	0.36
		2FM	0.04	0.27	0.30
		TBMTK	0.16	0.41	0.48
	7946	MTTK	0.04	0.15	0.10
		2FM	0.07	0.14	0.06
		TBMTK	0.08	0.26	0.31

Tablo 4'te ortalama ortalama ölçek dönüştürme yöntemi uygulandığında veri seti 1'de en düşük RMSE değerlerinin TBMTK modelinde 1000 örnekleme, en yüksek RMSE değerlerinin MTTK modelinde 1000 örnekleme olduğu görülmektedir. Veri seti 2'de en düşük RMSE değerlerinin TBMTK modelinde 1000 örnekleme, en yüksek RMSE değerlerinin 2FM modelinde 1000 örnekleme olduğu görülmektedir. Veri seti 2'de tüm modellerde genel olarak RMSE değerleri diğer veri setlerinden daha düşük bulunmuştur. Veri seti 3'te en düşük RMSE

değerlerinin MTTK modelinde 7946 örnekleme, en yüksek RMSE değerlerinin TBMTK modelinde 1000 örnekleme olduğu görülmektedir. Veri seti 4'te en düşük RMSE değerlerinin 2FM ve MTTK modellerinde 7946 örnekleme, en yüksek RMSE değerlerinin TBMTK modelinde 1000 örnekleme olduğu görülmektedir. Genel olarak tüm modellerde en düşük hatalar veri seti 2'de görülürken en yüksek hatalar veri seti 3'te bulunmuştur.

Örneklem büyüklüğü 1000 olduğunda en düşük RMSE değerlerinin veri seti 2'de TBMTK modelinde, en yüksek RMSE değerlerinin veri seti 3'te TBMTK modelinde olduğu görülmektedir. Örneklem büyüklüğü 7946 olduğunda en düşük RMSE değerlerinin veri seti 2'de TBMTK modelinde, en yüksek RMSE değerlerinin veri seti 1'de 2FM modelinde olduğu görülmektedir.

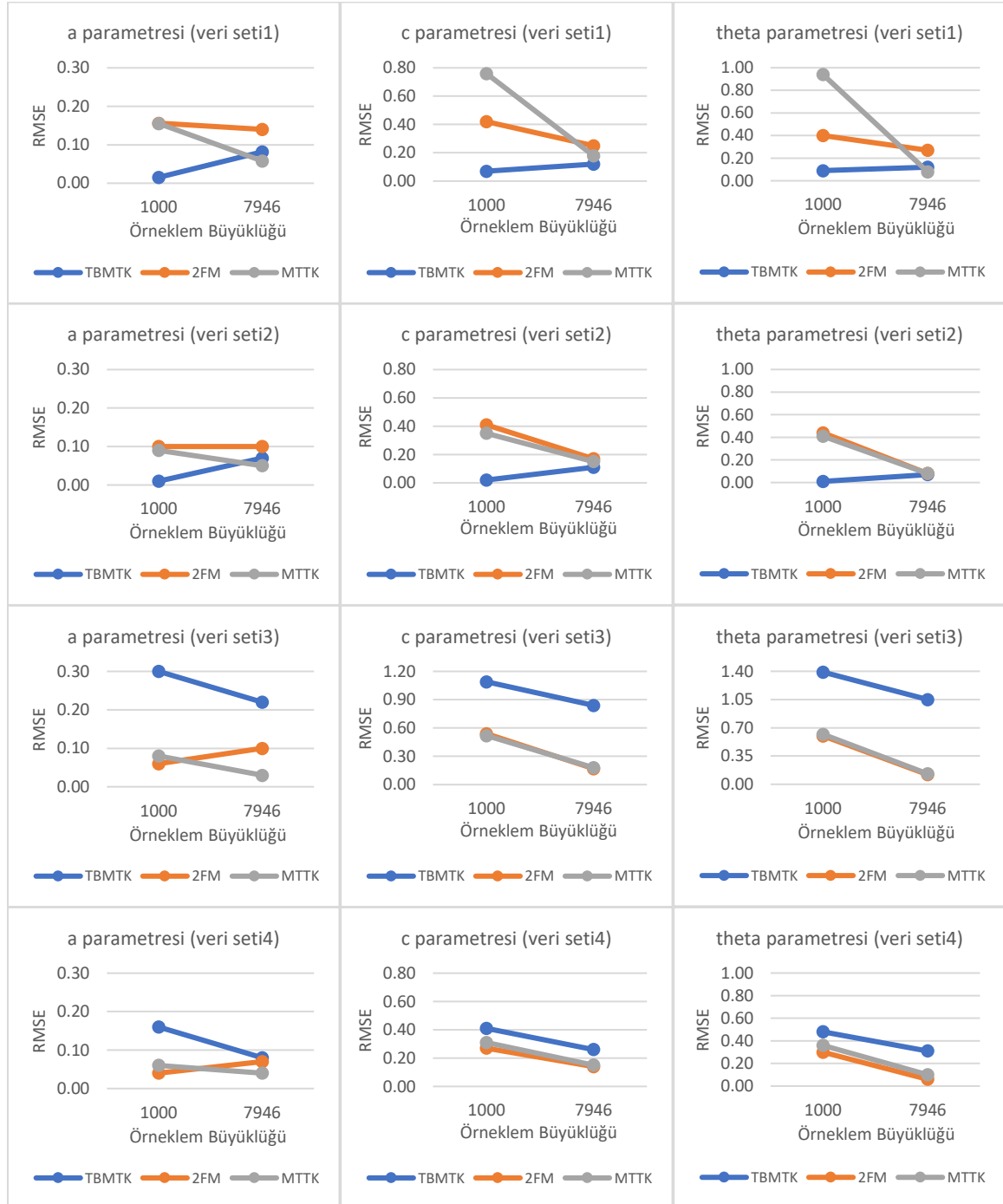
TBMTK için tüm parametrelerde en düşük RMSE değerleri veri seti 2'de 1000 örneklem büyüklüğünde elde edilmiştir. TBMTK için tüm parametrelerde en büyük RMSE değerleri ise veri seti 3'te 1000 örneklem büyüklüğünde elde edilmiştir. Madde takımı sayısının artması ve bağımsız madde sayısının azalmasının TBMTK'da hatayı artırıcı yönde etki yaptığı söylenebilir. 2FM için en düşük RMSE değerleri veri seti 4'te 7946 örneklem büyüklüğünde, en yüksek ise veri seti 3'te 1000 örnekleme elde edilmiştir. MTTK için en düşük RMSE değerleri veri seti 4'de 7946 örneklem büyüklüğünde, en yüksek ise veri seti 1'de 1000 örnekleme elde edilmiştir. 2FM ve MTTK'da madde takımı sayısının artmasının hatayı azaltıcı yönde etki yaptığı söylenebilir. Bağımsız madde sayısının artması ise tüm modellerde hatayı azaltıcı yönde etki yapmıştır. Bu bulgulara göre ortalama-ortalama ölçek dönüştürme yöntemi kullanıldığında madde takımı sayısı arttıkça 2FM ve MTTK modellerinin, madde takımı sayısı azaldıkça TBMTK modelinin daha az hatalı sonuçlar verdiği söylenebilir.

Bu bulgulara göre ortalama-ortalama ölçek dönüştürme yöntemi kullanıldığında madde takımı sayısı arttıkça 2FM ve MTTK modellerinin, madde takımı sayısı azaldıkça TBMTK modelinin daha az hatalı sonuçlar verdiği söylenebilir.

Ortalama-ortalama ölçek dönüştürme yöntemi uygulandığında elde edilen RMSE değerlerinin modellere ve koşullara göre değişimini daha iyi görebilmek için a, c ve theta parametrelerine ilişkin elde edilen RMSE değerleri Şekil 11’de grafik olarak gösterilmiştir.

Şekil 11

Ortalama-Ortalama Yönteminden Elde Edilen RMSE Değerlerinin Değişimi



Şekil 11'e göre ortalama-ortalama ölçek dönüştürme yöntemi kullanıldığında madde takımı sayısı az olan veri seti 1 ve 2'de 1000 örnekleme TBMTK'nın 2FM ve MTTK'ya göre daha az ölçek dönüştürme hatası verdiği bulunmuştur. Örnekleme büyüklüğü 7946 olduğunda TBMTK'dan elde edilen hata artarak, 2FM ve MTTK'dan elde edilen hatalar azalarak birbirine yakın değerler almıştır. Madde takımı sayısı fazla olan veri seti 3 ve 4'te ise tüm modellerde ölçek dönüştürme hataları örnekleme büyüklüğü arttıkça azalma eğilimi göstermiş ve 2FM ile MTTK modellerinden TBMTK'ya göre daha az hata elde edilmiştir.

Madde takım sayısının en fazla, bağımsız madde sayısının en az olduğu veri seti 3'te TBMTK'dan elde edilen hatanın 2FM ve MTTK modellerine göre belirgin bir şekilde yüksek olduğu görülmektedir. Bu bulgulardan yola çıkarak madde takımı sayısının fazla, bağımsız madde sayısının az olduğu madde takımı tabanlı testlerde 2FM ve MTTK modellerinin kullanımının hatayı belirgin oranda azaltacağı sonucu çıkarılabilir. Ayrıca 2FM ve MTTK modellerinin küçük örnekleme daha çok hata verdiği görülmektedir. Buradan bu modellerden ortalama-ortalama ölçek dönüştürme yöntemi kullanıldığında daha az hata elde etmek için büyük örneklere ihtiyaç olduğu sonucu çıkarılabilir.

Ortalama-Sigma Yönteminden Elde Edilen RMSE Değerleri

Bu bölümde TBMTK, 2FM ve MTTK modelleri kullanılarak elde edilen madde ve yetenek parametrelerine ortalama-sigma ölçek dönüştürme yöntemi uygulanmış ve bulunan RMSE değerleri koşullara göre tablo ve grafiklerle yorumlanmıştır. Ortalama-sigma ölçek dönüştürme yöntemi uygulandığında elde edilen RMSE değerleri Tablo 5'te verilmiştir.

Tablo 5*Ortalama-Sigma Yönteminden Elde Edilen RMSE Değerleri*

Veri Seti	Örneklem Büyüküğü	Model	RMSE		
			a	c	theta
Veri Seti 1 (2 madde takımı) (11 bağımsız madde)	1000	MTTK	0.59	0.76	0.74
		2FM	0.57	0.70	0.70
		TBMTK	0.1	0.26	0.33
	7946	MTTK	0.01	0.21	0.20
		2FM	0.12	0.39	0.50
		TBMTK	0.1	0.29	0.37
Veri Seti 2 (2 madde takımı) (22 bağımsız madde)	1000	MTTK	0.27	0.42	0.20
		2FM	0.35	0.48	0.19
		TBMTK	0.03	0.06	0.07
	7946	MTTK	0.05	0.15	0.08
		2FM	0.02	0.11	0.09
		TBMTK	0.02	0.08	0.07
Veri Seti 3 (4 madde takımı) (11 bağımsız madde)	1000	MTTK	0.38	0.54	0.50
		2FM	0.79	1.05	1.10
		TBMTK	0.12	0.24	0.10
	7946	MTTK	0.03	0.18	0.13
		2FM	0.02	0.15	0.12
		TBMTK	0.18	0.29	0.14
Veri Seti 4 (4 madde takımı) (22 bağımsız madde)	1000	MTTK	0.23	0.36	0.18
		2FM	0.24	0.34	0.17
		TBMTK	0.09	0.16	0.07
	7946	MTTK	0.06	0.16	0.08
		2FM	0.04	0.12	0.08
		TBMTK	0.08	0.17	0.09

Tablo 5'e göre ortalama-sigma ölçek dönüştürme yöntemi uygulandığında TBMTK için tüm parametrelerde en düşük RMSE değerleri 2 madde takımı ve 22 bağımsız madde içeren veri seti 2'de 1000 örneklem büyüklüğünde elde edilmiştir. TBMTK için tüm parametrelerde en büyük RMSE değerleri ise 4 madde takımı ve 11 bağımsız madde içeren veri seti 3'te 1000 örneklem büyüklüğünde elde edilmiştir. Madde takımı sayısının artması ve bağımsız madde sayısının azalmasının TBMTK'da hatayı artırıcı yönde etki yaptığı söylenebilir. Örneklem

büyükklüklerine göre bakıldığında örneklemin büyümesi TBMTK'da hatayı artırıcı yönde etki yapmıştır.

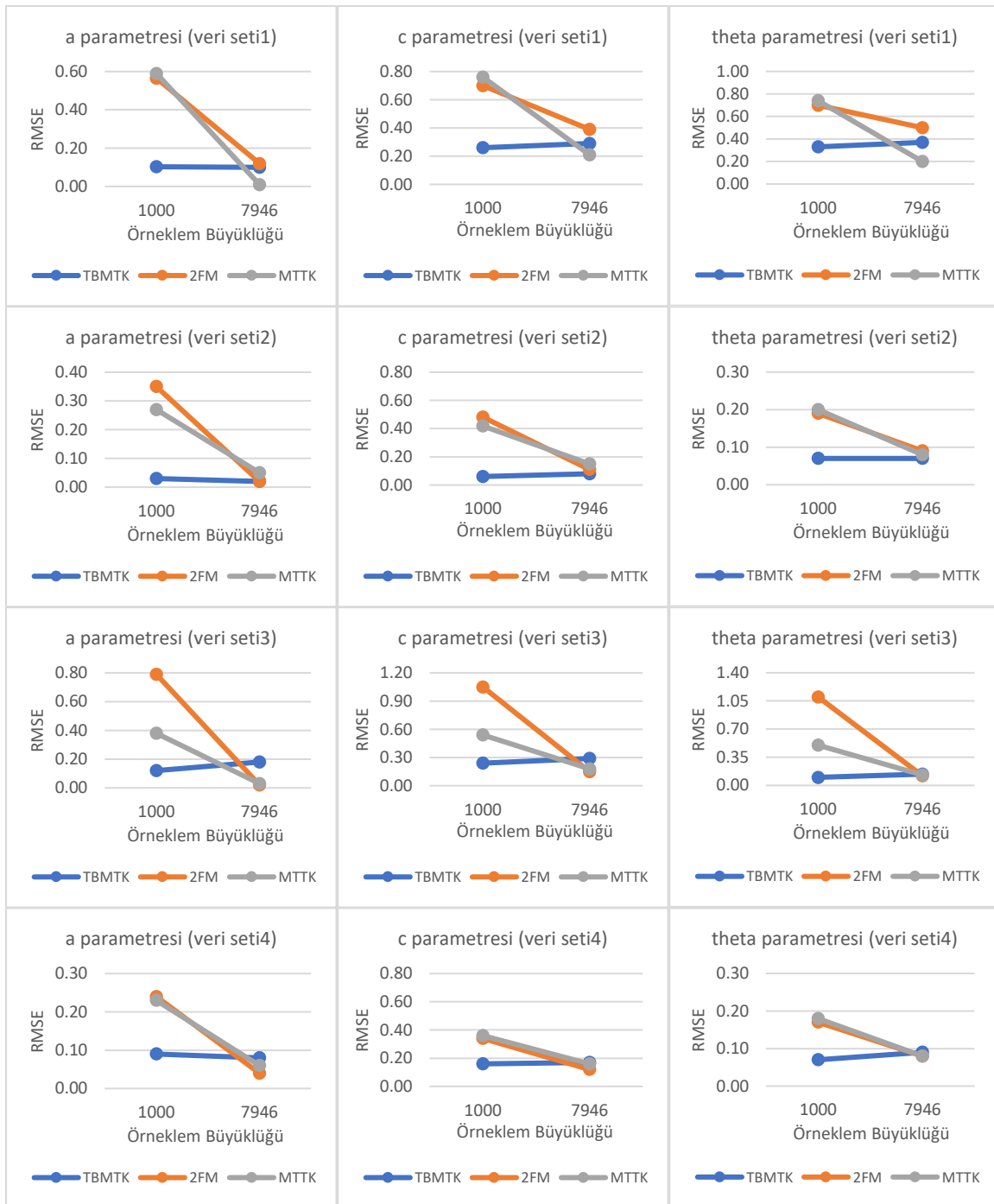
2FM'de en düşük RMSE değerleri 2 madde takımı ve 22 bağımsız madde içeren veri seti 2'de 7946 örneklem büyüklüğünde elde edilmiştir. En büyük RMSE değerleri ise 4 madde takımı ve 11 bağımsız madde içeren veri seti 3'te 1000 örneklem büyüklüğünde elde edilmiştir. MTTK'da en düşük RMSE değerleri 2 madde takımı ve 22 bağımsız madde içeren veri seti 2'de 7946 örneklem büyüklüğünde elde edilmiştir. En büyük RMSE değerleri ise 2 madde takımı ve 11 bağımsız madde içeren veri seti 1'de 1000 örneklem büyüklüğünde elde edilmiştir. TBMTK'nın aksine 2FM ve MTTK'da madde takımı sayısının artmasının hatayı azaltıcı yönde etki yaptığı söylenebilir. Bağımsız madde sayısının artması ise tüm modellerde hatayı azaltıcı yönde etki yapmıştır. Ayrıca 2FM ve MTTK'da örneklem büyüklüğünün azalması hatayı artırıcı yönde etki yapmıştır.

Bu bulgulara göre ortalama-sigma ölçek dönüştürme yöntemi kullanıldığında madde takımı sayısı arttıkça 2FM ve MTTK modellerinin, madde takımı sayısı azaldıkça TBMTK modelinin daha az hatalı sonuçlar verdiği söylenebilir.

Ortalama-sigma ölçek dönüştürme yöntemi uygulandığında elde edilen RMSE değerlerinin modellere ve koşullara göre değişimini daha iyi görebilmek için a, c ve theta parametrelerine ilişkin elde edilen RMSE değerleri Şekil 12'de grafik olarak gösterilmiştir.

Şekil 12

Ortalama-Sigma Yönteminden Elde Edilen RMSE Değerlerinin Değişimi



Şekil 12'ye göre ortalama-sigma ölçek dönüştürme yöntemi kullanıldığında tüm veri setlerinde 1000 örnekte TBMTK'nın 2FM ve MTTK'ya göre daha az ölçek dönüştürme hatası verdiği bulunmuştur. Örneklem büyüklüğü 7946 olduğunda 2FM ve MTTK'dan elde

edilen hatalar azalarak TBMTK'dan elde edilen hatalara yakın deęerler almıştır. Veri seti 1, 3 ve 4'te 2FM ile MTTK modellerinden TBMTK'ya göre daha az hata elde edilmiştir.

Veri setleri incelendiğinde tüm modellerde en az hatanın madde takım sayısı ve bağımsız madde sayısı en fazla olan veri seti 4'te elde edildiği görülmektedir. Bu bulgulardan yola çıkarak madde takımı sayısının fazla, bağımsız madde sayısının az olduğu madde takımı tabanlı testlerde 2FM ve MTTK modellerinin kullanımının hatayı belirgin oranda azaltacağı sonucu çıkarılabilir. Ayrıca 2FM ve MTTK modellerinin küçük örnekleme daha çok hata verdiği görülmektedir. Buradan bu modellerden ortalama-sigma ölçek dönüştürme yöntemi kullanıldığında daha az hata elde etmek için büyük örneklere ihtiyaç olduğu sonucu çıkarılabilir. Küçük örneklemlerde ortalama-sigma ölçek dönüştürme yöntemi kullanılması 2FM ve MTTK için hatayı artırıcı yönde etki yaparken, TBMTK modelinde hatayı olumsuz yönde etkilememiştir.

Stocking-Lord Yönteminden Elde Edilen RMSE Deęerleri

Bu bölümde TBMTK, 2FM ve MTTK modelleri kullanılarak elde edilen madde ve yetenek parametrelerine Stocking-Lord ölçek dönüştürme yöntemi uygulanmış ve elde edilen RMSE deęerleri çalışmada ele alınan koşullara göre tablo ve grafiklerle yorumlanmıştır. Stocking-Lord yöntemiyle yapılan ölçek dönüştürme işlemi sonucunda modellere göre madde ve yetenek parametrelerinin RMSE deęerleri Tablo 6'da verilmiştir.

Tablo 6*Stocking-Lord Yönteminden Elde Edilen RMSE Değerleri*

Veri Seti	Örneklem Büyüküğü	Model	RMSE		
			a	c	theta
Veri Seti 1 (2 madde takımı) (11 bağımsız madde)	1000	MTTK	0.10	0.34	0.44
		2FM	0.07	0.11	0.11
		TBMTK	0.02	0.11	0.09
	7946	MTTK	0.07	0.24	0.30
		2FM	0.04	0.09	0.11
		TBMTK	0.06	0.11	0.13
Veri Seti 2 (2 madde takımı) (22 bağımsız madde)	1000	MTTK	0.05	0.14	0.06
		2FM	0.06	0.10	0.05
		TBMTK	0.03	0.06	0.03
	7946	MTTK	0.03	0.13	0.09
		2FM	0.02	0.08	0.05
		TBMTK	0.01	0.05	0.05
Veri Seti 3 (4 madde takımı) (11 bağımsız madde)	1000	MTTK	0.04	0.19	0.13
		2FM	0.30	0.69	0.24
		TBMTK	0.17	0.70	0.87
	7946	MTTK	0.09	0.23	0.29
		2FM	0.01	0.02	0.01
		TBMTK	0.11	0.5	0.62
Veri Seti 4 (4 madde takımı) (22 bağımsız madde)	1000	MTTK	0.04	0.11	0.06
		2FM	0.08	0.12	0.10
		TBMTK	0.01	0.17	0.16
	7946	MTTK	0.01	0.10	0.11
		2FM	0.01	0.06	0.04
		TBMTK	0.02	0.14	0.15

Tablo 6'da Stocking-Lord ölçek dönüştürme yöntemi uygulandığında veri seti 1'de en düşük RMSE değerlerinin TBMTK modelinde 1000 örnekleme, en yüksek RMSE değerlerinin MTTK modelinde 1000 örnekleme olduğu görülmektedir. Veri seti 2'de en düşük RMSE değerlerinin TBMTK modelinde 7946 örnekleme, en yüksek RMSE değerlerinin MTTK modelinde 1000 örnekleme olduğu görülmektedir. Veri seti 2'de tüm modellerde genel olarak RMSE değerleri diğer veri setlerinden daha düşük bulunmuştur. Veri seti 3'te en düşük RMSE

değerlerinin 2FM modelinde 7946 örnekleme, en yüksek RMSE değerlerinin TBMTK modelinde 1000 örnekleme olduğu görülmektedir. Veri seti 4'te en düşük RMSE değerlerinin 2FM modelinde 7946 örnekleme, en yüksek RMSE değerlerinin TBMTK modelinde 1000 örnekleme olduğu görülmektedir. Veri seti 1 ve 2'de modellerden elde edilen hatalar birbirine yakinken madde takımı ve bağımsız madde eklenmesiyle veri seti 3 ve 4'te hatalar 2FM ve MTTK modelleri lehine ayrılmaktadır.

Örnekleme büyüklüğü 1000 olduğunda en düşük RMSE değerlerinin veri seti 2'de TBMTK modelinde, en yüksek RMSE değerlerinin veri seti 3'te TBMTK modelinde olduğu görülmektedir. Örnekleme büyüklüğü 7946 olduğunda en düşük RMSE değerlerinin veri seti 3'de 2FM modelinde, en yüksek RMSE değerlerinin veri seti 3'te TBMTK modelinde olduğu görülmektedir. Örnekleme büyüklüğü genel olarak tüm modellerde hatayı azaltıcı yönde olumlu etki yapsa da özellikle madde takımı ve bağımsız madde sayısı arttığında 2FM ve MTTK modellerinde daha fazla olumlu etkiye sahip olmaktadır.

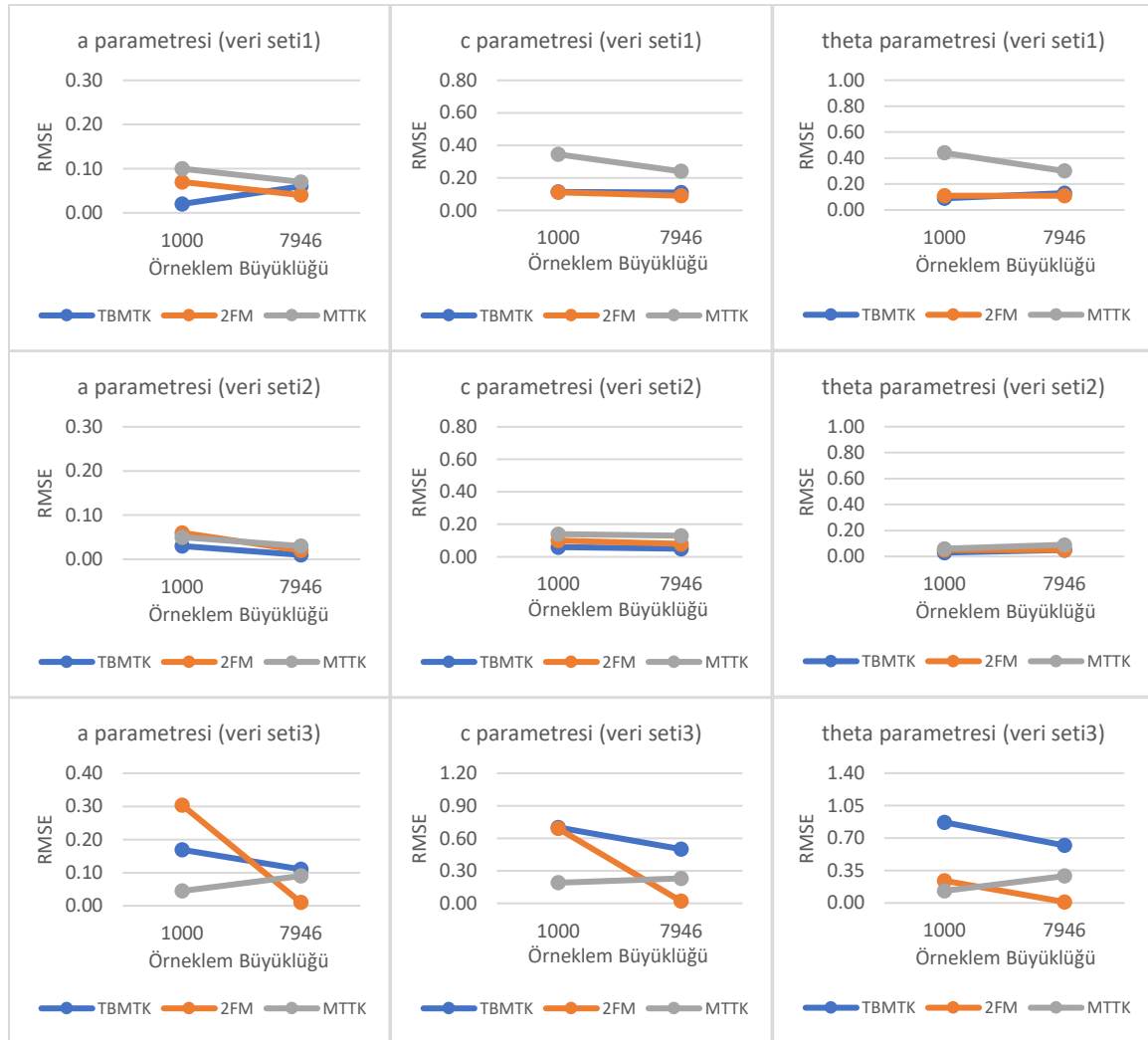
TBMTK için tüm parametrelerde en düşük RMSE değerleri veri seti 2'de 7946 örnekleme büyüklüğünde elde edilmiştir. TBMTK için tüm parametrelerde en büyük RMSE değerleri ise veri seti 3'te 1000 örnekleme büyüklüğünde elde edilmiştir. Madde takımı sayısının artması ve bağımsız madde sayısının azalmasının TBMTK'da hatayı artırıcı yönde etki yaptığı söylenebilir. Ayrıca hatanın azaldığı veri setlerinde 1000 örnekleme TBMTK'da hatayı azaltırken, hatanın arttığı veri setlerinde 1000 örnekleme hatayı artırıcı yönde etki yapmıştır. 2FM için en düşük RMSE değerleri veri seti 3'de 7946 örnekleme büyüklüğünde, en yüksek ise yine veri seti 3'te 1000 örnekleme elde edilmiştir. MTTK için en düşük RMSE değerleri veri seti 4'de 7946 örnekleme büyüklüğünde, en yüksek ise veri seti 1'de 1000 örnekleme elde edilmiştir. 2FM ve MTTK'da madde takımı sayısının artmasının hatayı azaltıcı yönde etki yaptığı söylenebilir. Bağımsız madde sayısının artması ise tüm modellerde hatayı azaltıcı yönde etki yapmıştır. Ayrıca 2FM ve MTTK'da örnekleme büyüklüğünün azalması hatayı artırıcı yönde etki yapmıştır.

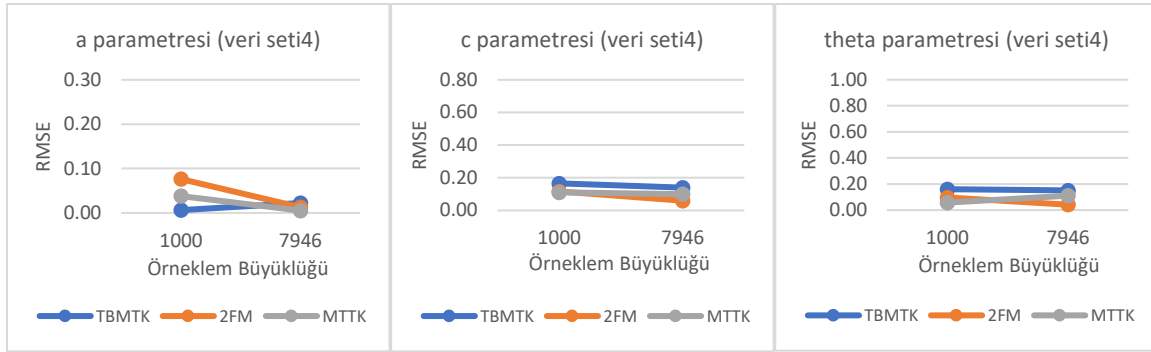
Bu bulgulara göre Stocking Lord ölçek dönüştürme yöntemi kullanıldığında madde takımı sayısı arttıkça 2FM ve MTTK modellerinin, madde takımı sayısı azaldıkça TBMTK modelinin daha az hatalı sonuçlar verdiği söylenebilir.

Stocking Lord ölçek dönüştürme yöntemi uygulandığında elde edilen RMSE değerlerinin modellere ve koşullara göre değişimini daha iyi görebilmek için a, c ve theta parametrelerine ilişkin elde edilen RMSE değerleri Şekil 13'te grafik olarak gösterilmiştir.

Şekil 13

Stocking Lord Yönteminden Elde Edilen RMSE Değerlerinin Değişimi





Şekil 13'e göre Stocking-Lord ölçek dönüştürme yöntemi kullanıldığında veri seti 1'de her iki örneklem büyüklüğünde en büyük RMSE değerleri MTTK'da elde edilmiştir. MTTK'da elde edilen RMSE değerleri örneklem büyüklüğü arttığında azalsa da hata miktarı 2FM ve TBMTK'ya göre yüksek bulunmuştur. Veri seti 2'de her iki örneklem büyüklüğünde tüm modellerde RMSE değerleri düşük bulunmuştur. Veri seti 3'te 2FM ve TBMTK modellerinde hata miktarı örneklem büyüklüğü artınca azalmış, MTTK'da ise artmıştır. 1000 örnekleme en düşük RMSE değeri MTTK'da 7946 örnekleme en düşük RMSE değeri 2FM modelinde elde edilmiştir. Örneklem büyüklüğünün artışı en çok 2FM modeli etkilemiştir. Veri seti 4'te tüm modellerde RMSE değerleri birbirine yakın ve düşük düzeyde edilmiştir. Hata değerleri örneklem büyüklüğünden çok fazla etkilenmemiştir.

Madde takım sayısının en fazla, bağımsız madde sayısının en az olduğu veri seti 3'te TBMTK'dan elde edilen hatanın 2FM ve MTTK modellerine göre belirgin bir şekilde yüksek olduğu görülmektedir. Bu bulgulardan yola çıkarak madde takımı sayısının fazla, bağımsız madde sayısının az olduğu madde takımı tabanlı testlerde 2FM ve MTTK modellerinin kullanımının hatayı belirgin oranda azaltacağı sonucu çıkarılabilir. Ayrıca Stocking-lord yöntemi kullanıldığında RMSE değerlerinin örneklem büyüklüğüne göre fazla miktarda değişmemesi ve hataların düşük bulunması dikkat çekici bir bulgu olmuştur.

Haebara Yönteminden Elde Edilen RMSE Değerleri

Bu bölümde TBMTK, 2FM ve MTTK modelleri kullanılarak elde edilen madde ve yetenek parametrelerine Haebara ölçek dönüştürme yöntemi uygulanmış ve elde edilen RMSE

değerleri tablo ve grafiklerle yorumlanmıştır. Haebara ölçek dönüştürme yöntemi uygulandığında elde edilen RMSE değerleri Tablo 7’de verilmiştir.

Tablo 7

Haebara Yönteminden Elde Edilen RMSE Değerleri

Veri Seti	Örneklem Büyüküğü	Model	RMSE		
			a	c	theta
Veri Seti 1 (2 madde takımı) (11 bağımsız madde)	1000	MTTK	0.21	0.47	0.47
		2FM	0.06	0.20	0.11
		TBMTK	0.09	0.14	0.07
	7946	MTTK	0.05	0.25	0.31
		2FM	0.06	0.14	0.05
		TBMTK	0.09	0.16	0.18
Veri Seti 2 (2 madde takımı) (22 bağımsız madde)	1000	MTTK	0.03	0.07	0.03
		2FM	0.02	0.06	0.03
		TBMTK	0.02	0.04	0.02
	7946	MTTK	0.02	0.11	0.09
		2FM	0.01	0.05	0.05
		TBMTK	0.02	0.06	0.07
Veri Seti 3 (4 madde takımı) (11 bağımsız madde)	1000	MTTK	0.03	0.09	0.12
		2FM	0.38	0.92	0.32
		TBMTK	0.23	0.7	0.88
	7946	MTTK	0.09	0.24	0.31
		2FM	0.02	0.11	0.13
		TBMTK	0.11	0.46	0.57
Veri Seti 4 (4 madde takımı) (22 bağımsız madde)	1000	MTTK	0.01	0.05	0.03
		2FM	0.01	0.07	0.06
		TBMTK	0.01	0.11	0.10
	7946	MTTK	0.01	0.09	0.09
		2FM	0.03	0.06	0.05
		TBMTK	0.01	0.12	0.13

Tablo 7’de Haebara ölçek dönüştürme yöntemi uygulandığında veri seti 1’de en düşük RMSE değerlerinin 2FM modelinde 7946 örnekleme, en yüksek RMSE değerlerinin MTTK modelinde 1000 örnekleme olduğu görülmektedir. Veri seti 2’de en düşük RMSE değerlerinin

TBMTK modelinde 1000 örnekleme, en yüksek RMSE değerlerinin MTTK modelinde 7946 örnekleme olduğu görülmektedir. Veri seti 2’de tüm modellerde genel olarak RMSE değerleri diğer veri setlerinden daha düşük bulunmuştur. Veri seti 3’te en düşük RMSE değerlerinin 2FM modelinde 7946 örnekleme, en yüksek RMSE değerlerinin 2FM modelinde 1000 örnekleme olduğu görülmektedir. Veri seti 4’te en düşük RMSE değerlerinin MTTK modelinde 1000 örnekleme, en yüksek RMSE değerlerinin TBMTK modelinde 7946 örnekleme olduğu görülmektedir. Genel olarak tüm modellerde en düşük hatalar veri seti 2’de görülürken en yüksek hatalar veri seti 3’te bulunmuştur.

Örneklem büyüklüğü 1000 olduğunda en düşük RMSE değerlerinin veri seti 2’de TBMTK modelinde, en yüksek RMSE değerlerinin veri seti 3’te TBMTK modelinde olduğu görülmektedir. Örneklem büyüklüğü 7946 olduğunda en düşük RMSE değerlerinin veri seti 2’de 2FM modelinde, en yüksek RMSE değerlerinin veri seti 3’te TBMTK modelinde olduğu görülmektedir. Örneklem büyüklüğü genel olarak tüm modellerde hatayı azaltıcı yönde olumlu etki yapsa da özellikle madde takımı ve bağımsız madde sayısı arttığında 2FM ve MTTK modellerinde daha fazla olumlu etkiye sahip olmaktadır.

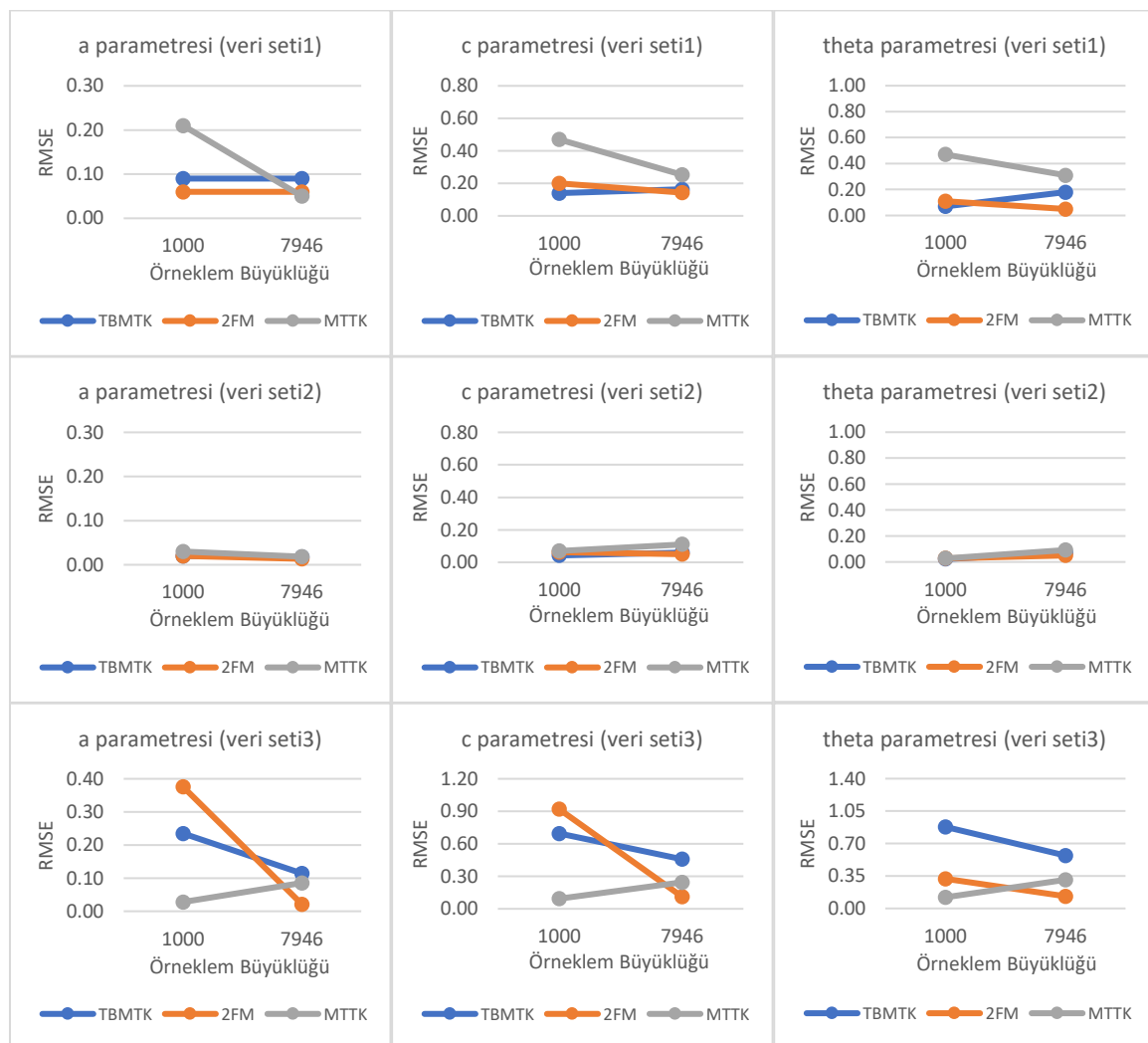
TBMTK için tüm parametrelerde en düşük RMSE değerleri veri seti 2’de 1000 örneklem büyüklüğünde elde edilmiştir. TBMTK için tüm parametrelerde en büyük RMSE değerleri ise veri seti 3’te 1000 örneklem büyüklüğünde elde edilmiştir. Madde takımı sayısının artması ve bağımsız madde sayısının azalmasının TBMTK’da hatayı artırıcı yönde etki yaptığı söylenebilir. Ayrıca hatanın azaldığı veri setlerinde 1000 örneklem TBMTK’da hatayı azaltırken, hatanın arttığı veri setlerinde 1000 örneklem hatayı artırıcı yönde etki yapmıştır. 2FM için en düşük RMSE değerleri veri seti 2’de 7946 örneklem büyüklüğünde, en yüksek ise veri seti 3’te 1000 örnekleme elde edilmiştir. MTTK için en düşük RMSE değerleri veri seti 4’de 1000 örneklem büyüklüğünde, en yüksek ise veri seti 1’de 1000 örnekleme elde edilmiştir. 2FM ve MTTK’da madde takımı sayısının artmasının hatayı azaltıcı yönde etki yaptığı söylenebilir. Bağımsız madde sayısının artması ise tüm modellerde hatayı azaltıcı yönde etki yapmıştır. Bu bulgulara göre Haebara ölçek dönüştürme yöntemi kullanıldığında

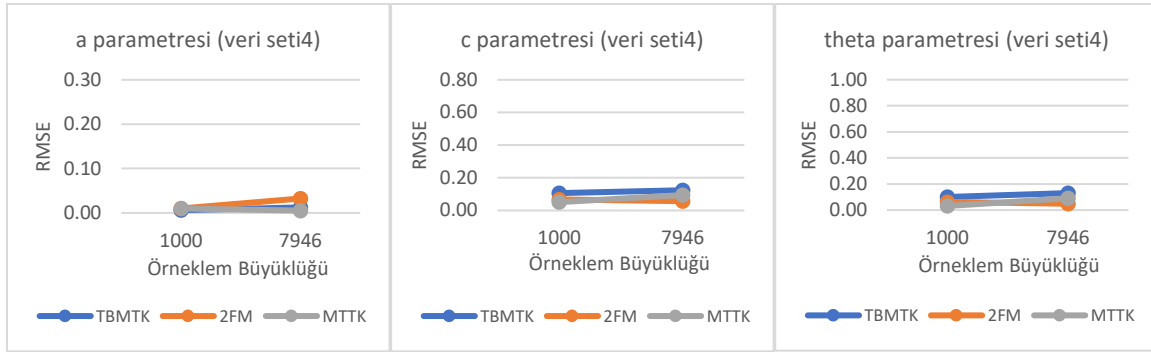
madde takımı sayısı arttıkça 2FM ve MTTK modellerinin, madde takımı sayısı azaldıkça TBMTK modelinin daha az hatalı sonuçlar verdiği söylenebilir.

Haebara ölçek dönüştürme yöntemi uygulandığında elde edilen RMSE değerlerinin modellere ve koşullara göre değişimini daha iyi görebilmek için a, c ve theta parametrelerine ilişkin elde edilen RMSE değerleri Şekil 14'te grafik olarak gösterilmiştir.

Şekil 14

Haebara Yönteminden Elde Edilen RMSE Değerlerinin Değişimi





Şekil 14'e göre Haebara ölçek dönüştürme yöntemi kullanıldığında veri seti 1'de genel olarak en büyük RMSE değerleri MTTK'da elde edilmiştir. Veri seti 2 ve 4'te her iki örneklem büyüklüğünde tüm modellerde RMSE değerleri düşük bulunmuştur. Hata değerleri örneklem büyüklüğünden çok fazla etkilenmemiştir. Veri seti 3'te 2FM ve TBMTK modellerinde hata miktarı örneklem büyüklüğü artınca azalmış, MTTK'da ise artmıştır. 1000 örnekleme en düşük RMSE değeri MTTK'da 7946 örnekleme en düşük RMSE değeri 2FM modelinde elde edilmiştir. Örneklem büyüklüğünün artışı en çok 2FM modeli etkilemiştir.

Madde takım sayısının en fazla, bağımsız madde sayısının en az olduğu veri seti 3'te TBMTK'dan elde edilen hatanın 2FM ve MTTK modellerine göre belirgin bir şekilde yüksek olduğu görülmektedir. Bu bulgulardan yola çıkarak madde takımı sayısının fazla, bağımsız madde sayısının az olduğu madde takımı tabanlı testlerde 2FM ve MTTK modellerinin kullanımının hatayı belirgin oranda azaltacağı sonucu çıkarılabilir. Ayrıca Haebara yöntemi kullanıldığında RMSE değerlerinin örneklem büyüklüğüne göre fazla miktarda değişmemesi ve hataların düşük bulunması dikkat çekici bir bulgu olmuştur. Habera yöntemi Stocking-Lord yöntemi ile birlikte hataların en düşük ve en az düzeyde değiştiği yöntemler olmuştur.

Modellere Göre Testin Tümünden Kestirilen Parametrelerden Elde Edilen Ortalama Hata Değerlerinin Karşılaştırılması

“TBMTK, 2FM ve MTTK modelleri testin tamamına uygulandığında elde edilen ölçek dönüştürme hatalarının koşullar arası etkileşimi ve anlamlılığı nasıldır?”

Alt problemin çözümü için, testin tümüne (4 madde takımı-22 bağımsız madde, 7946 örneklem) TBMTK, 2FM ve MTTK modelleri uygulanarak kestirilen madde ve yetenek

parametrelerine ilişkin ölçek dönüştürme hataları karşılaştırılmış ve hatalar arasındaki farkın anlamlılığı tek yönlü ANOVA testi ile incelenmiştir. Elde edilen sonuçlar sırasıyla a, c ve theta olmak üzere her parametre için ayrı tablolandırılmıştır.

Tablo 8'de a parametresine ilişkin TBMTK, 2FM ve MTTK modelleriyle kestirilen değerlerin ölçek dönüştürme hataları arasındaki ortalama farkın anlamlılığına ilişkin ANOVA testi sonuçları sunulmuştur.

Tablo 8

Modellere Göre Maddelerin a Parametresinden Kestirilen Ortalama Hata Değerlerinin Karşılaştırılması

Ölçek Dönüştürme Yöntemi	Model	N	Ort.	Ss	F	p	Eta-kare
OO	MTTK	42	0.035	0.016	32.387	.000	0.345
	2FM	42	0.070	0.028			
	TBMTK	42	0.074	0.026			
OS	MTTK	42	0.055	0.021	44.688	.000	0.421
	2FM	42	0.033	0.014			
	TBMTK	42	0.077	0.027			
SL	MTTK	42	0.005	0.002	98.433	.000	0.615
	2FM	42	0.012	0.005			
	TBMTK	42	0.021	0.007			
HA	MTTK	42	0.005	0.002	125.485	.000	0.671
	2FM	42	0.030	0.012			
	TBMTK	42	0.012	0.004			

Tablo 8 incelendiğinde, tüm ölçek dönüştürme yöntemleri için modellerden elde edilen a parametresine ilişkin ortalama ölçek dönüştürme hatalarının anlamlı biçimde farklılık gösterdiği görülmektedir ((F(2, 42) = [32.387], p = 0.00, $\eta^2=0.345$), (F(2, 42) = [44.688], p = 0.00, $\eta^2=0.421$), (F(2, 42) = [98.433], p = 0.00, $\eta^2=0.615$), (F(2, 42) = [125.485], p = 0.00, $\eta^2=0.671$)). Eta kare incelendiğinde a parametrelerine ilişkin ortalama hata farklarının tüm yöntemlerde büyük etkiye sahip olduğu görülmüştür. Anlamlılığın hangi gruplar arasında olduğunun tespiti için tüm yöntemlere post-hoc testi uygulanmıştır. Uygulanan Tukey testinin bulguları Tablo 9'da verilmiştir.

Tablo 9

Modellere Göre Maddelerin a Parametresinden Kestirilen Ortalama Hata Değerlerine İlişkin İkili Karşılaştırma Testi Sonuçları

Ölçek Dönüştürme Yöntemi	Model	Ort. Fark	SH	p
OO	MTTK-2FM	-0.034	0.005	.000
	MTTK-TBMTK	-0.039	0.005	.000
	2FM-TBMTK	-0.005	0.005	.646
OS	MTTK-2FM	0.022	0.005	.000
	MTTK-TBMTK	-0.022	0.005	.000
	2FM-TBMTK	-0.044	0.005	.000
SL	MTTK-2FM	-0.007	0.001	.000
	MTTK-TBMTK	-0.016	0.001	.000
	2FM-TBMTK	-0.009	0.001	.000
HA	MTTK-2FM	-0.025	0.002	.000
	MTTK-TBMTK	-0.007	0.002	.000
	2FM-TBMTK	0.018	0.002	.000

Tablo 9 incelendiğinde, ortalama-ortalama yönteminde 2FM-TBMTK modelleri hariç (O= -0.005, SH= 0.005) tüm ölçek dönüştürme yöntemlerinde modellerden elde edilen a parametresine ilişkin ortalama ölçek dönüştürme hata ortalamalarının anlamlı biçimde farklılık gösterdiği görülmektedir. Ortalama farklar incelendiğinde genel olarak MTTK-2FM arasında 2FM yönünde, MTTK-TBMTK arasında TBMTK yönünde, 2FM-TBMTK arasında TBMTK yönünde ortalama farklar daha yüksek çıkmıştır. Buradan a parametresi için en iyi çalışan modelin MTTK, daha sonra 2FM olduğu sonucu çıkarılabilir.

Tablo 10'da c parametresine ilişkin TBMTK, 2FM ve MTTK modelleriyle kestirilen değerlerin ölçek dönüştürme hataları arasındaki ortalama farkın anlamlılığına ilişkin ANOVA testi sonuçları sunulmuştur.

Tablo 10

Modellere Göre Maddelerin c Parametresinden Kestirilen Ortalama Hata Değerlerinin Karşılaştırılması

Ölçek Dönüştürme Yöntemi	Model	N	Ort.	Ss	F	P	Eta-kare
OO	MTTK	42	0.140	0.068			
	2FM	42	0.106	0.099	18.432	.000	0.231
	TBMTK	42	0.232	0.121			
OS	MTTK	42	0.132	0.094			
	2FM	42	0.107	0.059	1.424	.245	0.023
	TBMTK	42	0.137	0.103			
SL	MTTK	42	0.116	0.010			
	2FM	42	0.051	0.022	124.176	.000	0.669
	TBMTK	42	0.133	0.036			
HA	MTTK	42	0.094	0.010			
	2FM	42	0.046	0.034	115.637	.000	0.653
	TBMTK	42	0.122	0.020			

Tablo 10 incelendiğinde, ortalama-sigma yöntemi hariç ($F(2, 42) = [1.424]$, $p = 0.245$) tüm ölçek dönüştürme yöntemleri için modellerden elde edilen c parametresine ilişkin ortalama ölçek dönüştürme hata ortalamalarının anlamlı biçimde farklılık gösterdiği görülmektedir (($F(2, 42) = [18.432]$, $p = 0.00$, $\eta^2=0.231$), ($F(2, 42) = [124.176]$, $p = 0.00$, $\eta^2=0.6669$), ($F(2, 42) = [115.637]$, $p = 0.00$, $\eta^2=0.653$)). Bununla birlikte eta kare değerleri incelendiğinde maddelerin c parametrelerine ilişkin ortalama hata değerlerinin etki büyüklüklerinin anlamlı fark bulunan tüm yöntemlerde büyük etkiye sahip olduğu görülmüştür. Anlamlılığın gruplar içinde hangi düzeyler arasında olduğunun tespiti için OO, SL ve HA yöntemlerine ilişkin post-hoc testi yapılmıştır. Uygulanan Tukey testinin bulguları Tablo 11'de verilmiştir.

Tablo 11

Modellere Göre Maddelerin c Parametresinden Kestirilen Ortalama Hata Değerlerine İlişkin İkili Karşılaştırma Testi Sonuçları

Ölçek Dönüştürme Yöntemi	Model	Ort. Fark	SH	p
OO	MTTK-2FM	0.034	0.022	.247
	MTTK-TBMTK	-0.092	0.022	.000
	2FM-TBMTK	-0.126	0.022	.000
SL	MTTK-2FM	0.065	0.005	.000
	MTTK-TBMTK	-0.017	0.005	.006
	2FM-TBMTK	-0.081	0.005	.000
HA	MTTK-2FM	0.047	0.005	.000
	MTTK-TBMTK	-0.029	0.005	.000
	2FM-TBMTK	-0.076	0.005	.000

Tablo 11 incelendiğinde, ortalama-ortalama yönteminde MTTK-2FM modelleri dışında (O = 0.034, SH = 0.022) tüm ölçek dönüştürme yöntemlerinde modellerden elde edilen c parametresine ilişkin ortalama ölçek dönüştürme hata ortalamalarının anlamlı biçimde farklılık gösterdiği görülmektedir. Ortalama farklar incelendiğinde genel olarak MTTK-2FM arasında MTTK yönünde, MTTK-TBMTK arasında TBMTK yönünde, 2FM-TBMTK arasında TBMTK yönünde ortalama farklar daha yüksek çıkmıştır. Buradan c parametresi için en iyi çalışan modelin 2FM, daha sonra MTTK olduğu sonucu çıkarılabilir.

Tablo 12'de theta parametresine ilişkin TBMTK, 2FM ve MTTK modelleriyle kestirilen değerlerin ölçek dönüştürme hataları arasındaki ortalama farkın anlamlılığına ilişkin ANOVA testi sonuçları sunulmuştur.

Tablo 12

Modellere Göre Theta Parametresinden Kestirilen Ortalama Hata Değerlerinin Karşılaştırılması

Ölçek Dönüştürme Yöntemi	Model	N	Ort.	Ss	F	P	Eta-kare
OO	MTTK	7946	0.100	0.032			
	2FM	7946	0.046	0.033	56623	.000	0.826
	TBMTK	7946	0.304	0.075			
OS	MTTK	7946	0.071	0.043			
	2FM	7946	0.071	0.027	6.666	.000	0.001
	TBMTK	7946	0.069	0.050			
SL	MTTK	7946	0.110	0.009			
	2FM	7946	0.040	0.009	146130	.000	0.925
	TBMTK	7946	0.150	0.018			
HA	MTTK	7946	0.090	0.009			
	2FM	7946	0.042	0.025	59318	.000	0.860
	TBMTK	7946	0.130	0.009			

Tablo 12 incelendiğinde, tüm yöntemlerde modellerden elde edilen theta parametresine ilişkin ortalama ölçek dönüştürme hata ortalamalarının anlamlı biçimde farklılık gösterdiği görülmektedir (($F(2, 23835) = [56623]$, $p = 0.00$, $\eta^2=0.826$), ($F(2, 23835) = [6.666]$, $p = 0.00$, $\eta^2=0.001$), ($F(2, 23835) = [146130]$, $p = 0.00$, $\eta^2=0.925$), ($F(2, 23835) = [59318]$, $p = 0.00$, $\eta^2=0.860$)). Bununla birlikte eta kare değerleri incelendiğinde maddelerin b parametrelerine ilişkin ortalama hata değerlerinin etki büyüklüklerinin ortalama-sigma yöntemi hariç anlamlı fark bulunan tüm yöntemlerde büyük etkiye sahip olduğu görülmüştür. Ortalama-sigma yönteminde bulunan sonuç anlamlı görünse de örneklem büyüklüğünün etkisi göz önüne alındığında çıkan sonucun etkisinin olmadığı söylenebilir. Anlamlılığın gruplar içinde hangi düzeyler arasında olduğunun tespiti için OO, SL ve HA yöntemlerine ilişkin post-hoc testi yapılmıştır. Uygulanan Tukey testinin bulguları Tablo 13'de verilmiştir.

Tablo 13

Modellere Göre theta Parametresinden Kestirilen Ortalama Hata Değerlerine İlişkin İkili Karşılaştırma Testi Sonuçları

Ölçek Dönüştürme Yöntemi	Model	Ort. Fark	SH	p
OO	MTTK-2FM	0.054	0.001	.000
	MTTK-TBMTK	-0.203	0.001	.000
	2FM-TBMTK	-0.258	0.001	.000
SL	MTTK-2FM	0.070	0.001	.000
	MTTK-TBMTK	-0.039	0.001	.000
	2FM-TBMTK	-0.109	0.001	.000
HA	MTTK-2FM	0.048	0.001	.000
	MTTK-TBMTK	-0.039	0.001	.000
	2FM-TBMTK	-0.088	0.001	.000

Tablo 13 incelendiğinde, tüm ölçek dönüştürme yöntemlerinde modellerden elde edilen theta parametresine ilişkin ortalama ölçek dönüştürme hata ortalamalarının anlamlı biçimde farklılık gösterdiği görülmektedir. Ortalama farklar incelendiğinde genel olarak MTTK-2FM arasında MTTK yönünde, MTTK-TBMTK arasında TBMTK yönünde, 2FM-TBMTK arasında TBMTK yönünde ortalama farklar daha yüksek çıkmıştır. Buradan theta parametresi için en iyi çalışan modelin 2FM, daha sonra MTTK olduğu sonucu çıkarılabilir.

Bölüm 5

Sonuç ve Öneriler

Bu çalışmada madde takımı tabanlı testlerden tek boyutlu madde tepki kuramı, iki faktör ve madde takımı tepki kuramı modelleri kullanılarak kestirilen madde ve yetenek parametrelerinin ölçek dönüştürme hatalarının çeşitli koşullar altında incelenmesi amaçlanmıştır. Çalışma kapsamında parametre kestirim modelleri ölçek dönüştürme yöntemleri (OO, OS, SL ve Haebara), madde takım sayısı (2,4), bağımsız madde sayısı (11,22), örneklem büyüklüğü (1000 ve 7984) alt koşulları olmak üzere toplam 32 alt koşulda incelenmiştir. Yapılan analizler sonucu elde edilen bulgular daha önceki bölümde tablo ve grafiklerle açıklanmıştır. Bu bölümde, çalışmanın bulguları özetlenmiş ve literatürdeki diğer çalışmalarla karşılaştırılarak önerilerde bulunulmuştur.

Sonuçlar

Literatürde, hem parametre kestiriminde hem eşitleme çalışmalarında yerel madde bağımsızlık varsayımının ihlal edildiği durumlarda 2FM ve MTTK modellerinin klasik madde tepki teorisi modelinden daha yüksek parametre kestirimi doğruluğuna sahip olduğunu gösteren birçok çalışma bulunmaktadır (Bradlow ve diğerleri, 1999; Cao ve diğerleri, 2014; Chen, 2014; DeMars, 2006; Lee ve diğerleri, 2001; Koziol, 2016; Tao ve diğerleri, 2011; Xu, 2016; Wainer & Wang, 2000; Wainer ve diğerleri, 2007; Zhang, 2007; Zhang, 2010)

Çalışma kapsamında yer alan koşullara göre elde edilen RMSE değerlerinin değişimi (ölçek dönüştürme yöntemi, madde takım sayısı, bağımsız madde sayısı, örneklem büyüklüğü) başlıklar halinde açıklanmış ve tartışılmıştır.

Ölçek Dönüştürme Yöntemine Göre Elde Edilen RMSE Değerleri

Madde takımı tabanlı testlerde TBMTK, 2FM ve MTTK modelleri uygulanmış, kestirilen madde ve yetenek parametrelerine ölçek dönüştürme işlemleri yapılmış, daha sonra elde edilen ortalama RMSE değerleri karşılaştırılmıştır. Analiz sonucunda TBMTK için ortalama-

sigma, Stocking-Lord ve Haebara kullanıldığında, 2FM ve MTTK için Stocking-Lord ve Haebara kullanıldığında RMSE değerleri genel olarak düşüş göstermiştir.

Çalışma sonucunda elde edilen karakteristik eğri dönüştürme yöntemlerinin (Stocking Lord, Haebara) moment yöntemlerden (ortalama-ortalama, ortalama-sigma) daha az hata miktarı verdiği bulgusu literatürde ölçek dönüştürme yöntemlerinin farklı koşullar altında karşılaştırıldığı benzer çalışmalarla (Baker & Al-Karni, 1991; Gök, 2012; Gül ve diğerleri, 2017; Hanson & Beguin, 2002; Karkee & Wright, 2004; Kilmen, 2010; Kim & Kolen, 2006; Li ve diğerleri, 2006; Meng, 2012; Gübeş & Kelecioğlu, 2016; Stocking & Lord, 1983; Uyar ve diğerleri, 2018; Uysal, 2014) tutarlılık göstermektedir.

Gök (2012), iki parametrelili model verileri üzerinde yaptığı simülasyon çalışmasında ortalama-sigma yönteminin en düşük hata oranına sahip olduğunu bulmuş ve Keller vd. (2004) yaptıkları çalışmada, tüm koşullar altındal en düşük hata oranının yine ortalama-sigma yöntemiyle elde edildiğini gözlemlemişlerdir. Ayrıca, Tate (2000) tarafından yapılan çalışmada, her iki yöntemin benzer sonuçlar verdiği gözlemlenmiştir. Bu çalışmalarda bulunan hata değerleri ele alınan koşullar nedeniyle literatürdeki birçok benzer çalışmaya göre farklı çıkmış olabilir.

Bu bağlamda, madde takımı tabanlı testlerin en az bilgi kaybı ve hatayla eşitlenebilmesi için seçilen ölçek dönüştürme yönteminin önemli olduğu sonucuna varılabilir.

Madde Takım Sayısına Göre Elde Edilen RMSE Değerleri

Madde takımı tabanlı testlerde TBMTK, 2FM ve MTTK modelleri uygulanmış, kestirilen madde ve yetenek parametrelerine ölçek dönüştürme işlemleri yapılmış, daha sonra elde edilen ortalama RMSE değerleri madde takım sayısına göre (2, 4 madde takımı) karşılaştırılmıştır. Analiz sonucunda TBMTK'da madde takım sayısının 2 olduğu durumda en az eşitleme hatası elde edilirken, MTTK'da madde takım sayısının 4 olduğu durumlarda en az eşitleme hatası elde edilmiştir. 2FM modeli ise madde takımı sayısının değişimine direnç göstermiş ve madde sayısından diğer iki model gibi etkilenmemiştir. Bu bulgulara göre madde

takımı tabanlı testlerde madde takımı sayısının artması TBMTK'da yerel madde bağımlılığı ihlal edildiği için hatayı artırıcı yönde etki yaptığı, 2FM ve özellikle MTTK'da hatayı azaltıcı yönde etki yaptığı söylenebilir Bu bulgu Koziol'un (2016) yaptığı çalışma ile tutarlılık göstermektedir.

Bu bağlamda, madde takımı tabanlı testlere kullanılan madde takımı sayısı artıkça parametre kestirimi ve eşitleme çalışmaları için TBMTK modelinin uygulanması hataları artıracak olup madde takımı sayısına göre kullanılacak modelin belirlenmesinin önemli olduğu sonucuna varılabilir.

Bağımsız Madde Sayısına Göre Elde Edilen RMSE Değerleri

Madde takımı tabanlı testlerde testlerde TBMTK, 2FM ve MTTK modelleri uygulanmış, kestirilen madde ve yetenek parametrelerine ölçek dönüştürme işlemleri yapılmış, daha sonra elde edilen ortalama RMSE değerleri bağımsız madde sayısına göre (11, 22 bağımsız madde) karşılaştırılmıştır. Bunun sonucunda bağımsız madde sayısı artışının genel olarak tüm modellerde hatayı azaltıcı etkisi olduğu gözlenmiştir. Özellikle TBMTK'da bağımsız madde sayısı arttıkça eşitleme hatası dramatik bir şekilde azalırken, MTTK ve 2FM modellerindeki hata değerleri madde takımı sayısının değişimine daha çok direnç göstermiş ve daha kısıtlı azalmıştır. Buradan madde takım sayısının testteki oranı arttıkça yerel bağımlılığı göz önüne alan 2FM ve MTTK modellerinin, yerel bağımlılığı göz önüne almayan TBMTK gibi modellerden daha az hatalı sonuçlar elde edeceği sonucuna ulaşılabilir. Bu bulgu benzer çalışmalarla (Cao ve diğerleri, 2014; Chen, 2014; Tao ve diğerleri, 2011; Xu, 2016; Zhang, 2010) tutarlılık göstermektedir.

Bu bağlamda, madde takımı tabanlı testlerde kullanılan madde takımı sayısı, kullanılan bağımsız madde sayısına oranla arttıkça parametre kestirimi ve eşitleme çalışmaları için TBMTK modelinin uygulanması hataları artıracak olup bağımsız madde ve madde takımı sayısına göre kullanılacak modelin belirlenmesinin önemli olduğu sonucuna varılabilir.

Örneklem büyüklüğüne göre Elde Edilen RMSE değerleri

Madde takımı tabanlı testlerde TBMTK, 2FM ve MTTK modelleri uygulanmış, kestirilen madde ve yetenek parametrelerine ölçek dönüştürme işlemleri yapılmış, daha sonra elde edilen ortalama RMSE değerleri örneklem büyüklüğüne göre (1000, 7946) karşılaştırılmıştır.

Örneklem büyüklüğünün artmasının TBMTK'da özellikle madde takımı ve bağımsız madde sayısı az olan veri setlerinde hatayı artırıcı etkisi gözlenirse de kullanılan tüm modellerde örneklem büyüklüğündeki artış ile genel olarak daha düşük ölçek dönüştürme hatası değerleri elde edilmiştir. Bu bulgu literatürde benzer çalışmalardan elde edilen bulgularla uyumludur (Atar & Yeşiltaş, 2017; Bökeoğlu ve diğerleri, 2022; Cho, 2007; Gök & Kelecioğlu, 2014; Hanson & Beguin, 2002; Kogar, 2016; Zhang, 2010; Zor, 2023). Çalışmada ayrıca 2FM ve MTTK modellerinin örneklem büyüklüğünden TBMTK'ya göre daha fazla etkilendiği görülmüştür. Buradan, ikincil boyutun parametrelerinin doğru kestirilebilmesi için daha büyük örneklemelere ihtiyaç olduğu sonucu çıkarılabilir. Çalışmada TBMTK'da bazı koşullarda örneklem büyüklüğü azalırken hatanın da azalması rastgele seçilen 1000 örnekte yer alan yanıtlayıcıların daha homojen bir grup olmasından, verinin tek boyutluluğa daha yakın olmasından ve seçilen ölçek dönüştürme yönteminden kaynaklanmış olabilir.

Bu bağlamda, madde takımı tabanlı testlerin en az bilgi kaybı ve hatayla eşitlenebilmesi için yeterli örneklem büyüklüğünün kullanılmasının önemli olduğu sonucuna varılabilir. Ölçek dönüştürme hatası örneklem büyüklüğü arttıkça genel olarak düşme eğiliminde olsa da testin amacı ve verimliliği göz önünde bulundurularak çalışmada yer alan bazı koşullarda 1000 örneklem büyüklüğündemadde takımı tabanlı testlerin ölçek dönüştürme işlemleri gerçekleştirilebilir.

Modellere göre RMSE değerlerinin değişimi

Araştırmada ölçek dönüştürme yöntemleri testin tümüne (4 madde takımı, 22 bağımsız madde, 7946 örneklem) uygulandığında madde ve yetenek parametrelerinin kullanılan modellere göre (TBMTK, 2FM, MTTK) değişen hataları incelenmiştir.

Madde parametrelerinden a parametresi incelendiğinde ortalama-ortalama yöntemi dışındaki tüm ölçek dönüştürme yöntemlerinde modellerden elde edilen a parametresine ilişkin ortalama ölçek dönüştürme hata ortalamalarının anlamlı biçimde farklılık gösterdiği ve etki büyüklüklerinin yüksek olduğu görülmektedir. Ortalama farklar incelendiğinde genel olarak MTTK-2FM arasında 2FM yönünde, MTTK-TBMTK arasında TBMTK yönünde, 2FM-TBMTK arasında TBMTK yönünde ortalama farklar daha yüksek çıkmıştır. Buradan a parametresi için en iyi çalışan modelin MTTK, daha sonra 2FM olduğu sonucu çıkarılabilir.

Madde parametrelerinden c parametresi incelendiğinde ortalama-ortalama yöntemi dışındaki tüm ölçek dönüştürme yöntemlerinde modellerden elde edilen c parametresine ilişkin ortalama ölçek dönüştürme hata ortalamalarının anlamlı biçimde farklılık gösterdiği ve etki büyüklüklerinin yüksek olduğu görülmektedir. Ortalama farklar incelendiğinde genel olarak MTTK-2FM arasında MTTK yönünde, MTTK-TBMTK arasında TBMTK yönünde, 2FM-TBMTK arasında TBMTK yönünde ortalama farklar daha yüksek çıkmıştır. Buradan c parametresi için en iyi çalışan modelin 2FM, daha sonra MTTK olduğu sonucu çıkarılabilir.

Yetenek parametresi incelendiğinde ortalama-sigma hariç tüm ölçek dönüştürme yöntemlerinde modellerden elde edilen theta parametresine ilişkin ortalama ölçek dönüştürme hata ortalamalarının anlamlı biçimde farklılık gösterdiği ve etki büyüklüklerinin yüksek olduğu görülmektedir. Ortalama farklar incelendiğinde genel olarak MTTK-2FM arasında MTTK yönünde, MTTK-TBMTK arasında TBMTK yönünde, 2FM-TBMTK arasında TBMTK yönünde ortalama farklar daha yüksek çıkmıştır. Buradan theta parametresi için en iyi çalışan modelin 2FM, daha sonra MTTK olduğu sonucu çıkarılabilir.

Öneriler

Bu bölümde çalışmadan elde edilen bulgulara yönelik öneriler ve benzer çalışmalar yapan yeni araştırmacılara yönelik öneriler sıralanmıştır.

Çalışmanın Bulgularına İlişkin Öneriler

Çalışmanın sonucunda ortaya çıkan bulgular ışığında verilebilecek öneriler aşağıda sıralanmıştır.

- Bu çalışmada madde takımı tabanlı testler çeşitli koşullar altında sırasıyla TBMTK, 2FM ve MTTK modelleri kullanılarak ölçek dönüştürme işlemleri yapılmış ve ölçek dönüştürme sonucu elde edilen RMSE değerleri kullanılan modele göre karşılaştırılmıştır. Çalışmada TBMTK, 2FM ve MTTK modellerinden elde edilen RMSE değerleri arasında 2FM ve MTTK yöntemleri lehine istatistiksel olarak anlamlı fark bulunmuştur. Bu bağlamda eşitlenecek olan testlerde madde takımlarının olup olmamasına göre uygun koşulların, yöntemlerin ve modelin seçilmesi önerilmektedir.
- Testteki madde takım sayısı arttıkça TBMTK'da RMSE değerlerinin arttığı, MTTK'da ise RMSE değerlerinin düştüğü gözlenmiştir. 2FM diğer iki yöntemle göre madde takımı sayısına daha dirençli tepki vermiştir. Çalışmacıların testteki madde takımı sayısını göz önünde bulundurup eğer madde takım sayısı fazlaysa TBMTK yerine 2FM ve MTTK modellerini kullanabilir.
- Testteki bağımsız madde sayısı arttıkça TBMTK'da RMSE değerlerinin düştüğü, 2FM ve MTTK'da ise arttığı görülmüştür. Çalışmacıların testteki bağımsız madde sayısını ve madde takım sayısına oranını göz önünde bulundurarak kullanacakları modeli belirlemesi önerilebilir.
- Örneklem büyüklüğü arttıkça birçok koşulda ölçek dönüştürme yöntemlerinden elde edilen eşitleme hatalarının azaldığı bulgusuna ulaşılmıştır. 2FM ve MTTK gibi ikincil boyutu dikkate alan modellerde örneklem büyüklüğünün etkisi daha fazla görülmüştür. Kitapçıları alan tüm öğrencilerin çalışmaya dahil edildiği 7946 örnekleme hatalar bazı koşullarda 1000 örneklem ile benzer veya daha hatalı

sonuçlar vermiştir. Bu nedenle bu modellerin kullanıldığı çalışmalarda çalışmanın amacına ve koşullara göre örneklem seçimi yapılması tavsiye edilebilir.

Yeni Araştırmalara Yönelik Öneriler

Çalışmanın sonucunda ortaya çıkan bulgular ışığında benzer konularda yapılacak yeni araştırmalara ilişkin verilebilecek öneriler aşağıda sıralanmıştır.

- Bu araştırmada TIMSS 2019'dan elde edilen gerçek veriler kullanılmıştır. Gelecekte yapılacak araştırmalarda simülasyon üzerinden ölçek dönüştürme işlemleri yürütülerek koşullara göre elde edilen sonuçlar karşılaştırılabilir.
- İki faktör ve madde takımı kuramı modelleri için literatürde 2PLM parametre kestirim modellerinin daha az hatalı sonuçlar verdiği belirtilmiş bu nedenle bu çalışmada 2PL modeller kullanılmıştır. Yapılacak çalışmada 1PLM, 3PLM ve Rasch modelleri kullanılarak karşılaştırmalar yapılabilir.
- Çalışmada örneklem büyüklüğü koşulu 1000 ve 7946 olarak araştırma yürütülmüş ve örneklem büyüklüğü 7946 olduğunda ölçek dönüştürme hatalarının genel olarak düştüğü gözlenmiştir. Gelecekte yapılacak araştırmalarda daha küçük ve daha büyük örneklem büyüklüklerinde yapılacak ölçek dönüştürme işlemlerinin hataları karşılaştırılabilir.
- Bu çalışmada OO, OS, SL ve ve Haebara ölçek dönüştürme yöntemleri kullanılmıştır. Madde takımı tabanlı testlere farklı ölçek dönüştürme yöntemleri uygulanarak elde edilecek sonuçlar karşılaştırılabilir.
- Bu çalışmada, gerçek veriler kullanıldığı için madde takım sayısı ortak maddelerde 3, testin genelinde 4 ile sınırlanmıştır. Benzer çalışma daha fazla madde takımı içeren testlerle yapılabilir.
- Çalışmada gerçek veriler kullanıldığı için madde takımı uzunlukları 3, 4, 5 ve 8 ile sınırlıdır. Başka çalışmalarda daha kısa ve daha uzun madde takımları kullanılarak modeller karşılaştırılabilir.

- Çalışmada bağımsız madde sayısı da gerçek testin yapısından dolayı sınırlıdır. Bağımsız madde sayısının değiştiği farklı durumlarda madde takımları içeren testlerin kullanılan modellere olan tepkisi ölçülebilir.
- Araştırmada değerlendirme ölçütü olarak RMSE kullanılmıştır. Yeni araştırmalarda yanlılık gibi farklı teknikler kullanılarak sonuçlar kıyaslanabilir.

Kaynaklar

- Ackerman, T. A. (1987). *ACT research report series: The robustness of LOGIST and BILOG IRT estimation programs to violations of local independence*. PsycEXTRA Dataset. <https://doi.org/10.1037/e426132008-001>
- Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Educational Testing Service.
- Atar, B., & Yeşiltaş, G. (2017). Investigation of the performance of multidimensional equating procedures for common-item nonequivalent groups design. *Journal of Measurement and Evaluation in Education and Psychology-EPOD*, 8(4).
- Bao, H. (2007). *Investigating differential item function amplification and cancellation in the application of item response testlet models* [Unpublished doctoral dissertation]. University of Maryland.
- Babcock, B. G. E. (2009). *Estimating a Noncompensatory IRT Model Using a modified Metropolis algorithm* [Unpublished doctoral dissertation]. The University of Minnesota.
- Baker, F. (2001). *The basics of item response theory*. Retrieved from <http://ericae.net/irt/baker>
- Baker, F. B., & Al-Karni, A. (1991). A comparison of two procedures for computing IRT equating coefficients. *Journal of Educational Measurement*, 28(2), 147-162. <https://doi.org/10.1111/j.1745-3984.1991.tb00350.x>
- Baker, F. B., & Kim, S. H. (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). Taylor and Francis.
- Baldwin, S. G. (2007). A Review of Testlet Response Theory and Its Applications. *Journal of Educational and Behavioral Statistics*, 32(3), 333-336. <https://doi.org/10.3102/1076998607305834>
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395-479). Addison-Wesley.

- Bishop, N. S., & Omar, M. H. (2002). *Comparing vertical scales derived from dichotomous and polytomous IRT models for a test composed of testlets*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443-459. <https://doi.org/10.1007/bf02293801>
- Bolt, D. M. (1999). Evaluating the effects of multidimensionality on IRT true-score equating. *Applied Measurement in Education*, 12(4), 383-407. https://doi.org/10.1207/S15324818AME1204_4
- Bökeoğlu, Ö. Ç. (2022). Madde tepki kuramına dayalı gerçek puan eşitlemede ölçek dönüştürme yöntemlerinin incelenmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 55(1), 1-36, <https://doi.org/10.30964/auebfd.1001128>
- Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, 64(2), 153-168. <https://doi.org/10.1007/bf02294533>
- Brennan, R. L. (2001). An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement*, 38(4), 295-317.
- Brossman, B. G., & Lee, W. C. (2013). Observed score and true score equating procedures for multidimensional item response theory. *Applied Psychological Measurement*, 37(6), 460-481.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. The Guilford Press.
- Cai, L., du Toit, S. H. C., & Thissen, D. (2011). *IRTPRO: User guide*. Lincolnwood, IL: Scientific Software International.
- Cao, Y., Lu, R., & Tao, W. (2014). *Effect of item response theory (IRT) model selection on testlet-based test equating* (ETS Research Report No. RR-14-19). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/ets2.12017>

- Chen, J. (2014). *Model selection for IRT equating of Testlet-based tests in the random groups design* [Doctoral dissertation, University of Iowa]. <https://doi.org/10.17077/etd.9yfctk57>
- Chen, F. F., West, S., & Sousa, K. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41, 189–225. https://doi.org/10.1207/s15327906mbr4102_5
- Cook, L. L., & Eignor, D. R. (1991). IRT equating methods. *Educational Measurement: Issues and Practice*, 10(3), 37-45. <https://doi.org/10.1111/j.1745-3992.1991.tb00207.x>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth: Holt, Rinehart and Winston
- Davey, T., Oshima, T. C., & Lee, K. (1996). Linking Multidimensional Item Calibrations. *Applied Psychological Measurement*, 20(4), 405-416. <https://doi.org/10.1177/014662169602000407>
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. The Guilford Press.
- de la Torre, J., & Hong, Y. (2010). Parameter estimation with small sample size: A higher-order IRT model approach. *Applied Psychological Measurement*, 34, 267-285.
- DeMars, C. E. (2006). Application of the Bi-factor multidimensional item response theory model to testlet-based tests. *Journal of Educational Measurement*, 43(2), 145-168. <https://doi.org/10.1111/j.1745-3984.2006.00010.x>
- DeMars, C. (2016). *Madde Tepki Kuramı*. Ankara: Nobel Akademi Yayıncılık.
- Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polytomous item response models and standardized indices. *The British Journal of Mathematical Psychology*, 38, 67-86
- Dresher, A. R. (2002). *The examination of local item dependency of NAEP assessments using the testlet model* [Unpublished doctoral dissertation]. University of Pittsburgh.

- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.
- Gibbons, R. D., & Hedeker, D. (1992). Full-information item bi-factor analysis. *Psychometrika*, 57, 423 –436.
- Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., & Brenwald, S. (2009). *Highlights from TIMSS 2007: Mathematics and science achievement of U.S. fourth- and eighth-grade students in an international context (NCES 2009-001 revised)*. <http://ies.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001>
- Gök, B. (2012). *Denk olmayan gruplarda ortak madde deseni kullanılarak madde tepki kuramına dayalı eşitleme yöntemlerinin karşılaştırılması* [Unpublished doctoral dissertation]. Hacettepe Üniversitesi, Ankara.
- Gök, B., & Kelecioğlu, H. (2014). Comparison of IRT equating methods using the common-item nonequivalent groups design. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 10(1), 120-136.
- Gül, E., Doğan-Gül, Ç., Çokluk-Bökeoğlu, Ö., & Özkan, M. (2017). Temel eğitimden ortaöğretime geçiş matematik alt testi asıl sınav ve mazeret sınavlarının madde tepki kuramına göre eşitlenmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 17(4), 1900-1915. <https://dergipark.org.tr/tr/pub/aibuefd/issue/32772/363973>
- Gübeş, N. Ö., & Kelecioğlu, H. (2015). Karma testlerin eşitlenmesinde MTK eşitleme yöntemlerinin eşitlik özelliği korunumu ölçütüne göre karşılaştırılması. *Journal of Measurement and Evaluation in Education and Psychology*, 6(1), <https://doi.org/10.21031/epod.65039>
- Gübeş, N. Ö., & Kelecioğlu, H. (2016). The impact of test dimensionality, common-item set format, and scale linking methods on mixed-format test equating. *Educational Sciences: Theory & Practice*, 16(3).

- Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, 22(3), 144-149.
- Hambleton, R. K., & Traub, R. E. (1973). Analysis of empirical data using two parameter logistic latent trait models. *British Journal of Mathematics and Statistical Psychology*, 26, 195-211.
- Hambleton, R. K., & Van der Linden, W. J. (1982). Advances in item response theory and applications: An introduction. *Applied Psychological Measurement*, 6(4), 373-378.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Springer Science & Business Media.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Han, K. T. (2009). IRTEQ: Windows application that implements item response theory scaling and equating. *Applied Psychological Measurement*, 33(6), 491-493.
- Hanson, B. A., & Beguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, 26(3), 3-24.
- Huiqin Hu, Rogers, W. T., & Vukmirovic, Z. (2008). Investigation of IRT-Based Equating Methods in the Presence of Outlier Common Items. *Applied Psychological Measurement*, 32(4), 311-333. <https://doi.org/10.1177/0146621606292215>
- Kilmen, S. (2010). *Madde tepki kuramına dayalı test eşitleme yöntemlerinden kestirilen eşitleme hatalarının örneklem büyüklüğü ve yetenek dağılımına göre karşılaştırılması* [Yüksek Lisans Tezi], Ankara Üniversitesi.
- Kim, S., & Kolen, M. J. (2006). Robustness to format effects of IRT linking methods for mixed format tests. *Applied Measurement in Education*, 19(4), 357-381.

- Koziol, N. A. (2016). Parameter recovery and classification accuracy under conditions of testlet dependency: a comparison of the traditional 2PL, testlet, and bi-factor models. *Applied Measurement in Education*, 29, 184-195. <https://doi.org/10.1080/08957347.2016.1171767>
- Kogar, E. Y. (2016). *Madde takımları içeren testlerde farklı modellerden elde edilen madde ve yetenek parametrelerinin karşılaştırılması* [Yayımlanmamış Doktora Tezi]. Hacettepe Üniversitesi, Ankara.
- Kogar, E. Y., & Kelecioğlu, H. (2017). Examination of different item response theory models on tests composed of testlets. *Journal of Education and Learning*, 6(4), 113. <https://doi.org/10.5539/jel.v6n4p113>
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices*. Springer Science & Business Media.
- Lee, G., Kolen, M. J., Frisbie, D. A., & Ankenmann, R. D. (2001). Comparison of dichotomous and polytomous item response models in equating scores from tests composed of testlets. *Applied Psychological Measurement*, 25(4), 357-372. <https://doi.org/10.1177/01466210122032226>
- Li, Y., Bolt, D. M., & Fu, J. (2006). A comparison of alternative models for testlets. *Applied Psychological Measurement*, 30(1), 3-21. <https://doi.org/10.1177/0146621605275414>
- Lim, S. (2020). Review: A course in item response theory and modeling with stata, and using r for item response theory model applications. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(4), 657-659.
- Lord, F. M. (1952). *A theory of test scores* (Psychometric Monograph No. 7). Iowa City, IA: Psychometric Society.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

- Loyd, B. H., & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, 179-193.
- Marais, I. D., & Andrich, D. (2008). Effects of varying magnitude and patterns of local dependence in the Rasch model. *Journal of Applied Measurement*, 9, 105-124.
- Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems 1. *ETS Research Bulletin Series*, 1977(1), i-41.
- Martin, J., Mccaughtry, N., Hodges-Kulinna, P., & Cothran, D. (2008). The influences of professional development on teachers' self-efficacy toward educational change. *Physical Education & Sport Pedagogy*, 13(2), 171–190. <https://doi.org/10.1080/17408980701345683>
- Meng, Y. (2012). *Comparison of kernel equating and item response theory equating methods* (Publication No. 3518262) [Doctoral dissertation, University of Massachusetts, Amherst]. ProQuest Dissertations and Theses Global.
- Min, S., & He, L. (2014). Applying unidimensional and multidimensional item response theory models in testlet-based reading assessment. *Language Testing*, 31(4), 453-477. <https://doi.org/10.1177/0265532214527277>
- Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24(1), 50-64.
- Raykov, T., & Marcoulides, G. A. (2018). *A course in item response theory and modeling with Stata*. Stata Press.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.
- Reise, S. P., Morizot, J., & Hays, R. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16, 19–31. <https://doi.org/10.1007/s11136-007-9183-7>

- Rijmen, F. (2009). Three multidimensional models for testlet-based tests: Formal relations and an empirical comparison. *ETS Research Report Series*, 2009(2).
<https://doi.org/10.1002/j.2333-8504.2009.tb02194.x>
- Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28, 237–247.
- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7(2), 201-210.
<https://doi.org/10.1177/014662168300700208>
- Stout, W. F. (1987). A nonparametric approach for assessing latent trait dimensionality. *Psychometrika*, 52, 589-617.
- Tan, Ş. (2005). Küçük örneklerde beta4 ve polynomial loglineer ön düzgünleştirme ve kübik eğri son düzgünleştirme metotlarının uygunluğu. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 35(1), 123-151.
- Tate, R. (2000). Performance of a proposed method for the linking of mixed format tests with constructed response and multiple choice items. *Journal of Educational Measurement*, 37(4), 329-346.
- TIMSS (2019). “TIMSS 1999 Üçüncü Uluslararası Matematik ve Fen Bilgisi Çalışmaları Ulusal Rapor”, Milli Eğitim Bakanlığı, EARGED.
- Xu, T. (2016). *Equating with local dependence under the anchor test design* [Doctoral dissertation], University of Pittsburgh.
- Uyar, Ş., Aksekioğlu, B., & Öztürk Gübeş, N. (2018). PISA 2012 Matematik Okuryazarlığı Testinde Farklı Ölçek Dönüştürme Yöntemlerinin Karşılaştırılması. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi* (46), 121-148.
<https://doi.org/10.21764/maeuefd.330613>
- Uysal, İ. (2014). *Madde tepki kuramına dayalı test eşitleme yöntemlerinin karma modeller üzerinde karşılaştırılması* [Yüksek Lisans Tezi]. Abant İzzet Baysal Üniversitesi, Bolu.

- van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory*. Springer.
- Wainer, H., & Wang, X. (2000). Using a new statistical model for testlets to score TOEFL. *Journal of Educational Measurement*, 37(3), 203–220
- Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. Cambridge University Press.
- Yao, L. (2004). LinkMIRT: Linking of multivariate item response model [Computer software]. Monterey, CA: Defense Manpower Data Cent
- Yen, W. M. (1993). Scaling performance assessments Strategies for managing local item dependence. *Journal of Educational Measurement*, 30, 187–213.
- Yen, W. M., & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 111-153).
- Zhang, J. (2007). *Dichotomous or polytomous model? Equating of testlet-based tests in light of conditional item pair correlations* [Doctoral dissertation], University of Iowa, IA.
- Zhang, Z. (2010). *Comparison of different equating methods and an application to link testlet-based tests* [Doctoral dissertation], Chinese University of Hong Kong.
- Zhang, J. (2012). Calibration of Response Data Using MIRT Models with Simple and Mixed Structures. *Applied Psychological Measurement*, 36(5), 375-398. <https://doi.org/10.1177/0146621612445904>
- Zhao, Y. (2008). *Approaches for addressing the fit of item response theory models to educational test data* [Doctoral dissertation]. University of Massachusetts Amherst.
- Zor, Y. M. (2023). Investigation of multidimensional scale transformation methods applied to multidimensional tests according to various conditions. *Adiyaman University Journal of Educational Sciences*, 13(1), 41-53. <https://doi.org/10.17984/adyuebd.1239198>

EK-A: Araştırmada Yer Alan Koşullara İlişkin Elde Edilen RMSE Değerleri

ÖRN	SET	MODEL	OO			OS			SL			HA		
			a	c	θ	a	c	θ	a	c	θ	a	c	θ
1000	S1	TBMTK	0.02	0.07	0.09	0.10	0.26	0.33	0.02	0.11	0.09	0.09	0.14	0.07
1000	S1	2FM	0.16	0.42	0.40	0.57	0.70	0.70	0.07	0.11	0.11	0.06	0.20	0.11
1000	S1	MTTK	0.16	0.76	0.94	0.59	0.76	0.74	0.10	0.34	0.44	0.21	0.47	0.47
1000	S2	TBMTK	0.00	0.02	0.01	0.03	0.06	0.07	0.03	0.06	0.03	0.02	0.04	0.02
1000	S2	2FM	0.10	0.41	0.44	0.35	0.48	0.19	0.06	0.10	0.05	0.02	0.06	0.03
1000	S2	MTTK	0.09	0.35	0.41	0.27	0.42	0.20	0.05	0.14	0.06	0.03	0.07	0.03
1000	S3	TBMTK	0.30	1.09	1.39	0.12	0.24	0.10	0.17	0.70	0.87	0.23	0.70	0.88
1000	S3	2FM	0.06	0.54	0.60	0.79	1.05	1.10	0.30	0.69	0.24	0.38	0.92	0.32
1000	S3	MTTK	0.08	0.52	0.62	0.38	0.54	0.50	0.04	0.19	0.13	0.03	0.09	0.12
1000	S4	TBMTK	0.16	0.41	0.48	0.09	0.16	0.07	0.01	0.17	0.16	0.01	0.11	0.10
1000	S4	2FM	0.04	0.27	0.30	0.24	0.34	0.17	0.08	0.12	0.10	0.01	0.07	0.06
1000	S4	MTTK	0.06	0.31	0.36	0.23	0.36	0.18	0.04	0.11	0.06	0.01	0.05	0.03
7946	S1	TBMTK	0.08	0.12	0.12	0.10	0.29	0.37	0.06	0.11	0.13	0.09	0.16	0.18
7946	S1	2FM	0.14	0.25	0.27	0.12	0.39	0.50	0.04	0.09	0.11	0.06	0.14	0.05
7946	S1	MTTK	0.06	0.18	0.08	0.01	0.21	0.20	0.07	0.24	0.30	0.05	0.25	0.31
7946	S2	TBMTK	0.07	0.11	0.07	0.02	0.08	0.07	0.00	0.05	0.05	0.02	0.06	0.07
7946	S2	2FM	0.10	0.17	0.08	0.02	0.11	0.09	0.02	0.08	0.05	0.00	0.05	0.05
7946	S2	MTTK	0.05	0.15	0.08	0.05	0.15	0.08	0.03	0.13	0.09	0.02	0.11	0.09
7946	S3	TBMTK	0.22	0.84	1.05	0.18	0.29	0.14	0.11	0.50	0.62	0.11	0.46	0.57
7946	S3	2FM	0.10	0.17	0.12	0.02	0.15	0.12	0.01	0.02	0.01	0.02	0.11	0.13
7946	S3	MTTK	0.03	0.18	0.13	0.03	0.18	0.13	0.09	0.23	0.29	0.09	0.24	0.31
7946	S4	TBMTK	0.08	0.26	0.31	0.08	0.17	0.09	0.02	0.14	0.15	0.01	0.12	0.13
7946	S4	2FM	0.07	0.14	0.06	0.04	0.12	0.08	0.01	0.06	0.04	0.03	0.06	0.05
7946	S4	MTTK	0.04	0.15	0.10	0.06	0.16	0.08	0.01	0.10	0.11	0.01	0.09	0.09


Örn: Örneklem büyüklüğü, Set: Veri Setleri, Model: Parametre kestirim modelleri

OO: Ortalama-ortalama, OS: Ortalama-sigma, SL: Stocking- Lord, HA: Haebara

**EK-B: Arařtırma Etik Komisyon İzin Muafiyeti Formu/ Arařtırma Etik Komisyonu Onay
Bildirimi**



T.C.
HACETTEPE ÜNİVERSİTESİ
Rektörlük

Tarih: 02/06/2020
Sayı: 35853172-300-E.00001097851

0001097851

Sayı : 35853172-300
Konu : Ertunç UKŞUL (Etik Komisyon İzni)

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 21.04.2020 tarihli ve 51944218-300/00001080378 sayılı yazı.

Enstitünüz Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı Doktora programı öğrencisi Ertunç UKŞUL'un Prof. Dr. Hülya KELECİOĞLU danışmanlığında yürüttüğü "Madde Takımı Tabanlı Testlerin Eşitlenmesinde Madde Tepki Kuramı ve Madde Takımı Tepki Kuramı Modellerinin Karşılaştırılması" başlıklı tez çalışması Üniversitemiz Senatosu Etik Komisyonunun 12 Mayıs 2020 tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

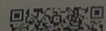
Bilgilerinizi ve gereğini saygılarımla rica ederim.

e-İmzalıdır
Prof. Dr. Rahime Meral NOHUTCU
Rektör Yardımcısı

Evrakın elektronik imzalı suretine <https://belgedogrulama.hacettepe.edu.tr> adresinden f7285ee8-e99d-4e94-a283-9301679154b0 kodu ile erişebilirsiniz.
Bu belge 5070 sayılı Elektronik İmza Kanunu'na uygun olarak Güvenli Elektronik İmza ile imzalanmıştır.

Hacettepe Üniversitesi Rektörlük 06100 Sıhhiye-Ankara
Telefon: (312) 305 3001-3002 Faks: (312) 311 9992 E-posta: yazimd@hacettepe.edu.tr İnternet

Sevda TOPAİ



EK-C: Etik Beyanı

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında,

- * tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- * görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- * başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- * atıfta bulunduğum eserlerin bütününe kaynak olarak gösterdiğimi,
- * kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- * bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

12/02/2024

Ertunç UKŞUL

EK-Ç: /Doktora Tez Çalışması Orijinallik Raporu

12/02/2024

HACETTEPE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı: Madde Takımı Tabanlı Testlerde Ölçek Dönüştürme Hatalarının İncelenmesi

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
08/02 /2024	103	129826	02/02/2024	%9	2289811849

Uygulanan filtreler:

- Kaynaklar hariç
- Alıntılar dâhil
- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esaslarını inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

Ad Soyadı: Ertunç UKŞUL

Öğrenci No.: N16146140

Ana Bilim Dalı: Eğitim Bilimleri

Programı: Eğitimde Ölçme ve Değerlendirme

Statüsü: Y.Lisans Doktora Bütünleşik Dr.

İmza

DANIŞMAN ONAYI

UYGUNDUR.

(Prof. Dr. Hülya KELECİOĞLU)

EK-D: Thesis/Dissertation Originality Report

12/02/2024

HACETTEPE UNIVERSITY
Graduate School of Educational Sciences
To The Department of Educational Sciences

Thesis Title: INVESTIGATION OF SCALE CONVERSION ERRORS IN TESTLET-BASED TESTS

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
08/02/2024	103	129826	02/02/2024	9%	2289811849

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Name Lastname: Ertunç UKŞUL

Student No.: N16146140

Department: Educational Sciences

Program: Measurement and Evaluation in Education

Status: Masters Ph.D. Integrated Ph.D.

Signature

ADVISOR APPROVAL

APPROVED
(Prof. Dr. Hülya KELECİOĞLU)

EK-E: Yayınlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge**" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü/Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren ... ay ertelenmiştir. ⁽²⁾
- Tezimle ilgili gizlilik kararı verilmiştir. ⁽³⁾

12 /02 /2024

(imza)

Ertunç UKŞUL

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

- (1) Madde 6.1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezinerişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6.2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internette paylaşılması durumunda 3 şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7.1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir
*Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

