

**ÇOK BOYUTLU KARMA-FORMAT TESTLERİN  
ÖLÇEKLENMESİNİ ETKİLEYEN FAKTÖRLERİN  
İNCELENMESİ**

**AN INVESTIGATION OF THE FACTORS AFFECTING THE  
SCALING OF MULTIDIMENSIONAL MIXED FORMAT  
TESTS**

**Akif AVCU**

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı İçin

Öngördüğü

Doktora Tezi

olarak hazırlanmıştır.

2016

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼'ne,

Akif AVCU'nun hazırladıđı "Çok Boyutlu Karma-Format Testlerin Ölçeklenmesini Etkileyen Faktörlerin İncelenmesi" başlıklı bu çalışma jürimiz tarafından Eđitim Bilimleri Anabilim Dalı, Eđitimde Ölçme ve Deđerlendirme Bilim Dalı'nda Yüksek Lisans/Doktora Tezi olarak kabul edilmiřtir.

Başkan Prof. Dr. Selahattin GELBAL



Üye (Danıřman) Prof. Dr. Hülya KELECIOđLU



Üye Doç. Dr. Nilüfer KAHRAMAN



Üye Doç. Dr. Burcu ATAR



Üye Doç. Dr. Neře GÜLER



ONAY

Bu tez Hacettepe Üniversitesi Lisansüstü Eđitim-Öđretim ve Sınav Yönetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından...../...../..... tarihinde uygun gör¼lmüş ve Enstitü Yönetim Kurulunca...../...../.....tarihinde kabul edilmiřtir.

Prof. Dr. Berrin AKMAN  
Eđitim Bilimleri Enstitüsü M¼d¼r¼

# ÇOK BOYUTLU KARMA-FORMAT TESTLERİN ÖLÇEKLENMESİNİ ETKİLEYEN FAKTÖRLERİN İNCELENMESİ

**Akif AVCU**

## ÖZ

Bu çalışma kapsamında karma format maddelerden oluşan bir, iki ve üç boyutlu testler DOGOM (Denk Olmayan Gruplarda Ortak Madde) deseni kullanılarak ölçeklendiğinde ortak madde setinin yapısı (yalnızca iki kategorili maddelerden oluşan ortak madde seti - iki ve çok kategorili maddelerin yer aldığı ortak madde seti), yetenek daralması (üst yetenek grubunda yetenek varyansının daralması - varyansın eşit kalması) ve parametre kestirim yöntemlerinin (EM - MHRM) ölçekleme sonuçları üzerindeki etkisi incelenmiştir. Ayrıca bu koşulların etkileşim içinde olup olmadığına bakılmıştır.

Çalışma, türetilmiş veriler kullanılarak gerçekleştirilmiştir. Ölçeklemenin niteliğinin değerlendirilmesinde ölçme hatası ve yanlılık değerleri kullanılmıştır. Veriler türetilirken yanıt matrisleri, içerisinde İKM (iki kategorili madde) ve ÇKM(çok kategorili madde)'ler yer alacak şekilde oluşturulmuştur. İKM'ler için parametre kestirimi 3 parametrelili modele (3PLM) göre, ÇKM'ler için ise aşamalı tepki modeline (ATM) göre gerçekleştirilmiştir. Veri türetme ve analizi sürecinde gerçekleştirilen işlem 50 defa tekrarlanmıştır. Ayrıca, araştırmada gerçekleştirilen veri türetme, testlerin kalibrasyonu ve ölçekleme işlemleri için R programı kullanılmıştır. Etkileşimleri incelemek için kullanılan iki ve üç yönlü analizler SPSS ile gerçekleştirilmiştir.

Araştırmada sonucunda ortak madde yapısının ölçekleme işlemi sonucunda ortaya çıkan hata ve yanlılık miktarını önemli ölçüde etkilediği görülmüştür. Buna göre karma format testlerde ortak madde setinin sadece İKM'lerden oluşması ölçekleme hatasını bazı istisnalar haricinde arttırmaktadır. Elde edilen bu bulgu, araştırmada kullanılan testlerin boyut sayısından bağımsız olarak kendini göstermektedir.

Varyans daralmasının etkisi incelendiğinde yetenek parametresi ve ÇKM'lere ait  $a$  parametreleri için farklılaşmalar olduğu görülmüştür. Bu farklılaşmalar iki boyutlu testler için hata ve üç boyutlu testler için ise yanlılık değerlerinde gözlenmiştir. ÇKM'lerin  $a$  parametreleri için ise üç boyutlu testlerin ikinci boyutu için hata

değerlerinde farklılaşmalar olduğu bulunmuştur. Her iki parametre için varyansın azaldığı durumda daha iyi sonuçlar elde edildiği görülmüştür.

Kullanılan kestirim yönteminin etkisi incelendiğinde ise tek boyutlu verilerde yetenek parametresi için hata ve yanlılık değerlerinin EM kestirim yöntemi için daha az olduğu, iki ve üç boyutlu testlerde ise bazı boyutlar için yanlılık değerlerinin MHRM kestirim yöntemi için daha az olduğu görülmüştür. Ayrıca İKM'lerin  $a$  ve  $b$  parametreleri ve ÇKM'lerin eşik parametreleri için bazı durumlarda kestirim yönteminin hata ve yanlılık değerlerini etkilediği görülmüştür. Gözlenen bu farklılaşmalar testlerin boyut sayısından bağımsızdır. ÇKM'lerin  $a$  parametresinin ise kestirim yönteminden etkilenmediği görülmüştür.

Son olarak, etkileşimler incelenmiştir. Buna göre, yetenek parametresi için iki ve üç boyutlu testlerde bazı koşullara göre yanlılık değerlerinin ikişerli ve üçerli etkileşimler gösterdiği bulunmuştur. İKM'lere ait  $a$  ve  $b$  parametreleri için bakıldığında  $b$  parametresine ait hata ve yanlılık değerlerinde yalnızca üç boyutlu testlerin birinci boyutunda varyans daralması ve kestirim yöntemi değişkenlerinin etkileşim içinde oldukları görülmüştür. İKM'lere ait  $a$  parametrelerine ait hata değerleri için iki boyutlu testler ile üç boyutlu testlerin ilk boyutunda üç koşulun etkileşim içinde olduğu bulunmuştur. Ayrıca, ÇKM'lere ait  $a$  parametreleri ile eşik parametreleri için iki boyutlu ve üç boyutlu testlerde etkileşim gözlenmemiştir. Üç boyutlu testlerde üç boyutun her birisi için OMY ve KY koşulları arasında etkileşim olduğu görülmüştür.

Sonuç olarak, etkisi incelenen koşullar içinde ölçekleme sonuçları üzerinde en fazla etkisi olan koşulun ortak madde yapısı olduğu sonucuna varılmıştır. Ayrıca, kestirim yöntemine göre gözlenen etkilerin tek boyutlu testler ile iki-üç boyutlu testler için farklılaştığı sonucuna varılmıştır.

**Anahtar sözcükler:** Test ölçekleme, çok boyutluluk, karma format testler, eşitleme hatası, yanlılık, çok boyutlu madde tepki kuramı

**Danışman:** Prof. Dr. Hülya KELECİOĞLU, Hacettepe Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı

# **AN INVESTIGATION OF THE FACTORS AFFECTING THE SCALING OF MULTIDIMENSIONAL MIXED FORMAT TESTS**

**Akif AVCU**

## **ABSTRACT**

In this study, the effects of the composition of common item set (common item set with only dichotomous items – common item set with mixed format items), scale shrinkage (shrinkage of ability variance for higher ability group – same ability variance) and estimation methods (EM – MHRM) on vertical scaling conducted by using Common Item Non-Equivalent Group Design (CINEG) on unidimensional, two dimensional and three dimensional tests were investigated. In addition, possible interactions between these factors were examined.

The study was conducted using simulated data. To evaluate the quality of scaling across study conditions, RMSE and bias values were computed. The simulation process involved generating response data matrices composed of both dichotomous and polytomous items. Simulation data were later used to calibrate item parameters for the Dichotomous items using the three parameter logistic model (3PLM) while polytomous items using the graded response model (GRM). For data generation and analysis, each process was iterated 50 times. Software written in R program was utilized to generate data, to calibration parameters and scaling. Two and three way ANOVA's were conducted to investigate the possible interactions by using SPSS.

The results suggested that composition of common items largely affects both RMSE and bias values obtained from scaling process. Accordingly, using only dichotomous items in common item set increases RMSE and bias values even when there are some exceptions. This result is consistently observed irrespective of the number of dimensions in the dataset.

As to the effect of scale shrinkage, it was found that there are differences observed for the ability parameters and the  $a$  parameters of polytomous items. These observed differences, belongs to RMSE values for two dimensional tests and to bias values for three dimensional tests. There was found another difference for  $a$  parameter of polytomous items in second dimension of three dimensional tests in

terms of RMSE values. For both parameters, it was found that when the variance of ability parameter shrink, better scaling results were obtained.

When investigating the effect of estimation methods, it was found that for ability parameter, the RMSE and bias values were smaller when using the EM method in one dimensional tests while the MHRM method give smaller bias values in some dimensions of two and three dimensional tests. Additionally, it was found that for the the  $b$  parameters of dichotomous items and threshold parameters of polytomous items, the estimation method was found to have an effect on RMSE and bias values on some occasions. These differences were observed irrespective of the number of dimensions in the tests. On the other hand, it was found that estimation method have no effect on the  $a$  parameters of both dichotomous and polytomous items.

Finally, possible interactions were investigated. It was found that, on some occasions, for bias values of the ability parameters in two and three dimensional tests, two and three way interaction effects were observed. When the  $a$  and the  $b$  parameters of dichotomous items were considered, bias and RMSE values of the  $b$  parameters in three dimensional tests have interaction effect of scale shrinkage and estimation methods. There was three way interaction for the  $a$  parameters in three dimensional tests and first dimension of three dimensional tests. In addition, some interactions observed for the  $a$  and the threshold parameters of polytomous items in two and three dimensional tests. In all dimensions of three dimensional tests, common item composition and estimation methods show interaction.

As a result, common item composition was found to have biggest effect on scaling results in comparison to the other factors investigated. In addition, scale results for dimensional tests were found to be affected more from estimation method when compared with two and three dimensional tests.

**Keywords:** Test scaling, multidimensionality, mixed format tests, equating error, bias, multidimensional item response theory

**Advisor:** Prof. Dr. Hülya KELECİOĞLU, Hacettepe University, Department of Educational Science, Division of Educational Measurement and Evaluation

## ETİK BEYANNAMESİ

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

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

beyan ederim.

Akif AVCU



## TEŞEKKÜR

Doktora öğrenimim ve tez çalışmam süresince desteğini her zaman bana hissettiren ve bu çalışmanın sonuçlanmasında yardımlarını hiçbir zaman esirgemeyen danışmanım Prof. Dr. Hülya KELECİOĞLU'na en içten teşekkürlerimi sunuyorum.

Doktora öğrenimim boyunca değerli görüşlerine ve geribildirimlerine sürekli başvurduğum ve birikimiyle, akademik gelişimimde önemli bir emeği olan Prof. Dr. Selahattin GELBAL'a sonsuz şükran duygularımı ifade etmek isterim.

Doktora tez izleme komitesinde ve tez savunma sınavı jürimde yer alan Doç. Dr. Nilüfer KAHRAMAN'a geri bildirimleri ve yardımları için çok teşekkür ederim.

Tez savunma sınavı jürimde yer alan Doç. Dr. Neşe GÜLER'e ve Doç. Dr. Burcu ATAR'a geri bildirimleri ve yardımları için çok teşekkür ederim.

Doktora tez çalışmasına aynı dönemlerde devam ettiğimiz ve sürekli fikir alışverişinde bulunduğumuz değerli dostlarım Dr. Ömür KALKAN ve Uzm. Eren ÖZBERK'e desteği ve yardımları için çok teşekkür ederim.

Eğitim hayatım boyunca iyi dileklerini ve desteklerini hiç eksik etmeyen sevgili ailemin üzerimdeki emekleri ve destekleri için minnettarım.



## İÇİNDEKİLER

ÖZ.....	iii
ABSTRACT.....	v
ETİK BEYANNAMESİ .....	vii
TEŞEKKÜR.....	viii
İÇİNDEKİLER.....	ix
ÇİZELGELER DİZİNİ.....	xi
ŞEKİLLER DİZİNİ.....	xii
SİMGELER VE KISALTMALAR DİZİNİ .....	xiii
1. GİRİŞ.....	1
1.1. Problem Durumu .....	1
1.2. Araştırmanın Amacı ve Önemi .....	5
1.3. Problem Cümlesi .....	7
1.3.1. Alt Problemler .....	7
1.4. Sayıtlılar .....	7
1.5. Sınırlılıklar .....	7
1.6. Tanımlar .....	8
1.7. Araştırmanın Kuramsal Temeli .....	8
2. İLGİLİ ARAŞTIRMALAR.....	26
2.1. İlgili Araştırmalar Özet .....	34
3. YÖNTEM.....	36
3.1. Araştırmanın Türü .....	36
3.2. Eşitleme Deseni .....	36
3.3. Test Yapılandırması .....	36
3.4. Verilerin Türetilmesi.....	39
3.5. Parametrelerin Kestirilmesi.....	42
3.6. Ölçekleme İşleminin Gerçekleştirilmesi .....	43
3.7. Değerlendirme Ölçütleri.....	43
3.8. Verilerin Analizi.....	44
5.BULGULAR VE TARTIŞMA .....	46
4.1. I. Alt Probleme Ait Bulgular ve Yorumlar .....	47
4.1.1. Ortak Madde Yapısının Etkisi .....	47
4.1.2. Yetenek Daralmasının Etkisi .....	48
4.1.3. Kestirim Yönteminin Etkisi .....	48
4.1.4. Tek Boyutlu Veriler için Koşulların Etkisinin ve Etkileşimlerin İncelenmesi .....	49
4.2. II. Alt Probleme Ait Bulgular ve Yorumlar .....	52
4.2.1. Ortak Madde Yapısının Etkisi .....	52
4.2.2. Yetenek Daralmasının Etkisi .....	54
4.2.3. Kestirim Yönteminin Etkisi .....	54
4.2.4. İki Boyutlu Veriler için Koşulların Etkisinin ve Etkileşimlerin İncelenmesi .....	55

4.3. III. Alt Probleme Ait Bulgular ve Yorumlar .....	59
4.3.1. Ortak Madde Yapısının Etkisi .....	59
4.3.2. Yetenek Daralmasının Etkisi .....	60
4.3.3. Kestirim Yönteminin Etkisi .....	61
4.3.4. Üç Boyutlu Veriler İçin Koşulların Etkisinin ve Etkileşimlerin İncelenmesi .....	62
5. SONUÇ ve ÖNERİLER .....	67
5.1. Sonuçlar .....	67
5.2. Öneriler .....	72
5.3. Uygulayıcılara Yönelik Öneriler .....	72
5.4. Araştırmacılara Yönelik Öneriler .....	73
KAYNAKÇA .....	76
EKLER DİZİNİ .....	82
EK 1. ETİK KURUL İZİNİ MUAFİYET FORMU .....	83
EK 2. ORJİNALLİK RAPORU .....	84
ÖZGEÇMİŞ .....	85

## ÇİZELGELER DİZİNİ

Çizelge 3.1. Verilerin Türetilmesinde Kullanılan Kovaryans Matrisleri.....	38
Çizelge 3.2. Ortak Maddelerin Boyutlara ve Formata Göre Temsil Durumu.	39
Çizelge 3.3. Veri Setlerinin Türetilmesi İçin Kullanılan Evren $\alpha$ Parametrelerine Ait Betimsel İstatistikler .....	40
Çizelge 3.4. Türetilen Evren Güçlük (b) ve Eşik (d) Parametrelerine Ait Betimsel İstatistiksel.....	42
Çizelge 4.1. Tek Boyutlu Veri Yapısı için Koşullara Göre Parametrelerin Hata ve Yanlılık Değerleri .....	47
Çizelge 4.2. Tek Boyutlu Veriler için ANOVA Sonuçları .....	49
Çizelge 4.3. İki Boyutlu Veri Yapısı için Koşullara Göre Hata ve Yanlılık Değerleri.....	52
Çizelge 4.4. İki Boyutlu Veriler İçin ANOVA Sonuçları .....	55
Çizelge 4.5. Üç Boyutlu Veri Yapısı için Koşullara Göre Hata ve Yanlılık Değerleri.....	59
Çizelge 4.6. Üç Boyutlu Veriler için ANOVA Sonuçları.....	62

## ŞEKİLLER DİZİNİ

Şekil 1.1: İki Farklı Test Formuna Ait Test Karakteristik Eğri Örneği .....	24
Şekil 3.1. İki Kademeli Model Örneği.....	37
Şekil 3.2. Veri Setlerinde Yer alan Maddelerin Ait Oldukları Boyutlar.....	41
Şekil 4.1. Tek Boyutlu Veriler için Elde Edilen Hata ve Yanlılık Değerlerinin Her Bir Parametre için Koşullara Göre Karşılaştırması .....	51
Şekil 4.2. İki Boyutlu Veriler için Elde Edilen Hata ve Yanlılık Değerlerinin Her Bir Parametre için Koşullara Göre Karşılaştırması .....	58
Şekil 4.3. Üç Boyutlu Veriler için Elde Edilen Hata ve Yanlılık Değerlerinin Her Bir Parametre için Koşullara Göre Karşılaştırması .....	66



## SİMGELER VE KISALTMALAR DİZİNİ

**MTK:** Madde Tepki Kuramı

**ATM:** Aşamalı Tepki Modeli

**İKМ:** İki Kategorili Madde

**ÇKM:** Çok Kategorili Madde

**DOGOM:** Denk olmayan gruplarda ortak madde deseni

**OMY:** Ortak Madde Yapısı

**KY:** Kestirim Yöntemi

**VD:** Varyans Daralması

# 1. GİRİŞ

## 1.1. Problem Durumu

Günümüzde büyük ölçekli test uygulamalarında birden fazla forma gereksinim duyulabilir. Bu gereksinim test uygulamalarında aynı testi birden fazla kullanılması ve bu sebeple testlerin güvenliğinin sağlanması ile ilgilidir. Kullanılan bu farklı test formlarından elde edilen puanların karşılaştırılabilir olabilmesi için formlar arasında fonksiyonel bir bağ oluşturulması gerekmektedir.

Test bağlama (linking) farklı test formları arasındaki bu ilişkinin kurulması işlemidir. Test bağlama için test formları arasındaki içerik ve zorluk seviyelerinin aynı olması gibi bir şart bulunmamaktadır. Test eşitleme (equating) ise bağlamanın özel bir formudur ve amacı farklı test formları arasındaki puanların birbirleri yerine kullanılmasıdır. Bu sebeple, test formlarının içerik ve güçlük olarak benzer olması gerekmektedir. Burada amaç, formlar arasında farklılaşması muhtemel form zorluklarının ayarlanması ve/veya gerçekleştirilen istatistiksel işlem sonucunda bir test formundan elde edilen puanın diğer formda karşılık gelen puanla dönüşümlü olarak kullanılabilmesidir (Kolen ve Brennan, 2004, s.2).

Test eşitlemede eşitlenen test formları aynı istatistiksel ve içerik özelliklerine sahip olma varsayımı üzerine inşa edilmiştir ve bir formdaki puan diğer formdaki puanın yerine kullanılabilir. Ölçeklemede (vertical scaling) ise farklı test formları birbirlerine bağlandığı için eşitleme ile benzerdir. Fakat, ölçeklemede test formları içerik ve güçlük olarak farklıdır çünkü formlar sınıflar arası ya da yaşa bağlı olarak ilerlemeyi yansıtmaktadırlar. Ölçeklemede amaç bireylerin bir eğitim programında elde ettikleri kazanımların mümkün olan en az hata ile modellenmesidir (Smith ve Yen, 2006). Bu kazanımlar, hesap verebilirlik kaygılarının ön plana çıktığı günümüzde eğitim reformlarını revize etmede ve beklenen düzeyde kazanım elde edememiş öğrencilerin belirlenmesinde kullanılmaktadır.

Ölçekleme farklı test formlarının karşılaştırılması için kullanılmakla birlikte her bir seviyedeki puanlar birbirlerinin yerine kullanılamazlar. Test ölçeklemesinde temel amaç farklı seviyelerdeki puanların karşılaştırılması ve bu yolla kazanımların belirlenmesidir. Seviye farklılığı bir öğrencinin bulunduğu sınıf, eğitim öğretim yılının bulunduğu aşama ya da yaştan kaynaklanabilir (Kolen ve Brennan, 2004). Bu

çalıřmada puanlar arasında baę kurulacak gruplar arasında yetenek farkı olacaęından gerekleřtirilecek iřlem leklemedir.

Günümüzde lekleme iřlemi iin kullanılan testlerde sadece doęru-yanlıř řeklinde puanlanan iki kategorili maddelerin yanında ok kategorili puanlamaya uygun farklı formata sahip maddelerin de yer aldıęı grlmektedir. Nitekim oktan semeli maddeler 20.yy'da başarı testlerinde en yoęun tercih edilen madde formatı olmasına (Koretz ve Hamilton, 2006) raęmen aık ulu soruların da uygulanan testlerde yer aldıęı karma-format (mixed format) testler artan sıklıkla tercih edilmektedir. Nitekim Lane (2005, Akt. Cao, 2008)'in belirttięi gibi ABD'de gerekleřtirilen ve ęrencilerin kazanımlarını belirlemeyi amalayan eyalet bazındaki durum belirleme sınavlarının %63'nde karma format maddeler kullanılarak testler oluřturulmaktadır.

Livingston (2009)'a gre oktan semeli sorular geniř yelpazede ierięi yksek gvenirlikle lebilmektedir, kısa srede ve dřk maliyet ile deęerlendirilebilmektedir. Dięer taraftan aık ulu sorular ise st dzey biliřsel dřnme becerilerini daha etkili bir řekilde lmektedir fakat bu maddelerden oluřan lme araları daha dar ieriklere sahiptir, deęerlendirilmesi maliyetlidir ve subjektif olması muhtemeldir. Karma-format testler ise farklı biimlerdeki maddelerin bu zelliklerinden aynı lme aracında faydalanılmasını mmkn kılmaktadır.

lekleme iřlemlerinin, geliřimin modellenmesi iin kullanılması, Madde Tepki Kuramı (MTK)'nın geliřimi ile paralel olmuřtur. MTK, psikoloji, eęitim, test geliřtirme, psikometri ve istatistik gibi farklı birok alandaki geliřen fikirlerin birleřmesinin sonucunda ortaya ıkmıřtır. MTK'ya dayalı eřitleme yntemleri genellikle  basamaktan oluřmaktadır: Madde kalibrasyonu, lek dnřm ve eřitleme. Birinci basamakta madde parametreleri kalibre edilir. İkinci basamakta her iki form iin kestirilen madde parametreleri ortak bir leęe baęlanır ve nc basamakta ise gerek puan ya da gzlenen puan yntemi kullanılarak eřitleme gerekleřtirilir. Uygulamada, MTK modelleri farklı formatta maddelerin olduęu karma-format testlerin kalibrasyonu ve leklenmesi iin de kullanılabilir. Nitekim

Testlerle ilgili bir dięer konu ise ok boyutluluktur. Bireylerin kazanımlarını lmek iin geliřtirilen testlerin ok boyutlu yapı gstermesi mmkndr. Geleneksel olarak ok boyutluluęun kaynaęı olarak test maddelerin ierik olarak llen yapının farklı alt alanlardan oluřması olduęu bilinmektedir. Fakat karma format testlerde

maddenin formatının da ayrı bir boyutluluk kaynağı olması mümkündür. Madde formatının başlı başına bir boyutluluk kaynağı olup olmadığı belirlenebilmesi için uygulanacak tekniklerden biri faktör analizidir. Bennett, Rock ve Wang (1991), Wainer ve Thissen (1993), Pollock ve Rock (1997) ve Lukhele, Thissen ve Wainer (1994) tarafından gerçekleştirilen çalışmalarda faktör analizi kullanılarak maddelerin formatına göre testin boyutlara ayrılıp ayrılmadığı incelenmiştir. Ulaşılan bulgulara göre madde formatının boyutluluğu etkileyen bir faktör olmakla birlikte tek başına boyutluluğa sebep olan bir etken olduğu varsayımı doğrulanmamıştır.

Ölçekleme çalışmalarında dikkate alınması gereken az araştırılmış konulardan birisi de yetenek daralmasıdır. Yetenek daralması, genellikle gelişim ve kazanımların ölçüldüğü testlerle gerçekleştirilen ölçekleme uygulamalarında test puanının varyansının ve puan aralıklarının ikinci uygulamada, ilk uygulamaya göre azalması durumudur (Yen, 1985). İkinci uygulamada yetenek dağılımlarına ait varyansın daralması, üst grupta yer alan bireylerin aldıkları eğitimin etkisiyle birbirlerine daha benzer edim düzeyleri ulaşmasından kaynaklanmaktadır. Ölçeklemenin öğrenci gelişiminin ve kazanımlarının modellenmesini amaçladığı düşünüldüğünde modellenen bu kazanımların, yani yeteneğin, daralması da muhtemeldir. Yani ölçekleme işleminin kullanıldığı durumlarda yetenek daralmasının gözlemlendiği şartlar tamamen aynıdır. Fakat yetenek daralmasının ölçekleme işlemini ne ölçüde etkilediği ile ilgili ilgili alan-yazında bir çalışmaya rastlanmamıştır.

Test puanlarının bağlanması konusunda güncel alan-yazında dört farklı alana odaklandığı görülmektedir (Kim ve Lee, 2006):

- a) Güncel ölçekleme yöntemlerinin teknik olarak gözden geçirilmesi ve bu yöntemlerin karma-formatlı testlerdeki ölçeklemelere uyarlanması,
- b) Farklı ölçme yöntemlerinin göreceli etkinliklerinin karşılaştırılması,
- c) Eşitleme sonuçları üzerindeki puanlayıcı etkisinin giderilmesi,
- d) Farklı durumların etkisinin incelenmesi (test uzunluğu, örneklem büyüklüğü, grup yetenek dağılımları, ortak maddelerin sayısı, farklı formatlardaki maddelerin toplam teste olan oranı), madde formatının etkisinden kaynaklanan boyutluluğun eşitleme sonuçları üzerindeki etkisinin incelenmesi,



Alan-yazında bu dört alanda sınıflanabilecek birçok çalışma yapıldığı görülmektedir. İlgili alan-yazında yer alan çalışmalar incelendiğinde bunların bir kısmında (Kim ve Lee, 2006; Kirkpatrick, 2005 ve Tate, 2000) karma-format testlerin tek formatlı testlerle karşılaştırıldığı görülmektedir. Karma format testlerle ilgili bazı çalışmalarda (Li, Lissitz ve Yang, 1999; He, 2011; Hagge, 2010) ise çok boyutluluk ele alınmamıştır. Karma-format testlerde çok boyutluluğun ele alındığı çalışmalarda ise puanlayıcı etkisinden kaynaklanan boyutluluğun (Kamata ve Tate, 2005) ya da format etkisinden kaynaklanan boyutluluğun (Kim ve Kolen, 2006) etkilerinin incelendiği görülmektedir. Fakat ilgili alan-yazında karma-format testlerde hem içerik hem de format etkisinden kaynaklanan boyutluluğun ve yetenek daralması gibi gerçek test uygulamalarında sıklıkla karşılaşılan bir durumun çok boyutlu testlerin ölçeklenmesi bağlamında incelendiği bir çalışmaya rastlanmamıştır. Gerçekleştirilecek bu çalışma ile ilgili alan-yazında belirtilen bu boşluğun doldurulacağı düşünülmektedir.

Ayrıca, Türkiye’de, önümüzdeki dönemde büyük ölçekli testlerin çok boyutlu ve karma-format yapılarda oluşturulma yöneliminin artacağı varsayılabilir. Nitekim Öğrenci Seçme ve Yerleştirme Merkezi (ÖSYM) tarafından 3 Kasım 2013 tarihinde Ankara ilinde 20 farklı okulda öğrenim gören ve 2648 öğrencinin katıldığı bir pilot uygulaması gerçekleştirilmiştir. Bu sınavda 30 adet açık uçlu ve 30 adet çoktan seçmeli olmak üzere 60 soru yer almıştır. Gerçekleştirilen bu sınavın içeriği YGS (Yükseköğretime Geçiş Sınavı) ile aynı tutularak Türkçe, Matematik, Fizik, Kimya, Biyoloji, Tarih, Coğrafya, Felsefe ve Din Kültürü ve Ahlak Bilgisi alanlarından sorulara yer verilmiştir. Bu pilot uygulamanın Türkiye’de karma-format testler ile gerçekleştirilen ilk ciddi uygulama olduğu söylenebilir. Önümüzdeki yıllarda bu yönde çalışmaların devam edeceği de yine ÖSYM tarafından belirtilmiştir (OSYM, 2016). Diğer taraftan, Türkiye’de ölçeklemenin kullanıldığı bir uygulama bulunmamakla birlikte Çetin (2009) tarafından İlköğretim Öğrencilerinin Başarılarının Belirlenmesi Sınavı (ÖBBS)’na ait 6., 7. ve 8. sınıf öğrencileri ve kullanılarak gerçekleştirilen dikey ve Altun (2013) tarafından Seviye Belirleme Sınavı (SBS)’na ait 6., 7. ve 8. sınıf öğrencileri kullanılarak gerçekleştirilen ölçekleme çalışmaları bulunmaktadır. Ölçeklemenin öğrenci gelişimini takip etmede ve eğitim politikalarını düzenlemede sağladığı faydalar dikkate alınacak olursa gelecekte ölçekleme uygulamalarından Türkiye’de de yararlanılabileceği

düşünülmektedir. Diğer taraftan, bu çalışmalarda kullanılan maddeler karma format değildir. Bu varsayımdan hareketle karma-formatlı testlerde ölçeklemenin incelendiği çalışmaların alanda çalışan uzmanlar için bu testlerin geliştirilmesi ve uygulanması sürecinde yararlı olacağı düşünülmektedir.

## **1.2. Araştırmanın Amacı ve Önemi**

Testlerden elde edilen puanlar birçok başlık altında alınan önemli kararlar için temel bilgi kaynakları arasındadır. Alınacak önemli kararlardan bağımsız olarak, test puanlarının mümkün olan en kesin bilgiyi sunması gerekmektedir. Daha kesin bilgi daha iyi kararların alınabilmesi için önemlidir (Kolen ve Brennan, 2004). Bununla birlikte uygulamada test güvenliği ve öğrenci gelişiminin takip edilebilmesi gibi birtakım gerekçeler yüzünden aynı testin farklı formları kullanılmakta veya farklı zamanlarda uygulanan testlerde ortak maddeler kullanılarak testler ölçeklenmektedir. Farklı formlardan elde edilen puanlar daha sonrasında eşitlenmekte ya da ölçeklenmektedir. Bu işlemin mümkün olduğunca az hata barındırması gerçekleştirilen sınavların daha adil olması ve öğrencilerin geleceği ile ilgili doğru kararlar verebilmek için önemlidir. Buna göre, ulusal çapta uygulanan geniş ölçekli testlere uygulanan ölçekleme yöntemlerinin psikometrik olarak savunulabilir olması önemlidir. Bu sebepten dolayı ölçekleme gerçekleştirilirken uygulayıcıların kararlarını dayandıracakları kuramsal çalışmalar büyük önem taşımaktadır.

İlgili alan-yazın incelendiğinde sadece çoktan seçmeli maddelerin karma-format testlerde ortak madde (anchor item) olarak kullanılmasının tek boyutlu testlerde daha iyi sonuçlar verdiği görülmüştür (Kim ve Lee, 2002). Fakat ortak madde setinin karma-format olduğu çok boyutlu testlerin ölçeklenmesi durumunda bu durumun ölçekleme sonuçlarını nasıl etkileyeceği ile ilgili yeterli araştırmaya rastlanmamıştır. Diğer taraftan, yetenek daralması da Camilli (1999) tarafından çok boyutlu testlerin ölçeklenmesi bağlamında ele alınmış ve bu konudaki ilgili alan-yazın kapsamlı olarak özetlenmiştir. Fakat yetenek daralmasının karma-format testlerin ölçeklenmesinde hata ve yanlışlık üzerinde nasıl bir etkisinin olacağı ile ilgili alan-yazında herhangi bir çalışmanın bulunmadığı görülmüştür.

Bu çalışmada ele alınacak bir diğer koşul ise ölçeklenecek test formları arasındaki ilişkinin yapısı olacaktır. Çok boyutlu karma-format testlerin ölçeklenmesi ile ilgili

çalıřmalarda boyutluluęun iki farklı kaynaęının olduęu grlmektedir. İerięe dayalı boyutluluk bu faktrlerden ilkidir. İerikten kaynaklı testlerde boyutluluk, testin birden ok yapıyı lmesinden kaynaklanmaktadır (Hagge, 2010). İkinici olarak ise maddelerin farklı formatta olmasından kaynaklanan boyutluluktur. Format etkisinden kaynaklı boyutluluęun da yine ierikteki boyutluluktan kaynaklandıęı savunulmaktadır (Thissen, Wainer, ve Wang, 1994). Thissen ve arkadaşlarına gre, formattaki farklılıklar farklı yapıların llmesine sebep olmaktadır. Yani, format etkisinin test boyutluluęunu daha karmařık hale getirdięi sylenebilir. Aık ulu maddelerin testlere eklenmesinin test boyutluluęunu nasıl etkiledięi ile ilgili birok alıřma vardır (Thissen, Wainer ve Wang, 1994; Wainer ve Thissen, 1993; Wainer, Wang, ve Thissen, 1994). Fakat bu alıřmaların tutarlı olmayan sonular verdięi grlmřtr. Bu bulgular formatın boyutluluk zerindeki etkisinin řartlara gre deęiřebildięini gstermektedir. Bu sebeple formatın boyutluluk zerinde etkisinin olduęu farklı boyut sayısındaki test formları kullanılarak bu durumun lekleme zerindeki etkisinin nasıl deęiřeceęinin belirlenmesinin uygulayıcılara nemli bilgiler saęlayacaęı dřnlmektedir.

Bu alıřmada ele alınacak dięer durumların (test formları arasındaki yetenek ortalaması farkı, ortak madde sayısının toplam teste oranı) lekleme hatası ve yanlılıęı zerindeki etkisi ile ilgili alan-yazında son 30 yılda yapılmıř ok sayıda alıřma olduęu bilinmektedir. Fakat bu durumların ortak madde yapısının zellięi, kestirim yntemi, boyut sayısı ve yetenek daralması ile nasıl bir etkileřim ierisinde olduklarının belirlenmesinin, test uygulayıcılarının doęru veri toplama desenlerini ve yntemlerini semelerini aısından nemli olacaęı dřnlmektedir.

Sonu olarak, MTK'ya dayalı farklı lekleme yntemleri farklı parametre kestirimleri sunmaktadır. Bu sebepten dolayı farklı yntemlerin karřılařtırılması ve farklı durumlar iin en az hata veren yntemlerin belirlenmesi gerekmektedir. Karma-format testler iin de bu tr alıřmaların gerekleřtirilmesi benzer řekilde nemlidir. nk bu tr testlerin kullanımı her geen gn artmaktadır ve doęru veri toplama desenlerinin ve lekleme yntemlerinin seimi daha az hata ieren lmeler yapılmasını saęlayacaktır.

### **1.3. Problem Cümlesi**

Karma format maddelerden oluşan bir, iki ve üç boyutlu testler DOGOM (Denk Olmayan Gruplarda Ortak Madde) deseni kullanılarak ölçeklendiğinde ölçme hatası ve yanlılık; ortak madde setinin yapısı, yetenek daralması ve parametre kestirim yöntemlerine göre nasıl değişmektedir?

#### **1.3.1. Alt Problemler**

- 1) Karma format maddelerden oluşan tek boyutlu testler DOGOM deseni kullanılarak ölçeklendiğinde, kestirilen parametrelerin hatası ve yanlılığı; ortak madde setinin yapısına, yetenek daralmasına ve parametre kestirim yöntemlerine göre nasıl değişmektedir?
- 2) Karma format maddelerden oluşan iki boyutlu testler DOGOM deseni kullanılarak ölçeklendiğinde, kestirilen parametrelerin hatası ve yanlılığı; ortak madde setinin yapısına, yetenek daralmasına ve parametre kestirim yöntemlerine göre nasıl değişmektedir?
- 3) Karma format maddelerden oluşan üç boyutlu testler DOGOM deseni kullanılarak ölçeklendiğinde, kestirilen parametrelerin hatası ve yanlılığı; ortak madde setinin yapısına, yetenek daralmasına ve parametre kestirim yöntemlerine göre nasıl değişmektedir?
- 4) Ölçekleme hatası ve yanlılık üzerinde etkisi incelenen bu değişkenler arasında etkileşimler bulunmakta mıdır?

### **1.4. Sayıtlılar**

Bilindiği üzere türetilmiş verilerden elde edilen bulguların gerçek koşullara genellenebilir olması gerekir. Bunu sağlamak için veri türetme aşamasında gerçek koşullara mümkün olduğunca benzer verilerin türetilmesi büyük önem taşımaktadır. Bu çalışmada türetilen verilerin gerçek test koşullarını yansıttığı varsayılmıştır.

### **1.5. Sınırlılıklar**

1. Gerçekleştirilen bu araştırmada ölçekleme işlemi karakteristik eğri yöntemlerinden biri olan ve Haebara (1980) tarafından geliştirilen yöntem ile sınırlandırılmıştır.
2. Araştırmada parametre kestirimleri 3PLM (Üç Parametrelili Lojistik Model ve ATM (Aşamalı Tepki Modeli) ile sınırlandırılmıştır.

3. Yetenek parametrelerinin kestiriminde MAP (Maximum a posteriori) yöntemi tercih edilmiştir.
4. Araştırmanın verileri tek boyutlu, iki boyutlu ve üç boyutlu veriler ile sınırlıdır.

### 1.6. Tanımlar

*Karma format test:* İçerisinde en az iki farklı formata sahip maddelerin olduğu testtir.

*İki kategorili madde (İKM):* Doğru ya da yanlış olarak puanlanan madde.

*Çok kategorili madde (ÇKM):* Üç ya da daha fazla kategoride puanlanan madde.

### 1.7. Araştırmanın Kuramsal Temeli

#### A. Madde Tepki Kuramı

##### *İki Kategorili Modeller*

MTK'nın temelinde lojistik modeller bulunmaktadır. Bu modeller içerisinde en yaygın bilinenler olarak İKM'lerin kullanıldığı tepki modelleri ve çok kategorili tepki modelleri ön plana çıkmaktadır. İki kategorili tepki modelleri Birnbaum tarafından 1968 yılında geliştirilen 3 parametrelili tepki modeline dayanır. Bu model aşağıda yer alan formül ile ifade edilmektedir (Lord ve Novick, 1968).

$$p_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

Burada  $\theta$  bireyin yetenek düzeyine,  $a$  maddenin ayırt edicilik parametresine  $b$  güçlük parametresine ve  $c$  ise maddenin sözde şans parametresine karşılık gelmektedir. Bu model ilk ortaya çıktığında tek boyutlu yapılar için kullanılsa da yapılan düzeltmelerle birlikte 1980'li yıllardan itibaren çok boyutlu modeller için de kullanılmaya başlanmıştır.

Birnbaum (1968)'un geliştirdiği modelin çok boyutlu durumlara genelleştirilmesi ise şu şekilde olmuştur:

*Mesela,*  $i = 1, \dots, N$  farklı katılımcılar olsun,  $j = 1, \dots, n$  test maddeleri olsun. Ayrıca,  $m$  tane gizil faktör olduğunu varsayalım.  $\theta = (\theta_{i1}, \dots, \theta_{im})$ . Boyutlarla ilintili eğim parametreleri ise  $\alpha_j = (\alpha_{j1}, \dots, \alpha_{jm})$ . Çok boyutlu 3PL modelleri için İKM'ye doğru yanıt verebilme olasılığı:

$$\phi ( x_{ij} = 1 \mid \theta_i, \alpha_j, d_j, \gamma_j ) = \gamma_j + \frac{(1 - \gamma_j)}{1 + \exp[-D(\alpha_j^T, \theta_i + d_j)]}$$

şeklinde ifade edilebilir. Bu modelde şans parametresinin 0 (sıfır)'a sabitlenmesiyle model iki 2PL modeli olarak kullanılabilir.

Burada  $d_j$  kesişme parametresine,  $\gamma_j$  "sözde" şans parametresine,  $D$  ise ölçekleme sabitine karşılık gelmektedir. Bu değer genellikle 1.702 olarak alınmaktadır ve lojistik metriği geleneksel normal ogive metriğine dönüştürmede kullanılmaktadır (Reckase, 2009).

Seçilen MTK modeli veriye uyum gösterdiğinde, Madde Karakteristik Eğrisi (MKE) testi alan gruplara göre değişmez çünkü MKE madde puanının yetenek puanı üzerindeki regresyonu olarak görülür. Bir diğer deyişle, bireyin maddeye doğru yanıt verme olasılığı evrendeki bireylerin yetenek dağılımlarından bağımsızdır. Bu durumda madde parametrelerinin gruplara göre değişmez olduğu varsayılır. Bu özelliğin test geliştirmede, madde havuzu oluşturmada, değişen madde fonksiyonu analizinde ve eşitleme/ölçeklemede önemli kazanımlar sağlar. Diğer taraftan MTK bağlamında yetenek puanı olan  $\theta$ 'nın tanımlandığı ölçek belirli ölçüde yapaydır.

### **Aşamalı Tepki Modeli**

ATM'ler çok kategorili yanıt modelleri için uygundur. Samejima (1969, 1972) tarafından geliştirilen bu modelde ayırt edicilik parametrelerinin sabit tutulduğu durumu düşündüğümüzde  $\tilde{P}_{ijk}$   $\theta_i$  yetenek seviyesine sahip  $i$  kişinin  $j$  maddesi için  $k$  kategorisinin ötesinde puan elde edebilmesine karşılık gelen kümülatif olasılığa karşılık gelir.  $K_j$  kategorileri için  $\tilde{P}_{ijk}$  şu şekilde ifade edilir:

$$\tilde{P}_{ijk} = \tilde{P}_{jk}(\theta_i) = \tilde{P}(\theta_i | a_j, b_{jk}) = \begin{cases} 1 & k = 1, \\ \frac{\exp[Da_j(\theta_i - b_{jk})]}{1 + \exp[Da_j(\theta_i - b_{jk})]} & 2 \leq k \leq K_j, \\ 0 & K > K_j \end{cases}$$

Burada  $a_j$  ayırt edicilik parametresine,  $b_{jk}$  2. Kategoriden  $K_j$  kategorisine kadar güçlük ya da kesişim kategorisine,  $D$  ölçekleme sabitine karşılık gelmektedir. Kategori tepki fonksiyonu olan  $P_{ijk}$  yan yana olan iki yığılmalı (cumulative) olasılık arasındaki farka karşılık gelmektedir ve şu şekilde ifade edilmektedir;

$$P_{ijk} = P_{jk}(\theta_i) = \tilde{P}_{ijk} - \tilde{P}_{ij(k+1)}$$

Samejima (1969)'nin ATM'si daha sonrasında Muraki ve Carlson (1995) tarafından çok boyutlu durumlara genelleyerek çok boyutlu ATM'yi geliştirmişlerdir. Burada  $j$  maddesine ait  $C_j$  kategorileri ve  $d_j = d_1, \dots, d_{(C_j-1)}$  kesim noktaları için yanıt kategorilerinin sınırları şu şekilde ifade edilir:

$$\begin{aligned} \phi ( x_{ij} \geq 0 | \theta_i, \alpha_j, d_j ) &= 1 \\ \phi ( x_{ij} \geq 1 | \theta_i, \alpha_j, d_j ) &= \frac{1}{1 + \exp [ -D ( \alpha_j^T, \theta_i, d_j ) ]} \\ \phi ( x_{ij} \geq 2 | \theta_i, \alpha_j, d_j ) &= \frac{1}{1 + \exp [ -D ( \alpha_j^T, \theta_i, d_j ) ]} \\ &\dots\dots \\ \phi ( x_{ij} \geq 0 | \theta_i, \alpha_j, d_j ) &= 0 \end{aligned}$$

Bu sınırlar  $x_{ij} = k$  tepkisinin koşullu olasılığının şu şekilde olmasına yol açar:

$$\phi ( x_{ij} = k | \theta_i, \alpha_j, d_j ) = \phi ( x_{ij} \geq k | \theta_i, \alpha_j, d_j ) - \phi ( x_{ij} \geq k - 1 | \theta_i, \alpha_j, d_j )$$

### **Parametrelerin Kestirimi**

MTK'nın ilk yıllarında temel odak noktası maddeye özel parametrelerin kestirimi olmuştur. Bock ve Aitkin (1981) EM (expectation – maximization) döngü temelli kestirim yöntemini kullanana kadar bu durum devam etmiş ve MTK küçük test durumları ile sınırlı kalmıştır. EM döngüleri kullanılarak gerçekleştirilen kestirim yöntemi sabit Gauss-Hermite quadrat'ını kullanmaktadır ve düşük boyutlu modeller için kabul edilebilir çözümler sunmaktadır. Diğer taraftan bu yöntem özellikle yüksek boyutlu verilerde etkinliğini kaybetmektedir. Bunun sebebi “*E-basamağı*” için gerekli olan quadrat noktaların boyut sayısı arttıkça çarpımsal olarak artmasıdır. Yani boyut sayısı arttıkça her bir boyut için gerekli olan quadrat nokta sayısı da çarpımsal olarak artmaktadır ve bu durum quadrat noktalarının kestirimler için yetersiz kalmasına yol açmaktadır (Cai, 2008).

Bu duruma geçici çözüm Schilling ve Bock (2005) tarafından getirilmiştir. Buna göre, kestirimlerde uyarlanabilir quadratlar kullanılmaktadır. Bu çözüm her bir boyut için gerekli quadrat sayısını azaltmakta fakat yine de yüksek boyutlu verilerde etkili olamamaktadır.

Daha sonrasında ise olasılıksal (stochastic) kestirim yöntemleri gerek açıklayıcı gerekse doğrulayıcı madde analizlerinde kullanılmaya başlanmıştır. Bayes Markov Zinciri Monte-Carlo (*Bayesian MCMC*) yöntemleri incelenmiş ve Metropolis-Hastings Robbins-Monro (MHRM) yöntemi açıklayıcı (Chai, 2010a) ve doğrulayıcı modeller (Chai, 2010b) için kullanılmaya başlanmıştır. MHRM, Markov chain Monte Carlo (MCMC)'nin prensiplerine dayalı olarak geliştirilmiş bir örnekleme yöntemidir (Houts ve Cai, 2013).

Cai ve Hout (2013) MHRM kestirim yönteminin işleyişini şu şekilde açıklamıştır;

MHRM kestirim yönteminde tekrarlamalar ağırlıklandırılmamış en küçük kareler faktör çıkarımı aşamasından elde edilen ilk değerleri kullanır (Aşama 1). Daha sonraki aşamada, birinci aşamada elde edilen değerler EM benzeri tamamlayıcı bir aşama ile daha da iyileştirilir. Üçüncü aşamada ise MHRM algoritması devreye girerek parametreler kestirilir. MHRM döngüleri yakınsama sağlanıncaya kadar sürer.

MHRM algoritmasının çok boyutlu MTK kestirimlerinde kullanılırken  $j+1$ 'inci döngü üç basamaktan oluşmaktadır.

*İmputasyon:* İlk basamaklarda, daha önceki döngülerden elde edilen geçici madde ve gizil yetenek parametre kestirimlerine ( $\beta^{(j)}$ ) bağlı olarak gizil özelliklerin tesadüfi örneklemeleri ( $\beta^{(j+1)}$ ) seçilir. Bu süreçte değişmez dağılım olarak bireye ait gizil özelliklerin ( $\theta | Y, \beta^{(j)}$ ) posterior dağılımına sahip olan Markov zincirinden MH örnekleyicisi kullanılır.

*Yaklaşma:* ikinci basamakta impute edilen veriye dayalı olarak, madde ve gizil yetenek parametrelerinin daha sonrasında belirlenebilmesi için "bütün veri" log olasılığı ve onun türevleri değerlendirilir. İlk olarak RM filtresi bütün veri bilgi matrisinin koşula dayalı beklentisine tekrarlanan olasılığa dayalı yaklaşma uygulanır. Daha sonra, RM filtresi yeni parametre kestirimlerinin güncellenmesi için kullanılır.

*Robins-Monro Güncellemesi:* üçüncü basamakta madde ve gizil özellik parametrelerinin kestiriminde RM olasılığa dayalı yaklaşma filtreleri kullanılır. MHRM yöntemi MH örnekleyicisinin ürettiği rastsal imputasyonlar tarafından yönlendirilen veri arttırıcı (data augmented) RM algoritmasıdır (Cai, 2010).



MHRM algoritması madde parametrelerini olasılığa dayalı olarak tahmin edilen tüm-veri olabirliği (imputed complete-data likelihood) ve varsayılan evren dağılım formunu kullanarak kestirmektedir. Çok değişkenli normal dağılım genellikle varsayılan normal dağılımdır. Bu kestirim şu şekilde ifade edilebilir.

$$L(\Psi|X, \theta) = \prod_{i=1}^N L_e(x_i | \Psi, \theta) g(\theta | \mu, \Sigma)$$

Açımlayıcı madde analizi için evren ortalama vektörü olan  $\mu$  genellikle  $m \times 1$  sıfırlar matrisine sabitlendiği ve  $\Sigma$  'nın  $m \times m$  birim matrisine sabitlendiği varsayılır. Bununla birlikte, doğrulayıcı madde analizinde  $\mu$  ve  $\Sigma$  birimleri klasik doğrulayıcı faktör analizi ile aynı şekilde kestirilir. Bu formülde görüleceği üzere tüm-veri log-olasılık iki ilave bileşenden oluşmaktadır. Bunlar çok değişkenli ordinal regresyon için log-olasılık ve faktörler arasındaki ilişki için log-olasılık bileşenleridir.

Bu yöntemde kestirim başlangıç  $\Psi$  değerlerinin sağlanmasıyla  $\theta^{(d)}$  değerlerinin hesaplanması ile başlar. Bu aşamada Metropolis-Hastings örnekleyicisi (sampler) kullanılarak gerçekleştirilir (Metropolis ve diğ., 1953; Hastings, 1970). Bu tüm veri eğim vektörü (*complete-data gradient vector*) ve Hessian matrisinin hesaplanmasına ve başlangıç parametrelerinin güncellenmesine olanak verir. Başlangıç parametreleri daha sonrasında Newton-Raphson düzeltmesi ile bir sonraki iterasyon için güncellenir ve bu parametreler  $\theta^{(d+1)}$ 'in örneklenmesi için kullanılır. Bu süreç tekrarlayarak devam eder ve yaklaşık başlangıç parametrelerinin hesaplanması için kullanılır. Bu süreçte en güncel parametreler Robbins-Monro (Robbins ve Monro 1951, Akt. Cai, 2008) kök bulucu (root finding) algoritması ile kontrol edilir. Bu yol ile MH aşamasındaki tutarsızlıklar bulunur. MHRM hem açıklayıcı hem de doğrulayıcı faktör analizi için kullanılabilir.

### ***Yetenek Parametrelerinin Kestirimi***

MTK temelli kestirimlerde bireylerin bir soruya doğru yanıt verebilmeleri, altta yatan yeteneğin ya da özelliğin işlevi olarak tanımlanmıştır. Maddelerin kalibrasyonu tamamlandıktan ve kabul edilebilir uyum sağlandıktan sonra farklı yetenek kestirimleri yapılır. Yetenek değerleri -3 ile 3 arasında olacak şekilde ölçeklenir ve ortalaması 0 (sıfır)'dır. Bu olasılığın bireyin yetenek değerlerine göre nasıl değiştiği MTF ile gösterilmektedir. MTF'nin tersi ise (" $1-P(X)$ " *fonksiyonu*) bir soruya doğru yanıt verememe olasılığına karşılık gelmektedir (Hambleton ve Swaminathan, 1985)

Bireylerin yetenek puanlarının hesaplanmasında bu iki fonksiyonun çarpılması ile elde edilen olabilirlik fonksiyonu (*the likelihood function*) kullanılmaktadır. Bu olabilirlik fonksiyonu şu şekilde tanımlanmaktadır:

$$L(\mathbf{u}|\theta_j) = \prod_{i=1}^n P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}}$$

Burada  $\mathbf{u}$  terimi  $j$  bireyinin 1'den  $n$ 'e kadar sıralı  $i$  maddelerine vermiş olduğu yanıt vektörüne karşılık gelmektedir. Belirtilen bu taslak dâhilinde, yetenek parametrelerinin kestirimi için üç farklı yöntem kullanılmaktadır. Bu yöntemlerden biri en yüksek olabilirlik (*Maksimum Likelihood-ML*) yöntemidir. Bu yöntem adından da anlaşılacağı üzere olabilirlik fonksiyonunda en yüksek noktayı bulur ve bu noktayı yetenek parametresinin değeri olarak kabul eder.

Bir diğer yetenek kestirim yöntemi ise en yüksek sonsal (*maximum a posteriori-MAP*) yöntemi ise en yüksek olabilirlik yönteminin Bayes türevidir. En yüksek noktanın MAP yönteminde de kullanılması sebebiyle bu yöntem Bayes Modal kestirimi olarak adlandırılmaktadır. Bu yöntemde yetenek parametreleri olabilirlik fonksiyonu eğrisinin varsayılan evren dağılımına karşılık gelen bir başka eğri ile çarpılmasıyla elde edilir. Son olarak, yetenek kestiriminde kullanılan yöntemlerden bir diğeri ise beklenen sonsal (*expectation a posteriori, EAP*) yöntemidir. MAP'ta elde edilen ve Bayes yaklaşımla kestirilen eğriyi kullanmakta fakat en yüksek noktayı almak yerine fonksiyon ile ağırlıklandırılan ortalama değeri hesaplamaktadır. Bayes temelli yöntemler testi alanlar içerisinde en az bir kişinin tüm soruları doğru ya da yanlış yaptığı durumlarda yetenek parametrelerinin kestirilmesinde büyük avantaj sağlarlar.

Bayes temelli yaklaşımlar için iki önemli terim önsel ve sonsal dağılımlardır. Önsel dağılım verilerin analizinden önce yeteneğin evrende nasıl dağıldığına yönelik varsayımdır. Bu varsayım evren hakkında önceden sahip olunan bilgiye dayandırılır. Sonsal dağılım ya da ortak olabilirlik dağılımı (*joint likelihood distribution*) ise gerçek verinin dikkate alınmasıyla elde edilir ve önsel dağılım ile olabilirlik fonksiyonun çarpılması ile elde edilir.

Ayrıca, özellikle EAP kestirim yöntemi simetrik olmayan olasılık fonksiyonlarının kullanıldığı durumlarda bu çarpık dağılımı da dikkate alabilmesiyle ön plana çıkmaktadır. Diğer taraftan Bayes yöntemlerinin en büyük sorunu ise yanlı

olmalarıdır (Tong, 2005). Olabilirlik fonksiyonunun başka bir fonksiyon ile çarpılması sistematik hataya yol açmaktadır. Tong (2005) tarafından gerçekleştirilen ve türetilmiş verilerin kullanıldığı bir çalışmada MAP yönteminin yetenek parametrelerinin kestiriminde aynı yetenek düzeyinde yer alan bireylerde yetenek dağılımının daha fazla varyasyon göstermesine sebep olduğu bulunmuştur. Bu durumun farklı yetenek düzeylerindeki grupların ayrışmasını engellediği sonucuna varılmıştır.

## **B. Karma Format Testler**

İKM ve ÇKM'lerin kuramsal ve uygulamaya yönelik sahip olduğu farklılıklar Martinez (1999, s.216) tarafından detaylı olarak ele alınmıştır. Her madde formatının kendine özgü avantajları ve dezavantajları bulunmaktadır. Karma format testlerin farklı format türlerindeki maddelerin dezavantajlarını gidermede olumlu etkisinin olacağı ve bu testleri kullanmanın daha avantajlı olacağı söylenebilir. Bu durum gerçek test uygulayıcılarını karma format testler kullanmaya her geçen gün daha fazla itmektedir. Bu sebeple psikometrik açıdan yüksek nitelikte karma format testlerin oluşturulması daha da önemli hale gelmiştir.

## **C. Eşitleme**

Testlerin güvenliğine yönelik ortaya çıkan kaygılardan ötürü geniş ölçekli testler farklı zaman ve mekânlarda aynı testin farklı formları kullanılarak uygulanmaktadır. Aynı testin farklı formları zorluk açısından bir miktar farklılık gösterebilmektedir. Testi alanlara farklı test formları uygulandığında kolay formu almak bu kişiler için bir avantaj sağlamaktadır. Değerlendirmenin adil olmasını ve farklı formlardan elde edilen puanların birbirlerinin yerine kullanımını sağlamak için puanlar üzerinde dönüştürme yapmak gerekmektedir (Livingston, 2004).

Farklı test puanlarının dönüştürülmesi için ilgili alan-yazında farklı kavramların kullanıldığı görülmektedir. Bu konuda detaylı açıklamalar Kolen ve Brennan (2004) ile von Davier, Holland, ve Thayer (2004) tarafından yapılmıştır. İlişkilendirme (linking) bir testten elde edilen puanların başka bir testten elde edilen puanlara dönüştürülmesidir. İlişkilendirme işleminde kullanılan test formlarının içerik ya da zorluk olarak aynı olmaları gerekmemektedir. Burada amaç zorluk ve içerik eşitliğinden bağımsız olmak üzere test puanları arasında bir ilişki kurmaktır. Eşitleme (equating) ve yordama (predicting) ise ilişkilendirmenin özel türleridir.

Yordama eşitlemenin en eski formudur. Burada amaç bireyin bir testten elde edeceği puanı diğer teste ait bilgi ile yordamaktır. Mesela doğrusal eşitlemede bir testten elde edilen puanlar diğer test formu ile yordama yapılmaktadır. Kullanılan bilgi test puanları olabileceği gibi bireye ait demografik bilgiler ya da birden fazla testten elde edilen puanlar da olabilir. Yordama ile ilgili temel sorun bu yöntemin puanların birbiri yerine kullanılmasını amaçlamaması ve dolayısıyla böyle bir amaç için kullanılmasının mümkün olmamasıdır (Livingston, 2004).

Eşitleme ise testlerden elde edilen puanların birbirleri yerine kullanılabilmesini sağlamak için puanların ayarlanmasını içeren istatistiksel bir süreçtir. Eşitleme farklı formların zorluk farklılıklarının ayarlanmasıdır. Eşitlenen test formları içerik olarak ise benzerdir (Kolen ve Brennan, 2004, s. 2).

Test eşitlemeyi gerçekleştirebilmek için 5 farklı gereksinim vardır (Kolen ve Brennan, 2004). Bu gereksinimler:

- a) *Eşit Yapı Gereksinimi*: eşitleme işleminde kullanılacak test formlarının aynı yapısı ölçmesi gerekmektedir.
- b) *Eşit Güvenirlik Gereksinimi*: eşitleme işleminde kullanılacak test formlarının aynı seviyede güvenilirliğe sahip olması gerekmektedir.
- c) *Simetri Gereksinimi*: eşitleme dönüşümünün X formundan Y formuna gerçekleştirilmesi ile Y formundan X formuna dönüştürülme işleminin tersi olması, yani eşitleme işleminden elde edilecek sonuçları etkilememesi gerekmektedir.
- d) *Eşitlik Gereksinimi*: bir bireyin eşitlemede kullanılan test formlarından hangisini aldığı için önemsiz olması gerekmektedir.
- e) *Evren değişmezliği gereksinimi*: X ve Y test formları eşitlemede kullanılan eşitleme fonksiyonunun seçilen örneklem grubundan bağımsız olması gereklidir.

Bu 5 gereksinimin önemi halen yoğun olarak tartışılmaktadır. Kolen ve Brennan (2004, s.283)'a göre bu gereksinimler soyut, alakasız, pratikte karşılanmaktan uzak ve katıdır. Livingston (2004)'a göre ise özellikle d ve e maddelerinin gerçek test uygulamalarında sağlanması neredeyse imkânsızdır.

Eşitleme, testleri ilişkilendirmenin en güçlü ve en katı varsayımlara sahip olan formudur. Eşitlemede amaç farklı formlardan elde edilen test puanlarını birbiri yerine

kullanılabilir hale getirmektir. Eşitlemede formların içeriği ve güçlüğü benzer olmalıdır. Test güçlüğü ile ilgili farklar ayarlanarak eşitleme yapılır. Puan dönüştürülmesinin yapılmasından sonra aynı içeriğe sahip test formlarına ait puanlar birbiri yerine kullanılabilir.

Test eşitlemede eşitlemenin niteliğini etkileyen bir diğer unsur ise “içerik temsili”dir. Klein ve Jarjoura (1985) içerik temsilini farklı içerik alanlarındaki madde sayısı oranlarının ortak madde seti ve testin tamamı için benzer olması şeklinde tanımlamıştır. İçerik temsilinin eşitleme sonuçları üzerindeki etkisinin özellikle gruplar arası yetenek farklılıkları arttığında daha da önemli bir etkiye sahip olduğu belirtilmiştir. Klein ve Jarjoura (1985) tarafından gerçekleştirilen bir çalışmada içerik temsilinin eşitleme sonuçları üzerindeki etkisi incelenmiştir. Bu çalışmada elde edilen bulgulara göre rastlantısal olmayan gruplarda (non-random groups) içerik temsilinin eşitleme hataları üzerinde önemli etkisinin olduğu ifade edilmiştir.

Eşitlemenin gerçekleştirilebilmesi için farklı veri toplama desenleri geliştirilmiştir. Uygulamada ise üç tanesi yaygın olarak kullanılmaktadır; tek grup deseni, rastgele gruplar deseni ve DOGOM deseni (Kolen ve Brennan, 2004). Tek grup deseni ve rastgele gruplar deseni için, testin farklı formlarını alan grupların aynı ya da eşdeğer olduğu varsayılır. Bu sebeple formlar arası puan farklılıkları doğrudan formların güçlüklerini yansıtır. DOGOM deseninde ise formların güçlük farklılığı ile grupların yetenek farklılığını ayırt etmek teknik açıdan önemli bir zorluktur. Bu zorluğu gidermek için her iki formda da yer almak üzere ortak maddeler farklı formlar arasında bağ kurmak için kullanılır.

Tek grup deseninde yanıt veren tek bir grup vardır ve her iki formu birlikte alırlar. Zaman içinde tecrübe kazanma ya da yorgunluğun etkisini azaltmak için verilen formların sırası denkleştirilmiştir. Yani testi alan grup rastgele iki alt gruba ayrılır. Alt gruplardan birisi X formunu daha önce alırken diğer alt grup önce Y formundaki yanıtları cevaplar. Bunun sonucunda her iki form da aynı bireyler tarafından yanıtlanır.

Rastgele gruplar deseninde bireyler rastgele gruplara ayrılırlar ve X ya da Y formundan birisini alırlar. Gruplara atama rastgele olduğundan her iki grubun denk olduğu varsayılır. Sonuç olarak X formunu alan bireyler ile Y formunu alan bireylerin elde ettikleri puan farklılıklarının form güçlüğüne yansıttığı varsayılır.

Denk olmayan gruplarda ortak madde (DOGOM) deseninde ise gruplar aynı ölçeğin farklı formlarını almaktadırlar fakat bu formların zorluk derecelerinin eşit olduğu varsayılmaz. Bu farklılaşmanın ortaya çıkarılması için her iki forma da ortak maddeler konulur ve formların eşitlenmesi için kullanılır. Ayrıca, DOGOM deseninde iki farklı grup vardır ve bu iki grubun denk olduğuna yönelik bir varsayım söz konusu değildir. Bu sebeple formlar arasındaki puan farklılıkları hem gruplar arasındaki yetenek farklılığının hem de formlar arasındaki maddelerin güçlük farkının sonucudur (Kolen ve Brennan, 2004).

Rastgele grup deseni ile DOGOM deseni arasındaki temel fark, eşitlemenin gerçekleştirildiği grupların aynı evrenden gelip gelmemesidir. Rastgele grup deseninde evren aynı iken DOGOM'da gruplar yetenek düzeyleri açısından farklılaşmaktadır bu sebeple evrenlerden gelmektedir.

DOGOM deseni kullanılarak gerçekleştirilen eşitleme yöntemleri sıkı istatistiksel varsayımlara sahiptir. Kolen ve Brennan (2004, s. 22)'in da belirttiği gibi gruplar arası farklılıklar arttıkça, istatistiksel yöntemler için grup ve form farklılıklarını ayırmak daha zorlaşmaktadır. Bununla birlikte farklı gruplar farklı test formlarını aldığı için önemli grup farklılıkları olabilir. Gruplar arası farklılıklar yüksek olduğunda eşitleme yöntemleri farklı ve bazen hatalı sonuçlar verebilir (Kolen ve Brennan, 2004).

#### **D. Ölçekleme**

Ölçeklemede öğrenci ham puanlarının başka bir puan ölçeğine dönüştürülmesi amaçlanır (Kolen ve Brennan, 2004). MTK bağlamında ise durum farklıdır. Buna göre bireylerin bir test formundan elde ettikleri ham puanlar hali hazırda ortalaması 0 (sıfır) ve standart sapması 1 olan bir ölçeğe aktarılmıştır. Bu ölçekteki puanlar theta ( $\theta$ ) yani bireyin yetenek puanı olarak bilinmektedir. Fakat farklı yaş gruplarında ve farklı yetenek düzeyine sahip öğrencilerin aynı yetenek ölçeğinde olduğundan bahsetmek mümkün değildir. Bu sebeple üst yetenek grubunda yer alan bireylerin yetenek değerlerinin alt yetenek grubunda yer alan bireylerin yetenek ölçeğine dönüştürülmesi gerekmektedir. Bu durum MTK'ya dayalı elde edilen parametrelerinin dönüştürülmesi gerektirir.

Ölçekleme içerik alanlarına ait farklı seviyelerdeki test formlarından elde edilen puanların tek bir metriğe dönüştürülmesidir. Ölçeklemenin uygulamadaki temel

faydası öğrencinin gelişimi hakkında bilgi sahibi olmaktır. Elde edilen puanlar sayesinde öğrenci gelişimsel süreçte bir yere yerleştirilir. Farklı zaman dilimlerinde gerçekleştirilen ölçümler aynı zamanda farklı seviyelere denk gelmektedir. Burada amaç öğrencilerin her bir seviyeye özel içerik için yeteneğini kestirmek ve seviyeler arasındaki (grade-to-grade) gelişimini belirleyebilmektir. Eğitim öğretim ortamları bağlamında seviyeden kasıt farklı sınıflarda gerçekleştirilen ölçümlerdir. Bu konuda bir diğer yaklaşım ise içeriğin zaman içerisindeki gelişiminin belirlenebilmesidir. Haliyle, ölçeklemenin seviyeler arasındaki gelişimi belirlemeye çalışıldığında normatif, belirli bir içerikte sağlanan gelişimin belirlenmesini amaçladığında ise mutlak ölçme olduğu kabul edilebilir (Kolen and Brennan, 2004).

Patz ve Hanson (2002)'a göre ardışık iki sınıf arasında evrene ait yetenek seviyelerindeki fark aynı grup içerisindeki bireyler arasındaki farka kıyasla yeterli miktarda yüksek olmadığı ya da farklı seviyelerdeki bireyler için beklenen çıktılar ve standartlar çok fazla örtüştüğünde bu seviyeleri ortak bir ölçekte ilişkilendirmek anlamlı bilgiler sunar. Yani müfredatın farklı sınıflarda geçerli ortak amaçlara sahip olması gerekir. Uygulamada ise fen, matematik ve okuma gibi alanlarda bu şartların fazlasıyla karşılandığı görülmektedir.

Dikey ölçeklerin oluşturulması test formlarının ortak bir ölçekte ilişkilendirilmesine dayanır. Teknik olarak, paralel olmayan fakat ortak bir yeteneği ölçen ölçekler ilişkilendirilebilir. Farklı sınıf seviyelerindeki içerik standartlarının örtüşmesi ortak bir yapıyı ölçtüklerine işaret eder. Fakat farklı sınıflara ait içerikler ve psikometrik özelliklerdeki farklılıklar bu formların paralel olmadığını gösterir. Bu durumda bu formlar ilişkilendirilebilir fakat eşitlenemez. Ölçekleme işleminde ise formlar arasında bilinçli olarak oluşturulan güçlük farklılıkları dikey ilişki kurmayı gerektirir ve bu işlem ölçekleme olarak ifade edilir (Kolen ve Brennan, 2004). Diğer taraftan ilgili alan-yazın incelendiğinde ölçekleme ile dikey eşitlemenin genellikle aynı şeyi kastettiği ve birbirleri yerine kullanıldıkları görülmüştür.

Ölçekleme sonucunda elde edilen sonuçlar bireyin gelişimi ile ilgili çıkarımlar yapmakta kullanılmaktadır. Ölçekleme eşitlemeye göre daha karmaşıktır. Bunun durumun sebebi farklı seviyelerde uygulanan testlerden elde edilen puanlar arasındaki kararlılık ve benzerliktir. Bu ilişki testlerin güçlüğü, güvenilirliği ve içeriği gibi kaynaklardan etkilenmektedir. Eşitlemede grupların eşdeğer olduğu varsayıldığı

için test formları arasındaki güçlük farklılıklarının yönetilmesi daha kolaydır (Kolen ve Brennan, 2004).

Ölçekleme uygulamada birçok avantaj sağlamaktadır. Her şeyden önce, ölçekleme bireylerin gelişimini görmeyi ve süreç içerisinde sağlıklı bir şekilde takip edebilmesini sağlar. Bu yönüyle tekrarlı ölçümlere benzemektedir. Ayrıca, ölçekleme farklı sınıfları, farklı okulları ve farklı çağlarda öğrenim görmüş öğrencilerin başarı olarak karşılaştırılabilmesine olanak sağlar. Benzer şekilde, ölçekleme sayesinde maddelere dayalı önemli karşılaştırmalar gerçekleştirilebilir. Hangi maddelerin hangi sınıf seviyesi için daha uygun olduğu tespit edilebilir. Son olarak, ölçekleme çalışmaları standart belirleme çalışmalarının öğrencilerin gelişimsel özellikleri ile daha fazla uyumlu bir şekilde gerçekleştirilebilmesine olanak sağlar.

Diğer taraftan, ölçekleme sonucunda ulaşılan çıkarımlar bir takım kısıtlılıklara da sahiptir. Ölçeklemenin gerçekleştirildiği test formları paralel olmadığı için eşitleme ile ilgili yapılan çıkarımlar ölçekleme için geçerli olmayabilir. Eşitlemeye kıyasla daha zayıf karşılaştırılabilirlik imkânı sunan ölçekleme sonucunda varılan çıkarımların da geçerliliği düşer. Ayrıca, ölçekleme ardışık sınıflar arasındaki gelişimi göstermek için faydalı olmasına rağmen, grupların seviyeleri arasındaki fark arttıkça yorum yapabilme olanağı azalır. Örneğin 3. sınıf ile 8. sınıf öğrencilerine ait test formları doğrudan ölçeklenemez (Haertel, 1991).

Ölçekleme için DOGOM deseni en yaygın kullanılan veri toplama desenidir. Bu yöntemde farklı yetenek seviyeleri ortak madde bloklarına yanıt verirler. Bu ortak madde blokları farklı yetenek seviyelerini ilişkilendirmede köprü görevi görürler (Kolen ve Brennan, 2004).

Eşitleme ve ölçekleme sürecinde gerçekleştirilen işlemler istatistikseldir ve bundan dolayı hem rastlantısal hem de sistematik hatalar ortaya çıkabilir. Rastlantısal hatalar örnekleme ile paralel gelişir. Eğer evrendeki tüm bireyler analize dâhil edilirse örnekleme hatası oluşmaz. Bununla birlikte, örnekleme dayalı hata olduğu için örneklem büyüklüğü arttırıldığında rastlantısal hatalar da buna bağlı olarak düşer. Diğer taraftan sistematik hata ya da modelin doğru tanımlanmamasına (model-mis-specification) bağlı ortaya çıkan hata varsa bu durum örneklem büyüklüğünün arttırılması ile düşmez. Bu durum yanlılık olarak ifade edilmekte ve istatistiksel varsayımların ihlal edildiğini göstermektedir. Ayrıca, model veri



uyumsuzluğu da sistematik hatanın bir diğer sebebi olabilir (Kolen ve Brennan, 2004).

MTK temelli ölçeklemede uygulayıcılar farklı birçok seçeneği dikkate almak zorunladılar. Madde parametrelerinin kestirimi için seçilecek modele karar verme, veri toplama deseninin belirlenmesi, eşitleme yönteminin seçimi ve yetenek parametrelerinin ortak ölçeğe yerleştirilmesi bir takım tercihler yapmayı gerektirmektedir (Tong ve Kolen 2007).

Cook ve Petersen (1987)'e göre grup yetenek seviyeleri birbirine benzer olduğu durumlarda farklı ortak madde setleri benzer eşitleme sonuçlarına sebep olmaktadır. Diğer taraftan gruplara ait yetenek seviyeleri farklılaştığında ortak madde setinin eşitleme sonuçlarını farklılaştırabildiği belirtilmiştir. Bu noktada ortak madde setinin seçiminde daha dikkatli olunması gerektiği ifade edilmiştir.

Harris (1991) ölçekleme işlemini gerçekleştirecek kişilerin dikkat etmeleri gereken konuları şu şekilde belirtmiştir:

- a) Ölçekleme belirli bir gereksinim sonucunda mı gerçekleştirilecek?
- b) Ölçekleme için gerekli olan varsayımlar testin özellikleri tarafından desteklenmekte midir?
- c) "Gelişme" nasıl tanımlanmaktadır?
- d) Ölçekleme işleminin gerçekleştirilebilmesi için örneklem büyüklüğü yeterli mi?

Eşitleme ve ölçeklemede kullanılacak yöntemler üzerine yapılacak çalışmalar önemlidir çünkü yöntemlerin etkisinin koşullara göre değiştiği görülmüştür. Ayrıca bu koşulların arasında etkileşim olması da muhtemeldir. Bu çalışmada ise ilgili alan-yazında ölçekleme hatası üzerinde etkisi tutarlı olarak doğrulanmış olan örneklem büyüklüğü, ortak madde setinin büyüklüğü, toplam madde sayısı gibi faktörler yerine daha az çalışılmış koşullardaki ölçeklemenin etkililiği incelenmiştir.

Bu koşullar yetenek dağılımlarına ait varyansın üst grupta daralması ya da her iki grup için ortak olması, kestirim yönteminde EM ya da MHRM döngülerinin kullanılması ve ortak madde setinin sadece iki kategorilerden oluşması ya da karma format maddelerden oluşması olarak belirlenmiştir.

## E. Ölçekleme Yöntemleri

Test formlarının ölçeklenebilmesi için KTK ve MTK'ya dayalı birçok yöntem bulunmaktadır. KTK'ya dayalı yöntemler geleneksel eşitleme yöntemleri olarak da bilinmekte ve ortalama eşitleme, doğrusal eşitleme ve eşit yüzde eşitleme olmak üzere üç alt başlıkta toplanmaktadır. Bu araştırmada MTK'ya dayalı yöntemlere ölçekleme yapıldığından bu yöntemler açıklanmış, KTK'ya dayalı yöntemlere yer verilmemiştir.

MTK temelli eşitleme yöntemlerinde ham puanlar ya da yüzdelik dilimler yerine kalibre edilmiş parametreler kullanılmaktadır. Burada test formlarından birisinden elde edilen parametreler eşitleme sabitleri kullanılarak diğer test formundan elde edilen parametrelerin yer aldığı ölçeğe aktarılır. Bu test formları "Form X" ve "Form Y" olarak ifade edilebilir. Ayrıca bu formların yer aldıkları ölçekler de yine X ve Y ölçekleri olarak ifade edilebilir. Burada X ölçeği dönüştürme işleminin uygulandığı eski ölçeğe, Y ölçeği ise dönüştürme işleminin sonucunda elde edilen yeni ölçeğe karşılık gelmektedir, Bu test formlarından elde edilen parametreler de geleneksel olarak  $F$  ve  $T$  sembolleri ile ifade edilmektedir. Buna göre  $F$  formuna ait yetenek parametreleri olan  $\theta_F$  değerlerini  $T$  formuna ait yetenek parametreleri olan  $\theta_T$  dönüştürmek için

$$\theta_T = A\theta_F + B = \theta_F^*$$

eşitliği kullanılmaktadır. Burada,  $A$  ve  $B$  eşitleme sabitleri ortalama ve standart sapma değerlerini ayarlamak için kullanılmaktadır.  $\theta_F^*$  ifadesi ise dönüştürülen değerlere karşılık gelmektedir. Madde parametreleri mutlak bir şekilde  $\theta$  ölçeğine bağlı olduğu için ölçeğin doğrusal dönüştürülmesi nihai olarak madde parametrelerini de, beklenen olasılıkları değiştirmeden, dönüştürmektedir.

Eşitleme işleminde kullanılan moment yöntemleri olan ortalama/sigma (Marco 1977) ve ortalama/ortalama (Loyd ve Hoover 1980) en basit yöntemlerdir ve  $A$  ve  $B$  sabitlerinin kestirilebilmesi için yalnızca parametrelerin bilinmesini gerektirir. Moment yöntemlerine alternatif olarak kullanılan yöntemler ise Haebara (1980) ve Stocking ve Lord (1983) tarafından gerçekleştirilen karakteristik eğri yöntemleridir. Bu yöntemlerin temelinde ortak maddelerin madde karakteristik eğrileri arasındaki kareler farkı toplamının en aza indirgenmesi mantığı yatmaktadır. Burada temel varsayım eşitlenen parametrelerin hatasız olarak kestirildiğidir. Haebara (1980)

yöntemi bu varsayımın sağlanması için bir ölçüt ortaya koymaktadır. Buna göre parametrelerin diğer ölçeğe dönüştürülmesi çift taraflı olarak dikkate alınmalıdır. Bunu sağlamak için “Form X” parametrelerinin “Form Y” parametrelerine dönüştürülmesi ile “Form Y” parametrelerinin “Form X” parametrelerine dönüştürülmesi eş zamanlı olarak dikkate alınır. Eski formdaki parametrelerin yeni forma dönüştürülmesi simetrik olmayan yaklaşım olarak adlandırılır. Diğer taraftan yeni formdaki parametrelerin eski forma dönüştürülmesi ise simetrik yaklaşım olarak kabul edilir. Haebara (1980) simetrik ve simetrik olmayan yaklaşımların her ikisi için ortak maddelerin madde karakteristik eğrileri arasındaki kareler farkı toplamının en aza indirgenmesine yönelik bir kriter belirlemiştir. Bu kriter  $Q = Q_1 + Q_2$  olarak ifade edilmiştir.  $Q_1$  simetrik olmayan kestirimlere,  $Q_2$  ise simetrik kestirime karşılık gelmektedir.

Haebara yöntemi  $Q = Q_1 + Q_2$  kriterini en aza indirger. Burada:

$$Q_1 = \frac{1}{L_1} \sum_{m=1}^M \sum_{s=1}^S \sum_{k=1}^{K_a} [P_{sk}(\theta_{mT}) - P_{sk}^*(\theta_{mT})]^2 W_1(\theta_{mT})$$

ve

$$Q_2 = \frac{1}{L_2} \sum_{m=1}^M \sum_{s=1}^S \sum_{k=1}^{K_a} [P_{sk}(\theta_{mF}) - P_{sk}^*(\theta_{mF})]^2 W_2(\theta_{mF})$$

Ayrıca:

$$L_1 = \sum_{m=1}^M W_1(\theta_{mT}) K_S \sum_{s=1}^S$$

ve

$$L_2 = \sum_{m=1}^M W_2(\theta_{mT}) K_S \sum_{s=1}^S$$

olarak ifade edilmektedir.

Bu eşitlikte

$\theta_{mT}$  : Y ölçeğinde yer alan M noktalar setidir ve beklenen olasılıklardaki farkların değerlendirilmesini sağlar.

$P_{sk}(\theta_{mT})$  : Y ölçeğinde yer alan ve dönüştürülmemiş ortak maddelerin parametrelerine dayalı beklenen olasılıklardır.

$P_{sk}^*(\theta_{mT})$  : X ölçeğinde yer alan ve dönüştürülmüş ortak maddelerin parametrelerine dayalı beklenen olasılıklardır.

$W_1(\theta_{mT})$  :  $\theta_{mT}$ 'ye karşılık gelen kuadrat ağırlıklar setidir.

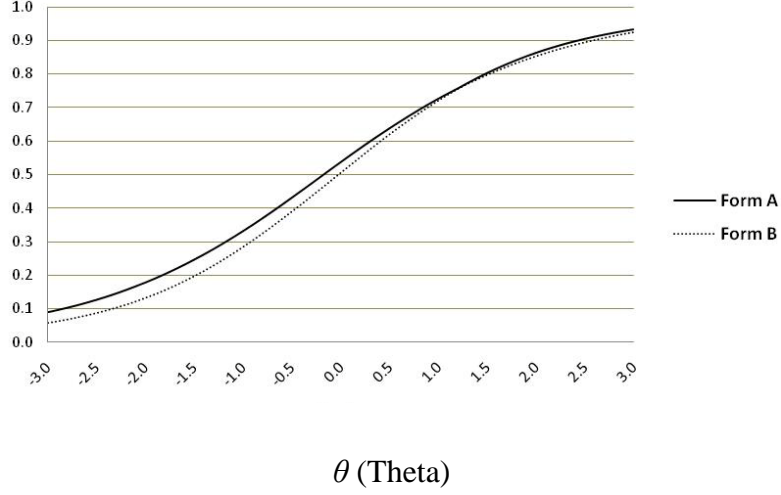
$(\theta_{mF})$  : X ölçeğinde yer alan M noktalar setidir ve beklenen olasılıklardaki farkların değerlendirilmesini sağlar.

$P_{sk}(\theta_{mF})$  : X ölçeğinde yer alan ve dönüştürülmemiş ortak maddelerin parametrelerine dayalı beklenen olasılıklardır.

$P_{sk}^\#(\theta_{mF})$  : Y ölçeğinde yer alan ve dönüştürülmüş ortak maddelerin parametrelerine dayalı beklenen olasılıklardır.

$W_2(\theta_{mF})$  :  $\theta_{mF}$ 'ye karşılık gelen kuadrat ağırlıklar setidir.

Burada  $L_1$  ve  $L_2$  sabitleri değeri optimizasyonun alt ve üst sınırlarından taşmasını önlemek için standartlaştırmada kullanılan ölçüt fonksiyonudur. Bu sabitler eşitleme sabitlerinin kestirilmesini etkilememektedir (Kim ve Lee, 2006). Görüldüğü gibi burada  $Q$  simetrik yaklaşımla en aza indirgenmektedir fakat yalnızca  $Q_1$  simetrik olmayan yaklaşımla en aza indirgenmektedir.



### Şekil 1.1: İki Farklı Test Formuna Ait Test Karakteristik Eğri Örneği

Daha basit bir ifade ile karakteristik eğri yöntemlerinde amaç farklı gruplardan elde edilen karakteristik eğrileri arasındaki farkın en aza indirgenmesidir. Şekilde de görüleceği üzere farklı yetenek düzeylerinde iki test formuna ait test karakteristik formları arasında boşluklar yer almaktadır. Karakteristik eğri yöntemleri test karakteristik formları arasındaki boşlukları en aza indirgeyecek A ve B katsayılarını bulmayı amaçlamaktadır. Haebara ve Stocking-Lord yöntemleri arasındaki temel fark ise test karakteristik formları arasındaki boşlukların nasıl tanımlandığı ile alakalıdır. Stocking-Lord yöntemi karakteristik eğrileri arasındaki boşlukları eğriler arasındaki farkların toplamının karesinin alınmasına dayanırken Haebara yöntemi ise farkların kareler toplamının hesaplanmasına dayanmaktadır.

### F. Karma Format Testlerde Eşitleme

Problem durumunda belirtildiği gibi karma format testler farklı puanlanan maddelerden oluşan testlerdir. Bazı örneklerde üç ya da dört farklı madde formatının aynı testte yer aldığı görülse de çoğunlukla uygulamada iki formatın yer aldığı testler tercih edilmektedir. Bu çalışma kapsamında da iki madde formatının yer aldığı testlerle gerçekleştirilen eşitleme senaryosu tercih edildiğinden ortak madde setinin oluşturulması için üç farklı yol bulunmaktadır: 1) Sadece İKM'lerle oluşturulan ortak madde setleri, 2) İKM ve ÇKM'ler ile oluşturulan ortak madde setleri, 3) Sadece ÇKM'ler ile oluşturulan ortak madde setleri. Uygulamada sadece İKM'lerin yer aldığı ortak madde setlerinin kullanıldığı çok sayıda test örneği ağırlıklı olmakla birlikte karma format ortak madde setlerinin kullanıldığı testler de bulunmaktadır. Karma format ortak madde setleri toplam testi hem içerik ve özellikleri itibariyle daha iyi

temsil edeceğinden ötürü kullanılmasının daha ideal olduğu belirtilmiştir (Kim ve Lee, 2004).

Karma format testlerin eşitlemesi DOGOM deseninde gerçekleştirildiğinde tatmin edici ortak madde setinin oluşturulması gruplar arasında göreceli olarak büyük farklılıklar varsa oldukça zor olabilir. Bu zorluklar eşitlemenin gerçekleştirilebilmesi için gerekli olan varsayımların sağlanmasını engelleyebilir.

Karma format testlerin DOGOM deseni kullanılarak eşitlenmesi sürecinde yapılandırılmış maddeler genellikle kullanılmaz. Bu durumun bazı sebepleri vardır. İlk olarak bu maddeler bir testte sayı olarak daha azdır. Bu sebeple ortak madde olarak kullanılması pratikte zordur. İkinci olarak bu maddelerin ezberlenmesi ve akılda tutulması daha kolaydır. Bu da testin güvenliğini ve eşitlemeden elde edilecek sonuçların kesinliğini tehlikeye düşürebilir. Son olarak, yapılandırılmış yanıtı dayalı maddeler insanlar tarafından puanlanmaktadır ve puanlayan kişinin yanlılığı eşitleme sonuçlarını olumsuz etkileyebilmektedir (Hagge, 2010). Diğer taraftan, eşitleme sürecinde yalnızca çoktan seçmeli maddelerin kullanılması da yine bir takım zorluklara sebep olmaktadır.

Bu zorluklardan en önemlisi ortak maddelerin içerik ve istatistiksel açıdan toplam testi temsil edebilmesinin zorunlu olmasıdır. Nitekim Kolen ve Brennan (2004) ortak madde setinin testin “mini” bir versiyonu olması gerektiği ifade etmiştir. Karma format testlerde, ortak madde setinin, toplam testi temsil edebilmesi için madde formatının etkisinin de dikkate alınması büyük öneme sahiptir (Kirkpatrick, 2005).

Ortak madde setine yönelik bir diğer önemli nokta ise ortak madde setinin uzunluğudur (Ricker ve von Davier, 2007). Ortak madde setinde daha fazla maddenin yer alması istenilen bir durumdur fakat fazla sayıda maddenin ortak madde setinde yer alması test güvenliği ve testin maliyeti açısından sorunlar doğurabilmektedir. Kolen ve Brennan (2004, s. 271) ortak madde setinin 20 maddeden daha fazla ya da toplam testin %20'si kadar olması gerektiğini belirtmiştir. Fakat Kolen ve Brennan tarafından önerilen bu kural ortak madde setinin çoktan seçmeli maddelerden oluştuğu durumlar düşünülerek yapılmıştır. Karma format testler için ise alan-yazında açık bir öneriye rastlanmamıştır.

Çok boyutluluk karma format testlerde ciddi bir sorun olarak ortaya çıkmaktadır. Çok boyutluluk bir testte birden çok yapının ölçülmesine karşılık gelmektedir (Hagge,

2010). Çok boyutlu karma format testlerde ise farklı formattaki maddelerin kullanılması ayrıca çok boyutluluğa sebep olan bir diğer etmenddir (Hanson ve Wang, 2002). Sadece çoktan seçmeli maddelerin ortak madde setinde yer alması ortak madde setinin ölçtüğü yapılar ile toplam testin ölçtüğü yapıyı temsil edememesi ile sonuçlanabilir. Ayrıca, yalnızca çoktan seçmeli maddelerin yer aldığı ortak madde setinden elde edilen puanlar karma format maddelerin ölçtüğü yapıya dair var olan gruplar arasındaki gerçek farkı tam olarak yansıtmayabilir. Bu durum eşitleme sonuçlarında sistematik hatalara sebep olur.

### **G. Eşitleme Çalışmalarında Kullanılan Veri Türleri**

Eşitleme ile ilgili çalışmalar eşitleme ilişkilerinin incelenmesi için farklı türlerde veriler kullanmışlardır. Bu veri türleri içerisinde en çok kullanılanlar gerçek verilerdir. Gerçek verileri kullanmanın en büyük yararı gerçek test uygulamalarında yer alan test madde ve bireylerle çalışmaktır. Bu çalışmada ise türetilmiş veri kullanılmıştır. Türetilmiş verilerle gerçekleştirilen çalışmalarda çalışmanın amaçları doğrultusunda maddelerin, testin ve testi alan bireylerin özellikleri kontrol edilebilir ve istendiğinde bu özelliklere müdahale edilebilir. Ayrıca, türetilmiş veriler evren madde parametreleri ve bilinen yetenek parametreleri kullanılarak elde edildiği için gerçek eşitleme ilişkileri bilinebilir. Bu durum türetilmiş veri kullanmanın temel avantajlarından birisidir. Ayrıca, türetilmiş verilerin kullanılması karıştırıcı etmenlerin sonuçları etkilemesini önleyebilir. Diğer taraftan türetilmiş veri kullanılmasının bir takım sakıncaları da olabilir. Bu sakıncaların başında ise gerçek test durumlarını yansıtmadaki olası başarısızlıktır.

Gerçek veriler ile karşılaştırıldığında türetilmiş verilerin bir takım kısıtlılıkları ve avantajları bulunmaktadır. Türetilmiş veriler gerçek test uygulamalarında hiçbir zaman olmayacak mükemmel durumları temsil etmektedir. Elde edilen bulguların geçerliği türetilen veri ile gerçek veriler arasındaki benzerliğin derecesine ve veri türetme işlemlerinin niteliğine bağlıdır. Diğer taraftan, gerçek bir veriye matematiksel bir model uygulandığında karıştırıcı faktörler devreye girmekte ve model uyumunu etkileyebilmektedir. Türetilmiş veri kullanmadaki amaç bu karıştırıcı faktörlerin etkisini ortadan kaldırmaktır.

## 2. İLGİLİ ARAŞTIRMALAR

Li, Lissitz ve Yang (1999) tarafından gerçekleştirilen çalışmada karma-format testlerde genişletilmiş Stocking-Lord Yönteminin performansı incelenmiştir. Ayrıca, farklı formatta maddelerin oranının eşitleme sonuçları üzerindeki etkisi üzerinde durulan bir başka noktadır. Çalışma kapsamında türetilmiş veri setleri kullanılmıştır ve örneklem büyüklüğü, eşitleme durumları, grup yetenek farklılıkları gibi durumlara müdahale edilerek bu etmenlerin etkisine bakılmıştır. Madde parametrelerini kestirirken, çoktan seçmeli maddeler için 3 PLM, açık uçlu seçmeli maddeler için ise genelleştirilmiş kısmi derece modeli kullanılmıştır. Elde edilen bulgulara göre Stocking-Lord yönteminin karma-format testlerde kullanılabilir etkili bir yöntem olduğu söylenilebilir.

Bastari (2000) gerçekleştirilmiş olduğu çalışmada test uzunluğu, farklı formatlardaki maddelerin oranı, ortak madde setinin büyüklüğü, örneklem büyüklüğü, grup yetenek dağılımları ve farklı eşitleme yöntemlerinin, eşitleme hatası üzerindeki etkisini incelemiştir. Araştırmada elde edilen bulgular göstermiştir ki testteki soru sayısı, çoktan seçmeli soruların test içindeki oranı ve ortak madde sayıları arttıkça, gruplar arasında yetenek seviyesi farkı azaldıkça eşzamanlı kalibrasyon yöntemleri, ayrı kalibrasyon yöntemlerine göre daha etkili eşitleme sonuçları vermektedir.

Tate (2000) tarafından gerçekleştirilen çalışmada 11 farklı durum için veri türetilmiştir. Bu farklı durumlarda örneklem büyüklüğü, grup yetenek farkları, ÇKM'lerin İKM'lere oranı, ortak madde setinin uzunluğu, ortak madde setinin yapısı ve çok boyutluluk gibi faktörlerin etkisi incelenmiştir. Parametrelerin kestiriminde 2PLM ile ATM'den yararlanılmıştır. Kullanılan eşitleme işleminin dayanıklılığını test etmek için kestirilen parametreler ile veri türetmede kullanılan gerçek parametreler karşılaştırılarak hata ve yanlılık değerleri hesaplanmıştır. Elde edilen bulgulara göre uzun ortak veri setinin bulunduğu durumlarda hata ve yanlılık miktarlarının düştüğü görülmüştür. Elde edilen bir diğer bulguya göre tek boyutlu veri seti için ortak madde setinin sadece iki kategorilerden oluştuğu senaryoda hata miktarları düşmekte fakat çok boyutlu test için hata miktarları oldukça fazla artmaktadır. Ayrıca, tek boyutlu veri setlerinde İKM ve ÇKM'lerin bir arada olmasının tek boyutluluk varsayımının ihlaline sebep olmadığı görülmüştür.



Skyes ve arkadaşları (2002) tarafından gerçekleştirilen bir diğer çalışmada ABD’de yer alan bir eyalette geniş ölçekli uygulanan bir matematik testine ait gerçek verilerden yararlanılmıştır. Kullanılan bu test iki boyuttan oluştuğu ve ilk boyutta yer alan maddelerin karma format olduğu, ikinci boyutta yer alan maddelerin ise iki kategorili olduğu belirtilmiştir. Bu testler kullanılarak dört farklı test formu oluşturulmuştur. Her bir formda aynı sorular yer almaktadır fakat ortak madde setleri farklı olasılıklara göre belirlenmiştir. İlk iki durumda ortak maddelerin her iki boyutu eşit şekilde temsil edecek şekilde seçilmiştir. Diğer iki durumda ise ortak maddelerin iki boyutun sadece birisinden yük alacak şekilde belirlenmiştir. Eşitleme işleminin gerçekleştirilmesi için Stocking-Lord yöntemi kullanılmıştır. Eşitleme işlemi ağırlıklandırılmış ortalama kare farkları kullanılarak gerçekleştirilmiştir. Elde edilen bulgular göstermiştir ki dengeli (balanced) yani ortak maddelerin her iki boyuttan seçildiği durumların daha az hata verdiği görülmüştür.

Kamata ve Tate (2005) gerçekleştirmiş oldukları çalışmada karma format testlerde puanlayıcı etkisini incelemişlerdir. Araştırmacıları böyle bir çalışmaya iten gerekçe ise karma-format testlerde puanlayıcıların farklılaşmasından ötürü, ortak açık uçlu soruların aslında ortak olmayabileceği düşüncesidir. Nitekim Pack ve Kim (2007) tarafından gerçekleştirilen çalışmada elde edilen bulgular göstermiştir ki sadece çoktan seçmeli soruların ortak madde olarak kullanıldığı karma-format testlerde hata oranları daha düşük çıktığı bulunmuştur.

Kirkpatrick (2005) tarafından gerçekleştirilen çalışmada ise türetilmiş ve gerçek veriler birlikte kullanılarak açık uçlu soruları ortak madde setine yerleştirmenin etkisini incelemiştir. Daha net olarak belirtmek gerekirse, bu çalışmada iki farklı durum oluşturulmuştur. Durumlardan birisinde iki açık uçlu soru kullanılmış, diğer durumda ise soruların tamamı çoktan seçmeli soru olarak seçilmiştir. Puanlayıcı etkisi kontrol edildiğinde gruplar arasındaki yetenek puanları farkının daha fazla etkili olduğu görülmüştür. Bu etkinin karma format maddelerin olduğu veri setlerinde olduğu görülmüştür.

Kim ve Lee (2006) sıklıkla kullanılan ayrı kalibrasyon (ortalama/ortalama, ortalama/sigma, Haebara ve Stocking-Lord) yöntemlerini genişleterek karma-format testlerin ölçeklenmesi için kullanımını farklı koşullar altında test etmiştir. Bu çalışmada iki farklı grubun tek bir ölçeği aldığı (ve bütün maddelerin ortak madde olduğu) bir durum tasarlamışlardır. Bu durum için veriler türetmişlerdir. Çoktan

seçmeli ve açık uçlu soruların eşzamanlı olarak ortak madde setinde yer aldığı, sadece çoktan seçmeli soruların ortak madde olduğu ve sadece açık uçlu soruların ortak madde olduğu durumlarda eşitlemeler gerçekleştirilmişlerdir. Ayrıca, örneklem büyüklüğü ve gruplar arasındaki yetenek farklılıkları da test edilen durumlar arasındadır. Araştırma neticesinde elde edilen bulgular göstermiştir ki karakteristik eğri yöntemleri genellikle daha az hatalar vermişlerdir. Ayrıca her iki formattaki soruların yer aldığı eşitleme durumlarının tek format ortak madde durumlarına göre daha az hatalar verdiği görülmüştür.

Kim ve Kolen (2006), Kim ve Lee (2006) tarafından geliştirilen karma-format eşitleme yöntemlerinin format etkisi sebebi ile MTK varsayımlarının ve model uyumunun sağlanmadığı durumlarda etkililiklerini test etmiştir. Veriler veri formatı sebebiyle çok boyutludur ve ortak maddeler eşit oranda karma-format madde içerecek şekilde oluşturulmuştur. Çalışma sonucunda elde edilen bulgulara göre karakteristik eğri yöntemlerinin moment yöntemlerine göre daha az ölçme hatası verdikleri ve bu durumun format etkisinden bağımsız olduğu görülmüştür. Ayrıca eşzamanlı kalibrasyonun diğer dört yöntemle göre daha az hata verdiği belirtilmiştir.

Kim, Walker ve McHale (2008) karma format testlerde trend puanlama eşitlemenin etkililiğini incelemiştir. Trend puanlama farklı sınıf seviyelerindeki puanlayıcıların yapılandırılmış maddeleri puanlamada gösterdikleri cömertliği önlemek için kullanılan bir yöntemdir. Bir soruyu farklı sınıf seviyelerinde ortak madde olarak kullanmak sorunun zorluğu ya da öğrencilerin yetenek düzeyinin yanında puanlayıcıların cömertliğinden de etkilenir. Üst sınıflarda puanlayıcıların beklentileri daha yüksek olacağından daha katı ölçütler kullanılabilir. Bu durum eşitleme sonuçlarını etkileyebilir. Trend puanlama farklı sınıf seviyelerine yerleştirilen ortak maddelerde puanlayıcı etkisinin kontrol edilmesini amaçlar. Çalışmada kullanılan testte 8 adet İKM, 4 adet ÇKM ve karma formatta ortak maddeler yer almıştır. Araştırmada DOGOM deseni kullanılmıştır. Araştırmada ortak madde yapısı kontrol edilerek iç farklı durum oluşturulmuştur. İlk durumda İKM'ler ile trend olmayan ÇKM'ler yer almıştır, ikinci durumda İKM'ler ile trend ÇKM'ler yer almıştır ve son olarak üçüncü durumda sadece İKM'lere yer verilmiştir. Ayrıca, ikinci eşitleme deseni olarak ortak madde olmayan hibrit desen kullanılmıştır. Hibrit desen tek grup desen ile denk gruplar deseninin birleşimi ile oluşturulmuştur. Eşitleme yöntemi olarak hibrit desende sadece zincirli doğrusal yöntem kullanılmıştır. Hibrit desenin

daha az hata ve yanlılığa sebep olduğu bulgulanmıştır. Ortak madde yapısının kullanıldığı üç farklı durum için ise iki kategorili ve trend çok kategorili durumun en az yanlılık ve hataya sebep olduğu görülmüştür.

Cao (2008) tarafından gerçekleştirilen türetilmiş verilerle gerçekleştirdiği çalışmasında ise ortak madde setinin temsil etme ve temsil etmeme durumlarının etkisi test edilmiştir. Bu çalışmada temsil etme ve etmeme istatistiksel, format ve içerik bağlamında tanımlanmıştır. Aynı zamanda tek boyutluluk varsayımının format etkisinden kaynaklanan çok boyutluluk durumundaki ihlalinin etkisi de incelenen bir diğer durumdur. Elde edilen bulgulara göstermiştir ki eşit gruplarda hata miktarı daha azdır, ortak maddenin temsil ediciliği hata miktarını düşürdüğü belirtilmiştir.

Walker ve Kim (2009) sadece İKM'lerin yer aldığı karma format testlerde eşitlemenin niteliği incelemiştir. Araştırmada 24 İKM ve 12 ÇKM den oluşan bir ölçme aracı kullanılmıştır. Bu test maddeleri kullanılarak iki paralel form oluşturulmuştur. Paralel formların her birisinde 16 İKM, 8 ÇKM yer almış, bu maddelerden 8 tane iki kategorili ve 4 tane çok kategorili olan madde ortak madde setine dâhil edilmiştir. Oluşturulan formlar tek bir testten türetildiği için testi alan bireyler tüm maddeleri yanıtlamışlardır ve bu sebeple tek grup deseni kullanılmıştır. Bu sebeple elde edilen sonuç sadece İKM'lerin ortak madde setinde yer alan desenin etkinliğinde kıstas olarak kullanılmıştır. Eşitleme işleminin gerçekleştirilmesi için zincirli doğrusal, zincirli eşit yüzdeli eşitleme yöntemlerinden yararlanılmıştır. Eşitleme işleminin etkinliği RMSE ve ölçmenin standart hatası indisleri kullanılarak test edilmiştir. Ayrıca eşitlemede iki farklı strateji kullanılmıştır. İlk strateji için sadece İKM'ler kullanılmıştır. İkinci stratejide ise İKM'ler ilk aşamada İKM'ler kullanılarak eşitlenmiş ikinci aşamada ise bütünleşik puanlar iki kategorili puanlara eşitlenmiştir. Elde edilen bulgulara göre her iki stratejinin de benzer sonuçlar verdikleri belirlenmiştir. Bu bulgular sadece İKM'lerin ortak veri setinde yer aldığı durumlarda, etkili eşitleme işlemlerinin gerçekleştirilebileceği belirtilmiştir.

Wu ve arkadaşları (2009) İKM'lerin dış ortak madde olarak yer aldığı durumlarda eşitleme işleminin etkililiğini nicelemiştir. Çalışmada türetilmiş veriler kullanılarak 8 tane ÇKM ve dış ortak madde olarak kullanılmak üzere 8 tane İKM kullanılmıştır. Ayrıca eşitleme sonuçları üzerindeki form güçlükleri arasındaki fark, İKM ve ÇKM'ler arasındaki ilişkinin derecesi ve örneklem büyüklüğü değişkenlerinin etkilerine bakılmıştır. Eşitleme işlemleri Levine ve Tucker yöntemleri kullanılarak

gerçekleştirilmiştir. Elde edilen sonuçlar ağırlıklandırılmış MSE, ağırlıklandırılmış ölçmenin standart hatası ve ağırlıklandırılmış yanlılık indisleri kullanılarak değerlendirilmiştir. Elde edilen bulgular göstermiştir ki ağırlıklandırılmış yanlılık indisleri en fazla formata bağlı ilişki ve grup farklılıklarından etkilenmiştir. Ayrıca ağırlıklandırılmış eşitlemenin standart hatası formata dayalı ilişki ve gruplar arası yetenek farkından etkilendiği bulunmuştur.

Yao ve Boughton (2009) tarafından gerçekleştirilen bir diğer çalışmada çok boyutlu karma format testlerde eşitleme işleminin niteliğini test karakteristik fonksiyonu/yanıt fonksiyonunun eşleştirilmesi yöntemi ile belirlemeyi amaçlamıştır. Bu çalışmada gerçek verilere ait parametrelerden türetilen veri setlerinden yararlanılmıştır. Eşitleme işlemi üzerindeki etkisi ortak madde setinin uzunluğu, evren yetenek dağılımı, maddelerin basit ya da karmaşık yapıda olmaları durumları kontrol edilerek incelenmiştir. Elde edilen bulgulara göre ortak madde seti uzunluğunun 6 maddeden oluştuğu durumlarda bile düşük hatalı eşitleme işlemlerinin gerçekleştirilebileceği belirtilmiştir.

Tan, Kim, Paek ve Xiang (2009) İKM'lerin ortak madde olarak kullanıldığı DOGOM desenini trend puanlama yöntemi ile karşılaştırmışlardır. Ayrıca, İKM'lerin ÇKM'lere oranı ve formata dayalı maddeler arası gözlenen ilişki durumlarını da ayrıca incelemiştir. Veri olarak sözde test formları kullanılmıştır. Öncelikli olarak trend puanlama eşitleme işlemi öncelikli olarak gerçekleştirilerek diğer yöntemle karşılaştırılmıştır. Değerlendirme sürecinde ortalama ağırlıklandırılmış farklar, RMSD ve durumsal RMSD değerleri hesaplanmış ve karşılaştırma işlemleri gerçekleştirilmiştir. Elde edilen bulgular göstermiştir ki çoktan seçmeli maddelerin oranı arttığında ve formata bağlı ilişki yüksek olduğunda her iki yöntem de nitelik olarak karşılaştırılabilir sonuçlar verdiği görülmüştür.

Hagge (2010) tarafından gerçekleştirilen bir diğer çalışmada karma format testlerin özellikleri ile ortak madde setinin eşitleme sonuçları üzerindeki etkisi incelenmiştir. Çalışmada veriler DOGOM uygun olarak toplanmıştır. Çalışmada gerçek verileri yerleştirme testine ait uygulamalardan edinilmiştir. Araştırmada MTK temelli yöntemlerden gerçek puan eşitlemesi ve gözlenen puan eşitlemeleri ile zincirli eşit yüzdeli eşitleme ile frekans kestirim yöntemi kullanılmıştır. Araştırmadan elde edilen bulgulara göre gruplar arası fark arttıkça eşitleme yanlılığının arttığı ve ortak madde

setinin yapısına dair tutarlı olmayan sonuçlara ulaşıldığı belirtilmiştir. Ayrıca, MTK temelli yöntemlerin daha az yanlılık ve hataya sebep olduğu belirtilmiştir.

Hagge, Liu, He, Powers, Wang ve Kolen (2010) tarafından gerçekleştirilen çalışmada DOGOM kullanarak gerçekleştirilen çalışmada farklı koşullara göre değişen grup farklılıklarının eşitleme sonuçları üzerindeki etkisi incelenmiştir. Çalışmada MTK gözlenen puan eşitlemesi, MTK gerçek puan eşitlemesi, frekans kestirimi yöntemi ve zincirli eşit yüzdeli eşitleme kullanılmıştır. Araştırmada kullanılan veriler ileri yerleştirme sınavına ait uygulamalardan elde edilmiştir. Grup farklılıklarına dair oluşturulan beş farklı durum test edilmiştir. Elde edilen bulgulara göre frekans kestirim yönteminin grup farklılıklarına yüksek derecede hassas olduğu görülmüştür. Ayrıca sistematik eşitleme hatasının özellikle MTK temelli yöntemler için daha düşük olduğu görülmüştür.

Lee, He, Hagge, Wang ve Kolen (2012) tarafından gerçekleştirilen çalışmada “*ileri yerleştirme*” sınavındaki verilere dayalı türetilmiş veriler kullanılmıştır. Bu çalışmada ortak madde setinde yalnızca İKM’lerin kullanılmasının etkinliğinin incelenmesi amaçlanmıştır. Ek olarak, farklı formattaki maddeler arasındaki ilişkinin ve gruplar arası yetenek farkındaki değişimin eşitleme sonuçlarını nasıl etkilediği üzerinde durulmuştur. Gerçek eşitleme ilişkileri basit yapılı çok boyutlu gözlenen puan eşitleme yöntemi kullanılarak hesaplanmıştır. Değerlendirme kıstasları olarak yanlılık, varyans ve MSE değerleri kullanılmıştır. Elde edilen bulgulara göre farklı formattaki maddeler arasındaki ilişki azaldıkça ya da gruplar arasındaki yetenek farkı arttıkça hata miktarının arttığı görülmüştür. Gruplar arasındaki yetenek farkının düşük olduğu senaryoda farklı eşitleme yöntemlerinin benzer sonuçlar sağladığı görülmüştür. Ayrıca, grup yetenek farkının fazla olduğu senaryoda kabul edilebilir ölçme sonuçları elde edebilmek için farklı formattaki maddeler arasında yüksek seviyede ilişki olmasının gerekli olduğu belirtilmiştir.

Suksuwan ve arkadaşları (2012) tarafından gerçekleştirilen çalışmada karma format çok boyutlu verilerde ortogonal olmayan Proctustes yöntemi kullanılarak gerçekleştirilen eşitlemelerde ortak maddelerin oranının etkisi incelenmesi amaçlanmıştır. Araştırma kapsamında ortak madde oranları %20, %25 ve %30 olarak belirlenmiştir. Bu kapsamda, Monte Carlo yöntemi kullanılarak veriler türetilmiş ve her bir verinin büyüklüğü N=3000 olarak belirlenmiştir. Eşitleme sonuçlarının etkinliğinin belirlenebilmesi için RMSE ve yanlılık değerleri

hesaplanmıştır. Elde edilen bulgulara göre ortak madde oranının farklılaştığı durumlarda elde edilen RMSE ve yanlılık değerlerinin .05 düzeyinde anlamlı seviyede farklılaştığı görülmüştür.

Ülkemizde gerçekleştirilen tez çalışmaları incelendiğinde eşitleme konusunda yapılan ilk çalışmanın ise Kelecioğlu (1994) tarafından gerçekleştirildiği ve YÖK'e ait ulusal tez merkezinde test eşitleme ile ilgili Ocak-2016 tarihi itibarıyla 15 farklı çalışmanın yer aldığı görülmüştür.

Ülkemizde yapılan ve karma format testlerin eşitlenmesine yönelik gerçekleştirilen bir çalışmada (Gübeş, 2014) ortak madde seti yapısının, gruplara ait  $\theta$  dağılımının ve eşitlemede kullanılan yöntemlerinin karma testlerle gerçekleştirilen eşitleme işleminde eşitlik özelliğinin korunuma etkisi araştırılmıştır. Araştırmada belirlenen amacın incelenmesi için türetilmiş veriler ve eşitleme deseni olarak DOGOM kullanılmıştır. Eşitleme işlemleri için MTK'ya dayalı gerçek puan eşitleme ve gözlenen puan eşitleme yöntemlerinden yararlanılmıştır. Eşitlik özelliği, birinci-sıra eşitlik ve ikinci-sıra eşitlik özelliğine göre D1 ve D2 indeksleri ile değerlendirilmiştir. Elde edilen bulgulara göre birinci sıra eşitlik özelliği ölçütüne göre tüm faktörlerin, ikinci sıra eşitlik özelliği ölçütüne göre ise ortak madde seti formatı dışında kalan diğer faktörlerin eşitleme sonuçları üzerinde anlamlı etkilerinin olduğu belirlenmiştir.

Uysal (2014) tarafından gerçekleştirilen bir diğer çalışmada karma format testlerde moment yöntemleri ile karakteristik eğri yöntemleri kullanılarak gerçekleştirilen eşitleme çalışmada ortak madde oranı ve yetenek dağılımının etkisi incelenmiştir. Gerçekleştirilen bu araştırmada türetilmiş veriler kullanılmıştır. Verilerin türetilmesi sırasında DOGOM'a uygun olarak N=1000 büyüklüğünde veri setleri türetilmiştir. İKM'lerin kalibrasyonu için 3 parametrelili lojistik model ve ÇKM'ler için genelleştirilmiş kısmi kredi modeli kullanılmıştır. Farklı durumların karşılaştırılması için eşitleme hataları (RMSE) kullanılmıştır. Elde edilen bulgulara göre en fazla hatayı moment yöntemleri verdiği görülmüştür. Ayrıca, en az hata miktarının sağa çarpık grupların eşitlendiği ve %30 ortak maddenin yer aldığı testlerde Stocking-Lord yöntemi ile elde edildiği belirtilmiştir. Ayrıca, eşitlenen test formları arasındaki yetenek dağılımlarının benzer olmasının daha az hatalı sonuçlara ulaşmayı sağladığı belirlenmiştir.

## 2.1. İlgili Araştırmalar Özeti

İncelenen çalışmalar ışığında etkisi incelenen değişkenler arasında gruplar arası yetenek farklı (Li, Lissitz ve Yang, 1999; Tate, 2000; Kirkpatrick, 2005; Lee, He, Hagge, Wang ve Kolen, 2012), ortak madde setinin toplam teste oranı (Bastari,2000; Tate, 2000; Yao ve Boughton, 2009; Sukswan ve diğ., 2012; Uysal, 2014) ve örneklem büyüklüğünün (Bastari, 2000; Tate, 2000; Wu ve diğ., 2009) ön plana çıktığı ve daha yoğun olarak incelendiği görülmüştür. Bu çalışmalarda tutarlı olarak gruplar arası yetenek farkının azaldığı, örneklem büyüklüğünün arttığı ve ortak madde setinin toplam teste oranının daha yüksek olduğu durumlarda ölçekleme sonuçlarının daha iyi sonuçlar verdiği bulgulanmıştır. Bu sebeple, gerçekleştirilen bu çalışmada bu değişkenler dahil edilmemiştir.

Gerçekleştirilen çalışmaların bazılarında ortak madde seti yapısının ölçekleme sonuçları üzerindeki etkisi incelenmiştir. (Gübeş, 2014; Tate, 2000; Skyes ve diğ. 2002 ve Kim ve Lee, 2006). Bu çalışmalarda elde edilen bulgular farklılık gösterdiği görülmüştür. Diğer taraftan incelenen çalışmalarda kestirim yöntemlerinin ve yetenek daralmasının etkisinin incelendiği çalışmaya rastlanmamıştır.

Ölçekleme işlemi için kullanılan yöntemlerin çeşitlilik gösterdiği görülmektedir. Tercih edilen ölçekleme yöntemlerinin Stocking-Lord yöntemi (Li, Lissitz ve Yang,1999; Skyes ve diğ.,2002), trend puan eşitleme yöntemi (Kim, Walker ve McHale,2008; Tan, Kim, Paek ve Xiang,2009), Levine ve Tucker eşitleme yöntemi (Wu ve diğ.,2009) ve ortogonal olmayan Proctustes yöntemi (Sukswan ve diğ.,2012) olduğu görülmüştür. Diğer bazı çalışmalarda ise ölçekleme yöntemleri karşılaştırılmıştır (Bastari, 2000; Kim ve Kolen, 2006 ve Kim ve Lee,2006).

Ayrıca araştırmada kullanılan veri türlerinin değişkenlik gösterdiği görülmektedir. Çalışmaların bazılarında gerçek veriler kullanılırken (Skyes ve diğ., 2002; Kim, Walker ve McHale, 2008 ve Walker ve Kim,2009) bazı çalışmalarda ise türetilmiş veriler tercih edilmiştir (Kim ve Lee, 2006; Cao, 2008 ve Wu ve diğ., 2009). Ayrıca, bazı çalışmalarda da gerçek verilerden elde edilen parametreler kullanılarak türetilmiş verilerin kullanıldığı görülmüştür (Yao ve Boughton, 2009; Tan ve diğ., 2009 ve Lee ve diğ.,2012). Gerçek veriler kullanılarak gerçekleştirilen çalışmaların bazılarında kullanılan testler araştırmacı tarafından geliştirilmiştir. Örneğin, Kim, Walker ve McHale (2008) çalışmasında kullandığı testte 8 İKM ve 4 ÇKM yer alırken

Walker ve Kim (2009) ise 24 İKM ve 12 ÇKM kullanmıştır. Kullanılan bu soru sayılarının karma format maddelerin yer aldığı “İleri Yerleştirme Sınavı” gibi geniş ölçekli sınavların soru sayılarına kıyasla daha az olduğu görülmüştür.





## 3. YÖNTEM

### 3.1. Araştırmanın Türü

Bu araştırmada, farklı koşullar altında gerçekleştirilen ölçeklemeler sonucunda parametrelere ait hata ve yanlılık değerlerinin nasıl etkilendiği incelenmiştir. Bu sayede hangi koşulların ölçeklemenin niteliğini daha fazla etkilediği belirlenerek alan-yazına katkı sağlanması amaçlanmıştır. Araştırma kapsamında kullanılan ön bilgilerin tamamı kontrol altında tutularak veriler türetildiği için kuramsal katkısı ön planda olan bir araştırmadır. Bu sebeple bu araştırma bir temel araştırmadır (Karasar, 2000).

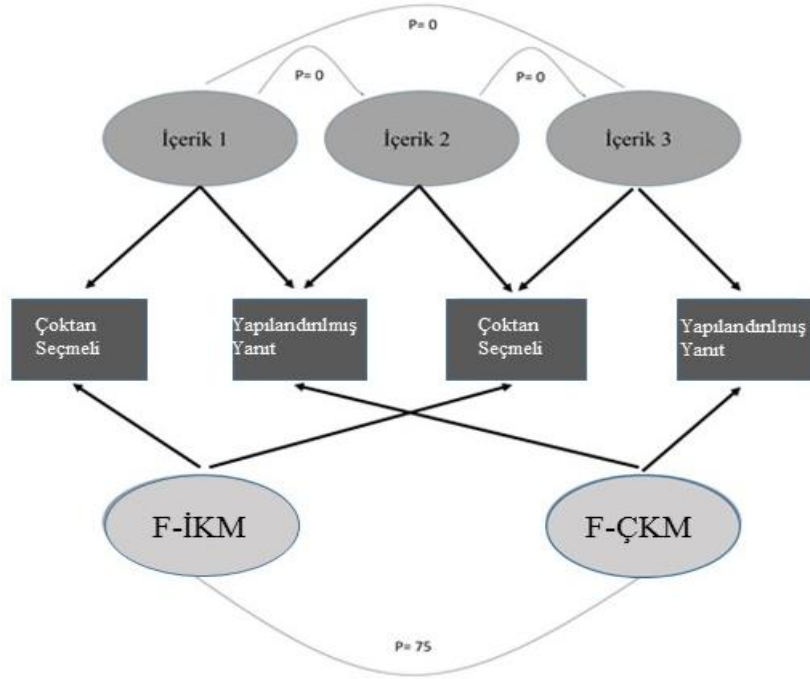
### 3.2. Eşitleme Deseni

Bu araştırmada denk olmayan gruplarda ortak madde (DOGOM) eşitleme deseni kullanılmıştır.

### 3.3. Test Yapılandırması

Simulasyon çalışmalarında veri setlerinin türetilmesi gerçek yetenek puanları ve gerçek madde parametreleri olmak üzere iki temel bilgiyi gerektirir. Fakat bu çalışmada boyutluluk yapısı da kontrol edilen bir koşul olduğundan dolayı boyutluluk yapısını gösteren kovaryans matrisinin de belirtilmesi gerekmektedir. Veri türetme işlemi gerçekleştirilirken boyutluluk yapısı iki kademeli (two-tier) model dikkate alınarak hazırlanmıştır.

İki kademeli modeller temel boyutlar ve özel boyutlar olmak üzere iki kademeye ayrılmaktadır. Boyutluluk kaynağının temel ya da özel olması o kaynağın kuramsal olarak daha önemli olması ya da boyutlar arasındaki kovaryans yapısından kaynaklanmaz. İki kademeli modelde temel ve özel boyutların arasında kuramsal olarak ilişki bulunmamaktadır. Ayrıca, özel faktörler karşılıklı olarak dik açıdır (ortogonal) ve herhangi bir madde en fazla bir boyuttan yük alır, diğer boyutlardan ise yük almaz. Diğer taraftan temel faktörler ise kendi aralarında ilişkili olabilirler. Bu belirtilenler haricinde ise model üzerinde başka herhangi bir kısıtlama bulunmamaktadır. Bu modelin genel yapısını gösteren ve bu araştırmaya uyarlanmış bir örnek Şekil 3.1'de yer almaktadır.



**Şekil 3.1. İki Kademeli Model Örneği**

Şekil 3.1'de altta yer alan F-İKM ve F-ÇKM boyutları temel boyutlara karşılık gelmektedir ve madde formatından kaynaklı boyutluluğu temsil etmektedir. Üst kısımda yer alan içerik boyutları ise içeriğe dayalı boyutluluğa karşılık gelen özel boyutları temsil etmektedir. Genel ve özel boyutlar arasındaki ilişkiler  $p$  ile gösterilmiştir. İki kademeli modellerin özelliği gereği özel boyutlar arasında ilişki olmayacağından dolayı  $p = 0$  olarak alınmıştır. Genel boyutlar arasındaki ilişki ise  $p=.75$  olarak gösterilmiş ve bu çalışmada da veri türetme işlemi bu değerler dikkate alınarak gerçekleştirilmiştir.

Bu şartlara göre tek boyutlu, iki boyutlu ve üç boyutlu veri yapıları için kovaryans matrisleri oluşturulmuştur. Oluşturulan bu matrisler Çizelge 3.1'de yer almaktadır.

**Çizelge 3.1. Verilerin Türetilmesinde Kullanılan Kovaryans Matrisleri**

<i>Tek Boyutlu Veri Yapısı</i>			<i>İki Boyutlu Veri Yapısı</i>				<i>Üç Boyutlu Veri Yapısı</i>							
	F1	F2	C1		F1	F2	C1	C2		F1	F2	C1	C2	C3
F1	1	0.75	0	F1	1	0.75	0	0	F1	1	0.75	0	0	0
F2	0.75	1	0	F2	0.75	1	0	0	F2	0.75	1	0	0	0
C1	0	0	1	C1	0	0	1	0	C1	0	0	1	0	0
				C2	0	0	0	1	C2	0	0	0	1	0
									C3	0	0	0	0	1

Daha önce de belirtildiği üzere bu çalışmada İKM ve ÇKM'lerin yer aldığı karma format veri setleri türetilmiştir. Buradan hareketle madde formatına göre boyutluluk iki kademeli modelde temel faktörler olarak alınmış ve Çizelge 3.1'de F1 ve F2 ile gösterilmiştir. Yani Çizelgede yer alan F1 ve F2 boyutları genel faktörlerdir. Bu çalışma için genel faktörler olan madde formatları arasındaki ilişki 0.75 olarak belirlenmiştir.

Bir diğer boyutluluk kaynağı ise maddelerin içeriğinden kaynaklanmaktadır. İçeriğe dayalı boyutluluk ise özel faktörler olarak alınmıştır ve C1, C2 ve C3 ile gösterilmiştir. Çizelgede de görüleceği üzere iki kademeli modele uygun olarak genel ve özel maddeler arasındaki ilişkinin olmadığı varsayılarak kovaryans değerleri 0 (sıfır) olarak alınmıştır. Ayrıca iki kademeli modele göre bir diğer kısıtlama ise özel faktörlerin arasında ilişki bulunmamasıdır. Yani içeriğe dayalı boyutluluk basit yapıdadır (simple structure).

Cook ve Petersen (1987) grup yetenek seviyeleri birbirine benzer olduğu durumlarda ortak madde setlerinin yapısı tek formatlı ya da karma olmasının herhangi bir etkisi olmadığı ve benzer ölçekleme sonuçları oluştuğu belirtmişlerdir. Bu noktada ortak madde setinin seçiminde daha dikkatli olunması gerektiği ifade edilmiştir. Bu sebeple alt grup ve üst grup arasındaki yetenek seviye farkı belirlenirken  $\theta$  ortalamalarının farkı 1 olarak belirlenmiştir. Tercih edilen bu değer yüksek kabul edilebilecek bir değerdir ve ortak madde setinin etkilerinin daha iyi görülebileceği varsayılmıştır.

Önemli bir diğer nokta ise veri setlerinde yer alacak toplam madde sayısı ve ortak madde sayılarının belirlenmesidir. Kolen ve Brennan (2004) ortak madde setinin

uzunluğunun 20 madde ya da toplam madde sayısının %20'si kadar olması gerektiğini belirtmişlerdir. Ayrıca, çok uzun testler için ortak madde sayısının 30'u geçmesine gerek olmadığını vurgulanmıştır. Diğer taraftan DOGOM deseni kullanılarak gerçekleştirilen araştırmalarda Amerika Birleşik Devletlerinde uygulanan ileri yerleştirme (advance placement) testlerine ait verilerinin tercih edildiği görülmektedir. Türetilmiş verilerle gerçekleştirilen çalışmalarda da bu testler referans alınmaktadır. Bu testlerde yer alan farklı alt alanlardaki madde sayılarının 120'ye kadar çıktığı görülmüştür. Bu bilgiler ışığında bu araştırmada madde sayısı 108, ortak madde sayısı ise 24 olarak belirlenmiştir. Ortak madde setinde yer alacak maddelere ait bilgiler Çizelge 3.2'de yer almaktadır.

**Çizelge 3.2. Ortak Maddelerin Boyutlara ve Formata Göre Temsil Durumu**

Boyutluluk	Sadece ÇKM	Karma Format
Tek Boyutlu	24	6 (ÇKM)/18(İKM)
İki Boyutlu	12/12	3(ÇKM)/9(İKM) - 3(ÇKM)/9(İKM)
Üç Boyutlu	8/8/8	2(ÇKM)/6(İKM) - 2(ÇKM)/6(İKM) - 2(ÇKM)/6(İKM)

Yao ve Boughton (2009) ÇBMTK modellerinde parametre kestirimlerinin kesin ve kararlı olarak gerçekleştirilebilmesi için örneklem büyüklüğünün en az 3000 olması gerektiğini belirtmişlerdir. Bu sebeple gerçekleştirilen bu çalışmada örneklem büyüklüğü her koşul için 3000 olarak belirlenmiştir.

Farklı yetenek seviyesine sahip test formlarının dikey ölçeklenmesinin gerçekleştirildiği gerçek uygulamalarda üst yetenek grubu için bireylerin  $\theta$  dağılımlarının varyansının azaldığı görülmüştür ve bu durum yetenek daralması (Scale Shrinkage) ya da bu çalışmada kullanılacağı şekliyle yetenek daralması olarak adlandırılmıştır (Yen, 1985). Yetenek daralmasının gözlemlendiği ölçeklere Stanford Başarı Testi ve Iowa temel beceriler testi örnek olarak gösterilebilir. Yen'in belirttiğine göre bu ölçeklerdeki yetenek daralması %38 ile %85 arasında değişmektedir. Bu çalışmada ise üst grup olarak belirlenen veri setleri için yetenek daralması gözlenen koşullarda varyans %40 daraltılmıştır.

### 3.4. Verilerin Türetilmesi

Veri türetme aşamasında ayırt edicilik parametreleri için evren değerlerine karşılık gelecek  $a$  matrisleri, güçlük parametreleri için  $b$  matrisleri ve yetenek parametresi

için  $\theta$  matrisleri türetilmesi gerekmektedir. Bu kısımda bu veri setleri türetilirken kullanılan evren parametrelerin hangi koşullarda türetildiğine değinilmiştir.

Her bir eşitleme analizi için alt ve üst gruplar için birer tane yanıt matrisi türetilmiştir. Yetenek parametreleri türetilirken boyut sayısı kadar  $\theta$  vektörü ( $N=3000$ ) normal dağılımdan rastgele olarak türetilmiştir. Alt grup için türetilen parametrelerin yetenek ortalaması 0, üst grup için ise 1 olarak alınmıştır. Türetilen “gerçek”  $\theta$  değerleri “ $N \times$  boyut sayısı” boyuta sahip olacak şekilde türetilmiştir.

Kovaryans matrisleri ile paralel olarak tek boyutlu veri setlerinin türetilmesi için  $3 \times N$ , iki boyutlu veri setlerinin türetilmesi için  $4 \times N$  ve üç boyutlu veri setlerinin türetilmesi için  $5 \times N$  boyutlu gerçek  $\alpha$  parametreleri matrisleri türetilmiştir. Her birisi 108 maddeden oluşan veri setlerinde ilk 45 madde İKM, devamındaki 18 madde ÇKM ve son 45 madde İKM olacak şekilde evren parametreleri türetilmiştir. Parametrelere ait betimleyici istatistikler Çizelge 3.2’de yer almaktadır

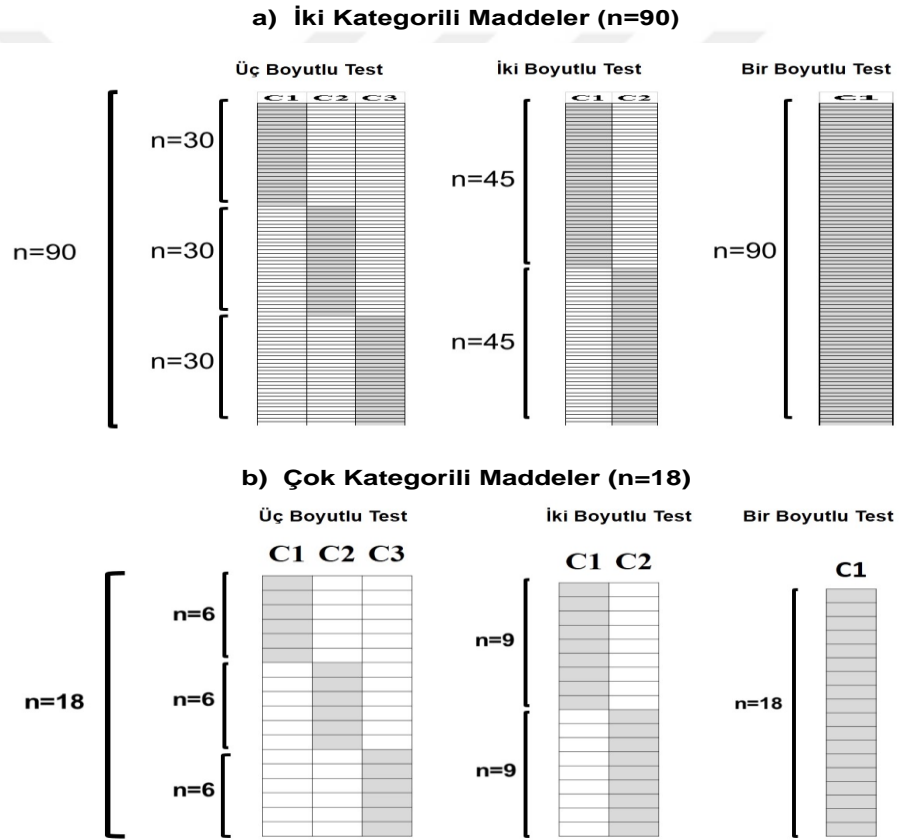
**Çizelge 3.3. Veri Setlerinin Türetilmesi İçin Kullanılan Evren  $\alpha$  Parametrelerine Ait Betimsel İstatistikler**

	<i>İKM (n=45)</i>	<i>ÇKM (n=18)</i>	<i>İKM (n=45)</i>
<b><i>Tek Boyutlu Veri Setleri</i></b>			
<b>F1</b>	(n=45, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )	(n=18, $\bar{x} = .2$ , $x_{std\ sap.}=.03$ )	(n=45, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )
<b>F2</b>	(n=45, $\bar{x} = .2$ , $x_{std\ sap.}=.03$ )	(n=18, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )	(n=45, $\bar{x} = .2$ , $x_{std\ sap.}=.03$ )
<b>C1</b>	(n=45, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )	(n=18, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )	(n=45, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )
<b><i>İki Boyutlu Veri Setleri</i></b>			
<b>F1</b>	(n=45, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )	(n=18, $\bar{x} = .2$ , $x_{std\ sap.}=.03$ )	(n=45, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )
<b>F2</b>	(n=45, $\bar{x} = .2$ , $x_{std\ sap.}=.03$ )	(n=18, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )	(n=45, $\bar{x} = .2$ , $x_{std\ sap.}=.03$ )
<b>C1</b>	(n=45, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )	(n=9, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ ) & (n=9, $\bar{x} = 0$ )	(n=45, $\bar{x} = 0$ )
<b>C2</b>	(n=9, $x = 0$ )	(n=9, $\bar{x} = 0$ ) & (n=9, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )	(n=45, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )
<b><i>Üç Boyutlu Veri Setleri</i></b>			
<b>F1</b>	(n=45, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )	(n=18, $\bar{x} = .2$ , $x_{std\ sap.}=.03$ )	(n=45, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )
<b>F2</b>	(n=45, $\bar{x} = .2$ , $x_{std\ sap.}=.03$ )	(n=18, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )	(n=45, $\bar{x} = .2$ , $x_{std\ sap.}=.03$ )
<b>C1</b>	(n=30, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ ) & (n=15, $x = 0$ )	(n=6, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ ) & (n=12, $x = 0$ )	(n=45, $x = 0$ )
<b>C2</b>	(n=30, $x = 0$ ) & (n=15, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )	(n=6, $x = 0$ ) & (n=6, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ ) & (n=6, $x = 0$ )	(n=15, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ ) & (n=30, $x = 0$ )
<b>C3</b>	(n=45, $x = 0$ )	(n=6, $x = 0$ )& (n=6, $x = 0$ )& (n=6, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )	(n=15, $x = 0$ )& (n=30, $\bar{x} = 1$ , $x_{std\ sap.}=.15$ )

Çizelgede madde formatına dayalı boyutluluğunu temsil eden F1 ve F2 temel boyutları için  $a$  değerleri veri setlerinin boyutluluk yapılarına göre F1 boyutu olan İKM’ler için  $a$  ortalamaları 1, standart sapma değerleri .15 olarak belirlenmiştir, F2

boyutu olan ÇKM'ler için  $a$  ortalamaları 1, standart sapma değerleri .15 olarak belirlenmiştir. F1 ve F2 boyutunda yer alan maddeler diğer boyutlarda ise  $a$  ortalamaları .2, standart sapma değerleri .03 olarak belirlenmiştir.

İçeriğe dayalı boyutluluğunu temsil eden C1, C2 ve C3 özel boyutları için  $a$  değerleri veri setlerinin boyutluluk yapılarına göre madde bulunduğu boyutta  $a$  ortalamaları 1, standart sapma değerleri .15'tir. Eğer madde o boyutta yük almıyorsa  $a$  değeri sıfır (0) olarak belirlenmiştir. Son olarak, türetilen  $a$  parametreleri normal dağılım gösterecek şekilde belirlenmiştir. Oluşturulan boyutluluk yapısı ve maddelerin yer aldığı boyutlar Şekil 3.1'de yer almaktadır.



**Şekil 3.2. Veri Setlerinde Yer Alan Maddelerin Ait Oldukları Boyutlar**

Şekil 3.2. de görüldüğü üzere türetilen testler 90 tane İKM ve 18 tane ÇKM içerecek şekilde türetilmiştir. Şekilde yer alan C1, C2 ve C3 içeriğe dayalı oluşturulan boyutları temsil etmektedir. Türetilen testin sahip olduğu boyut sayısına göre bu

maddeler her bir boyuta eşit olacak şekilde dağıtılmıştır. Yani üç boyutlu testler için her boyutta 30 İKM ve 6 ÇKM yer almaktadır. İki boyutlu testler için her boyutta 45 ÇKM ve 6 ÇKM yer almaktadır. Bir boyutlu verilerde ise maddelerin tamamı aynı boyutta yer almaktadır. Şekilde belirtilen bu boyutluluk yapıları yalnızca içeriğe dayalı boyutluluğun nasıl oluştuğunu yansıtmaktadır.

Diğer taraftan, maddenin formatına dayalı olarak F1 ve F2 boyutları oluşturulmuştur. F1 boyutu İKM'lere ve F2 boyutu ise ÇKM'lere karşılık gelmektedir. Yani, İKM'ler F1 boyutundan yük almakta, F2 boyutundan yük almamaktadır. Diğer taraftan ÇKM'ler ise F2 boyutundan yük almakta fakat F1 boyutundan yük almamaktadır.

Çok boyutlu MTK modellerinde geleneksel tek boyutlu modellerde kullanılan ve madde gücü olarak bilinen  $b$  parametresi yerine eşik parametreleri ( $d$  parametresi) bulunmaktadır. Gerçekleştirilen bu çalışmada ÇKM'ler 5 kategoriden oluşacak şekilde türetilmiştir. Bu sebeple her bir madde için 4 adet eşik noktası vardır. Veri setlerinin türetilmesi aşamasında her bir ÇKM'ye ait 4'er evren eşik parametresi türetilmiştir. Ayrıca İKM'ler için evren madde gücü ( $b$ ) parametreleri türetilmiştir. Türetilen evren madde parametrelerine ait betimleyici istatistikler Çizelge 4.3'de yer almaktadır.

**Çizelge 3.4. Türetilen Evren Güçlük ( $b$ ) ve Eşik ( $d$ ) Parametrelerine Ait Betimsel İstatistiksel**

		<i>Alt Grup</i>		<i>Üst Grup</i>	
		$\bar{X}$	$X_{Std Sap.}$	$\bar{X}$	$X_{Std Sap.}$
<b><i>İKM</i></b>	<b><i>b</i></b>	0	1	1	1
<b><i>ÇKM</i></b>	<b><i>b<sub>1</sub></i></b>	1.5	0.1	2	0.1
	<b><i>b<sub>2</sub></i></b>	0.5	0.1	1	0.1
	<b><i>b<sub>3</sub></i></b>	-0.5	0.1	0	0.1
	<b><i>b<sub>4</sub></i></b>	-1.5	0.1	-1	0.1

### 3.5. Parametrelerin Kestirilmesi

Evren parametrelerinin türetilmesi işleminden sonra veri setleri türetilmiştir. Daha sonra ise türetilen veri setleri kalibre edilmiştir. Bu çalışmada karma format testlerin kalibrasyonu için 3PLM ve ATM'yi kullanılacaktır. Bu kombinasyon farklı birçok çalışmada tercih edilmiştir. Rosa, Swygert, Nelson, ve Thissen (2001) İKM'lerin kalibrasyonu için 3PLM tercih edilmesinin sebebini maddeye dair daha fazla

parametrenin dikkate alındığı bir model olması ve dolayısıyla daha fazla bilgi vermesi olarak ifade etmiştir. İlgili alan-yazın incelendiğinde ÇKM'lerin kalibrasyonunda ise aşamalı tepki modeli ile dereceli kısmi bilgi modellerinin tercih edildiği görülmektedir(Kim ve Cohen, 2002; Bastari, 2000; Tate, 2000). Dodd (1984) tarafından gerçekleştirilen bir çalışmada bu iki modelin kavramsal ve matematiksel olarak oldukça farklı olmalarına rağmen uygulamadaki sonuçların oldukça benzer olduğu sonucuna varılmıştır. Nitekim (Cao, Yin ve Gao, 2007) tarafından gerçekleştirilen bir diğer çalışmada da benzer sonuçlar elde edilmiş ve her iki modelin benzer sonuçlar verdiği bulunmuştur.

Ayrıca, bu çalışmada evren yetenek parametrelerine ait varyanslar kontrol edilmektedir ve bazı durumlarda ikinci test formundaki varyans aynı kalacak şekilde veri türetilmiş, bazı durumlarda ise varyansın azalacağı şekilde türetilmiştir. Varyansın azaldığı senaryoda varsayılandan daha fazla bir azalma olmaması için her iki senaryo için de sınıf içi varyansı daha yüksek olan MAP yönteminin kullanılması tercih edilmiştir.

Ölçekleme işleminin gerçekleştirilebilmesi için her bir veri seti ayrı kalibre edilmiştir. EM döngüleri için her bir veri setinin analizinde yakınsaklık değeri 0,001 ve döngü sayısı 500 olarak belirlenmiştir. MH-RM kestirim tekniği için ise yakınsaklık değeri 0,0001 ve döngü sayısı 2000 olarak belirlenmiştir.

### **3.6. Ölçekleme İşleminin Gerçekleştirilmesi**

Bilindiği üzere MTK temelli eşitleme çalışmalarında karakteristik eğri yöntemleri moment yöntemlerine göre daha az hataya sebep olmaktadır. Diğer taraftan aynı aileden gelen Stocking-Lord yöntemi ile Haebara yöntemleri farklı koşullar altında bile oldukça benzer sonuçlar vermektedir. Bu sebeple gerçekleştirilen bu çalışmada Haebara yöntemi tercih edilmiştir.

### **3.7. Değerlendirme Ölçütleri**

Önemli kararlar alınan testlerde adil, geçerli ve güvenilir sonuçların alınması oldukça önemlidir. Bundan dolayı eşitleme sonuçlarının kesinliğinin değerlendirilmesi kritik öneme sahiptir. Puanlara dair yorumların uygunluğunu arttırmak için ölçme uzmanları ölçme hatasını belirlemek için farklı psikometrik işlemler kullanmaktadırlar. Karma format testlerle ilgili psikometrik alan-yazın eşitleme sonuçlarının niteliğinin sistematik hatalar (Bias) ve rastlantısal hataların (RMSE)



üzerinde durmaktadır. RMSE değerleri ise kestirilen değerler ile bilinen değerler arasındaki tutarsızlığın miktarını belirlemek için kullanılırlar. Bu indeks kestirim hatasının toplam ölçüsüdür.

Yetenek parametrelerinin hata miktarlarını belirlemek için şu formülden yararlanılmıştır.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{\theta}_i - \theta_i)^2}{n}}$$

Burada  $\hat{\theta}_i$  ölçekleme işlemi sonucunda elde edilen yetenek parametrelerine  $\theta_i$  ise evren parametreleri kullanılarak türetilen verilerin kestirimi ile elde edilen yetenek parametrelerine karşılık gelmektedir. Formülde yer alan  $n$  değeri ise toplam birey sayısına karşılık gelmektedir.

Yetenek parametrelerine ait Yanlılık (Bias) değerleri ise şu formülle hesaplanmıştır.

$$\text{Bias} = \frac{\sum_{i=1}^n (\hat{\theta}_i - \theta_i)}{n}$$

Madde parametrelerine ait hata ve yanlılık değerlerinin hesaplanması için ise yukarıda yer alan formüllerde  $\theta$  değerlerinin yerine  $a$  ya da  $b$  değerleri konulmuştur. ÇKM'lere ait eşik parametreleri için hata ve yanlılık değerleri hesaplanırken her bir  $b$  parametresi için elde edilen değerlerin ortalaması alınmıştır.

### 3.8. Verilerin Analizi

Verilerin analizleri R istatistik programı (R Core Team, 2015) kullanılarak gerçekleştirilmiştir. Bu süreçte farklı R paketleri yüklenerek analizler gerçekleştirilmiştir. Öncelikli olarak d matrisleri türetilirken Trautmann ve arkadaşları (2014) tarafından geliştirilen “*truncnorm*” paketi kullanılmıştır. Bu paket ÇKM'ler için evren eşik parametrelerinin yer aldığı matrisleri türetilirken türetilen değerlerin alt ve üst sınırlarını kontrol etmek için tercih edilmiştir. Bu sayede ardışık eşik parametrelerinin arasındaki farkın 0,3'ün altına düşmemesi ve model-veri uyumunun bozulmaması sağlanmıştır. Diğer parametre türetme işlemleri ise R programında var olan “*norm*” komutu kullanılarak gerçekleştirilmiştir.

Daha sonra, Chalmers (2012) tarafından geliştirilen “*mirt*” paketi kullanılmıştır. Bu pakette yer alan “*sim*” komutu ile yanıt matrisleri türetilmiş, “*mirt*” komutu ile madde parametreleri türetilmiştir. Son aşamada ise “*fscores*” komutu kullanılarak yetenek parametreleri kestirilmiştir. Ölçekleme aşamasına geçildiğinde ise Weeks (2010) tarafından geliştirilen “*plink*” paketi kullanılmıştır.

Gerçekleştirilen analizlerin tamamı “*for*” döngüsü kullanılarak 50 defa tekrar edilmiştir. Analizler Windows işletim sistemi, 17 işlemcili, işlemci hızı 3.6 GHz, 16 GB DD3 RAM, 6MB ön belleğe sahip olan diz üstü bilgisayar ile gerçekleştirilmiştir. Analizlerin süresi her bir durum için farklılık göstermektedir. Farklılığa sebep olan etkenlerin test formlarının boyut sayısı ve kestirimde kullanılan yöntemler olduğu görülmüştür. Bu şartlarda 24 durumun her biri için analizler 6 saat ile 42 saat arasında sürelerde tamamlanmıştır.

Analizler tamamlandıktan sonra, parametrelerin hata ve yanlılık değerleri hesaplanarak ele alınan değişkenlere göre yorumlanmıştır. Daha sonra araştırmada ele alınan değişkenlerin etkisini ve etkileşimlerini görmek için ANOVA uygulanmıştır.

## 5. BULGULAR VE TARTIŞMA

Bu bölümde araştırma kapsamında elde edilen bulgular yer almaktadır. Her bir araştırma sorusu için ölçekleme hatası (RMSE), Yanlılık (Bias) değerlerine ve üç yönlü ANOVA'dan elde edilen sonuçlara yer verilmiştir. Çizelgelerde yer alan hata ve yanlılık değerleri yapılan tekrarların ortalamalarını yansıtmaktadır. ANOVA analizleri ise her bir durum için uygulanan 50 tekrardan elde edilen hata ve yanlılık değerleri ortalama alınmadan kullanılmıştır.

Ortak madde deseni kullanılarak gerçekleştirilen ölçeklemelerdeki önemli sorunlardan birisi ortak maddeler için hesaplanan hata ve yanlılık değerlerinin ortak olmayan maddeler için hesaplanan değerlere göre her zaman daha düşük olmasıdır (Yu, Sharon ve Osborn, 2005). Bu durumun sebebi, eşitleme ilişkilerinin ortak maddeler kullanılarak hesaplanmasıdır. Ortak maddelerin hata ve yanlılık değerlerinin hesaplanmasında kullanılması, eşitlemenin gerçekte olduğundan daha az hata ve yanlılığa sahip olduğu yanılgısını getirecektir. Bu sebeple, hata ve yanlılık değerleri hesaplanırken ortak madde olarak belirlenen maddeler hesaplamalara dâhil edilmemiştir. Hata ve yanlılık değerleri hesaplanması sürecinde bu değerler test formlarının boyutluluk durumu dikkate alınarak her bir boyut için ayrı ayrı hesaplanmıştır. Ayrıca, İKM'ler ve ÇKM'lere ait hata ve yanlılık değerleri de ayrı olarak hesaplanmıştır.

Hata ve yanlılık değerleri İKM'lerde  $a$  ve  $b$  parametreleri, ÇKM'ler için  $a$  ve eşik ( $b$ ) parametreleri ve yetenek parametreleri için ayrı olarak hesaplanmıştır. Ayrıca, iki ve üç boyutlu testler kullanılarak ölçeklemenin gerçekleştirildiği senaryolarda, her bir boyut için hata ve yanlılık değerleri ayrı olarak hesaplanmıştır. Örneğin üç boyutlu testlerle gerçekleştirilen ölçekleme işleminde yetenek parametresi için üç adet hata değeri, üç adet İKM'lere ve üç adet ÇKM'lere ait olmak üzere  $a$  parametresine ait altı adet hata değeri hesaplanmıştır. Aynı işlem  $b$  parametreleri için de tekrarlanmıştır. ÇKM'lere ait  $b_1$ ,  $b_2$ ,  $b_3$  ve  $b_4$  parametrelerinin ortalamaları alınarak hesaplamalar gerçekleştirilmiştir. Bu şekilde, İKM'lerin ve ÇKM'lerin  $b$  parametreleri için toplam 6 adet hata değeri hesaplanmıştır. Nihai olarak, toplamda hesaplanan hata değeri sayısı 15 adettir. Aynı şekilde 15 adet de yanlılık değeri hesaplanmıştır. Elde edilen hata değerlerine Çizelge 4.1, Çizelge 4.3 ve Çizelge 4.5'de, İki Yönlü ANOVA analizlerinden elde edilen sonuçlara ise Çizelge 4.2, Çizelge 4.4 ve Çizelge

4.6'da yer verilmiştir. Ayrıca Hata ve yanlılık değerleri için elde edilen bulgulara ilgili çizelgeler yorumlandıktan sonra Şekil 4.1 - Şekil 4.3'de yer verilmiştir.

#### 4.1. I. Alt Probleme Ait Bulgular ve Yorumlar

*Karma format maddelerden oluşan tek boyutlu testler DOGOM deseni kullanılarak ölçeklendiğinde ölçme hatası ve yanlılık; ortak madde setinin yapısına, yetenek daralmasına ve parametre kestirim yöntemlerine göre nasıl değişmektedir?*

Yöntem kısmında da belirtildiği üzere tek boyutlu veri setleri için maddelerin tamamı içeriğe dayalı aynı faktörden yük alacak şekilde ve diğer türetilmiştir. Tek boyutlu testler için elde edilen ölçme hatası ve yanlılık değerlerinin ortak madde setinin yapısı, yetenek daralması ve parametre kestirim yöntemleri koşullarına göre düzeylerinin yer aldığı bulgular Çizelge 4.1'de yer almaktadır. Bulgular ayırt edicilik parametreleri İKM'ler ve ÇKM'ler için ayrı hesaplanmış ve yorumlanmıştır.

**Çizelge 4.1. Tek Boyutlu Veri Yapısı için Koşullara Göre Parametrelerin Hata ve Yanlılık Değerleri**

		Eşit Varyans				Azalan Varyans			
		EM		MHRM		EM		MHRM	
		Hata	Yanlılık	Hata	Yanlılık	Hata	Yanlılık	Hata	Yanlılık
İki Kategorili Ortak Maddeler	Yetenek	0.047	2.591	0.052	2.782	0.048	2.718	0.054	2.791
	a parametresi	0.095	0.831	0.096	0.696	0.112	0.885	0.087	0.634
	b parametresi	0.301	2.417	0.300	2.413	0.300	2.410	0.314	2.530
	a parametresi	0.184	0.774	0.154	0.645	0.196	0.827	0.143	0.592
	b parametresi	0.484	2.054	0.498	2.111	0.486	2.063	0.531	2.254
Karma Format Ortak Maddeler	Yetenek	0.042	2.304	0.042	2.362	0.041	2,283	0.043	2.345
	a parametresi	0.015	0.042	0.016	0.060	0.011	0.063	0.012	0.052
	b parametresi	0.269	2,283	0.278	2.359	0.271	2,303	0.273	2.314
	a parametresi	0.020	0.038	0.038	0.056	0.026	0.059	0.025	0.049
	b parametresi	0.652	2,258	0.673	2.332	0.654	2,267	0.659	2.283

##### 4.1.1. Ortak Madde Yapısının Etkisi

Çizelge 4.1'de hata ve yanlılık değerleri ortak maddelerin yapısına göre incelendiğinde, karma ortak maddelerden oluşan testlerde yetenek, a parametreleri ve iki kategorili maddelerde b parametrelerinin her koşulda iki kategorili ortak

maddelerden oluşan testlere göre daha düşük hata ve yanlılık değerine sahip olduğu görülmektedir. Ancak çok kategorili maddelerin eşik değerlerine ait hata ve yanlılık değerleri, ortak maddelerin karma testlerden oluştuğu durumda iki kategorili maddelere göre daha yüksektir. Bu bulgu, karma testlerde ortak maddelerin de karma maddelerden oluşmasının yetenek,  $a$  ve iki kategorili maddelerde  $b$  parametrelerinin kestiriminin daha kesin olmasına yol açacağını göstermektedir. Elde edilen bu bulgular Cao (2008) tarafından tek boyutlu veriler kullanılarak gerçekleştirilen analizler sonucunda elde edilen bulgular ile tutarlıdır. Ortak madde setinin istatistiksel ve içerik açısından toplam testi temsil edebilme gücü ölçekleme sonuçları etkilemektedir (Hardy ve diğ., 2011). Çok kategorili maddelerde ise eşik değerleri, ortak maddelerin iki kategorili olduğu durumda daha düşük hata ve yanlılık değerleri elde edilmiştir.

#### **4.1.2. Yetenek Daralmasının Etkisi**

Eşitlenen testlerin varyanslarının eşit ya da azalan varyanslı olması durumuna göre incelendiğinde, ortak maddeleri iki kategorili maddelerden oluşan testler için yetenek parametreleri ile ÇKM'lerin eşik parametrelerine ait hata ve yanlılık değerlerinin varyansın azaldığı durumda daha düşük olduğu bulunmuştur. Ayrıca, EM kestirim yöntemi için İKM'lerin  $b$  parametrelerine ait hata ve yanlılık değerlerinin, MHRM kestirim yöntemi için ise İKM ve ÇKM'lerin  $a$  parametrelerine ait hata ve yanlılık değerlerinin varyansın eşit olduğu durumlarda daha düşüktür.

Ortak testin karma maddelerden oluştuğu durumda MHRM kestirim yöntemine göre, yetenek parametresine ait hata değerleri hariç, diğer parametrelerin tamamı için varyansın düşük olduğu testler için daha düşük hata ve yanlılık değerlerinin elde edildiği görülmüştür. Aynı durum EM kestirim yöntemi için incelendiğinde yetenek parametresine ait hata ve yanlılık değerleri ile İKM'lerin  $a$  parametresine ait hata değerlerinin yanlılık değerlerinin varyansın eşit olduğu testlerde daha düşük olduğu bulunmuştur.

#### **4.1.3. Kestirim Yönteminin Etkisi**

Tek boyutlu verilerle parametrelerin hata ve yanlılık değerleri EM ve MHRM kestirim yöntemlerine göre incelendiğinde, varyansın eşit ve iki kategorili ortak maddelerin kullanıldığı koşulda yetenek parametresi, İKM'lere ait  $b$  parametresi ve ÇKM'lere ait eşik parametresi için MHRM kestirim yönteminin daha az hata ve yanlılığa yol açtığı

görülmüştür. Azalan varyansın olduğu koşulda ise yetenek parametresi, İKM'lere ait *b* parametresi ve ÇKM'lere ait eşik parametresi için MHRM kestirim yönteminin daha az hata ve yanlılığa sebep olduğu bulunmuştur.

Karma format ortak maddelerin yer aldığı testler için bakıldığında ise eşit varyans koşulunda yetenek parametreleri ve azalan varyans koşulunda ise ÇKM'lere ait *a* parametreleri hariç, diğer parametrelerde MHRM kestirim yöntemine göre daha az hata ve yanlılık değerleri elde edildiği görülmüştür.

#### 4.1.4. Tek Boyutlu Veriler için Koşulların Etkisinin ve Etkileşimlerin İncelenmesi

Daha sonrasında gözlenen bu farklılaşmaların test edilen koşullara göre etkileşim içerisinde olup olmadıklarının belirlenebilmesi için üç yönlü ANOVA gerçekleştirilmiştir. Elde edilen bulgular Çizelge 4.2'de yer almaktadır.

**Çizelge 4.2. Tek Boyutlu Veriler için ANOVA Sonuçları**

		İKM			ÇKM	
		Yetenek	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>
Ortak maddenin yapısı	Hata	31.418**	512.023**	99.866**	165.084**	339.157**
	Yanlılık	183.61**	188.888**	24.229**	189.215**	23.489**
Varyans daralması (VD)	Hata	0.214	0.006	1.434	0.287	0.961
	Yanlılık	0.692	0.114	1.421	0.082	1.173
Kestirim yöntemi (KY)	Hata	5.424*	2.291	3.673	1.867	5.809*
	Yanlılık	10.843**	3.785	4.014*	3.846	5.840**
I.Boyut OMY*VD	Hata	0.588	1.256	1.745	0.033	2.022
	Yanlılık	2.180	0.298	1.758	0.247	2.150
OMY*KY	Hata	2.396	3.106	0.013	8.046**	0.895
	Yanlılık	1.534	5.377	0.001	5.452*	1.302
VD*KY	Hata	0.378	2.839	0.250	0.378	0.051
	Yanlılık	0.998	0.212	0.265	0.196	0.134
OMY*VD*KY	Hata	0.012	2.794	4.129*	0.012	2.412
	Yanlılık	1.070	0.152	4.216*	0.144	2.291

\* $p < 0.05$ ; \*\* $p < 0.01$ ;  $F(1,396)$

Gerçekleştirilen üç yönlü ANOVA sonuçlarına göre OMY'nin ölçekleme sonuçları üzerindeki ana etkisinin yetenek parametreleri ve madde parametrelerinin tamamı için hata ve yanlılık miktarları üzerinde ana etkisinin anlamlı olduğunu göstermiştir ( $p < .01$ ). Buna göre, Bu farklılıklar yetenek parametreleri, İKM'lere ait *a* ve *b* parametreleri ve ÇKM'lere ait *a* parametreleri için karma format ortak madde setinin

yer aldığı koşul lehinedir. ÇKM'lere ait eşik parametresi için ise gözlenen bu farklılaşma iki kategorili karma format koşulu lehinedir. Cao (2008) tarafından gerçekleştirilen çalışmada ortak maddelerin format temsili, istatistiksel temsil ve içerik temsili açısından etkilerinin incelendiği çalışmada bu faktörlerin ölçekleme sonuçları üzerindeki etkileri üç yönlü ANOVA ile incelenmiş ve herhangi bir etkinin olmadığı görülmüştür. Bu yönüyle elde edilen bu bulgu, Cao (2008) tarafından elde edilen bulgular ile örtüşmemektedir.

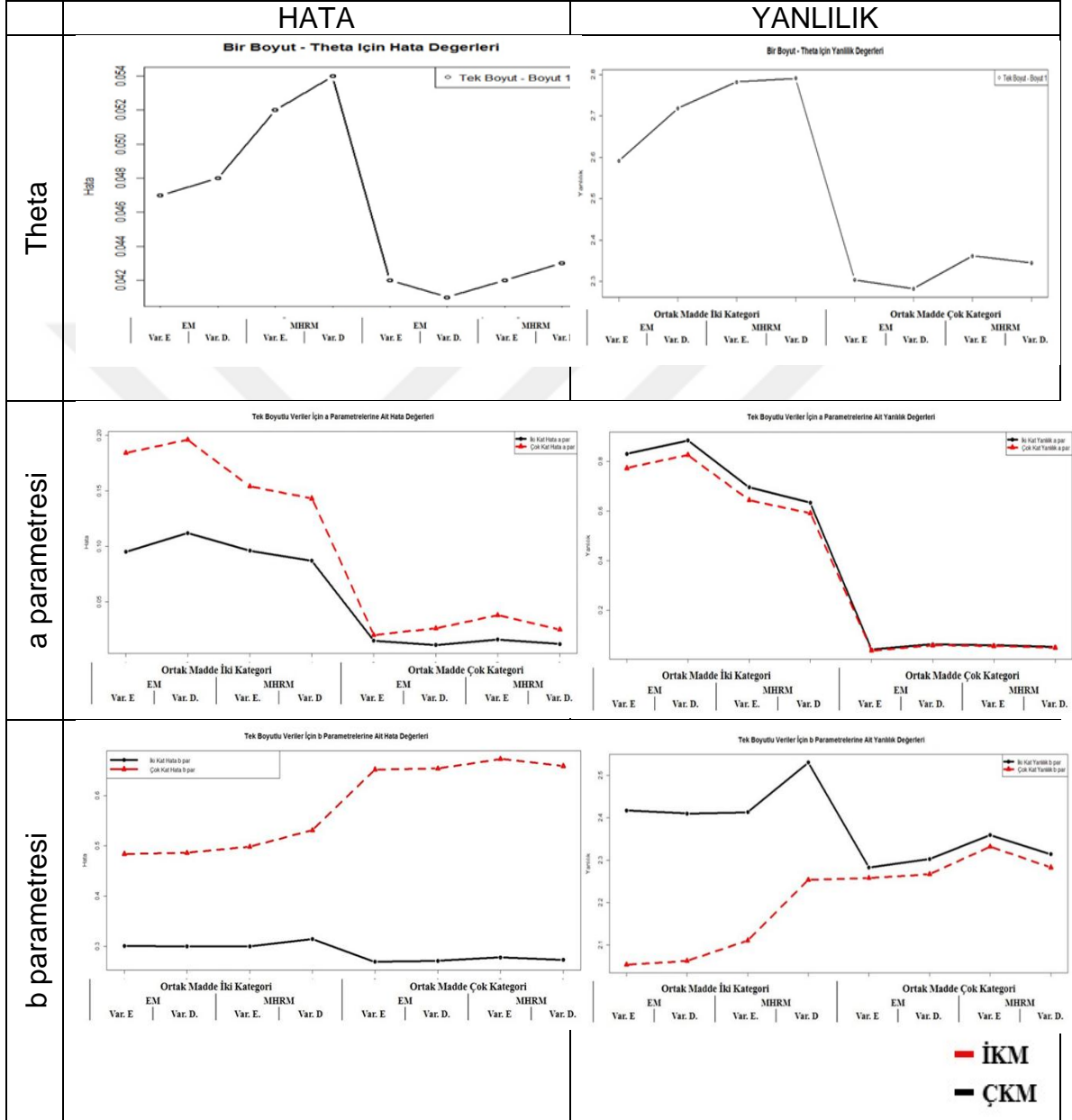
Diğer taraftan VD koşulunun ana etkiye sahip olmadığı görülmüştür ( $p < .05$ ). Buna göre hata ve yanlılık değerlerinin eşit ve azalan varyanslara göre anlamlı düzeyde farklılaşmadığı sonucuna varılabilir. KY koşulunun ana etkisi incelendiğinde ise yetenek parametresi ile ÇKM'lerin eşik parametresine ait hata ( $p < .05$ ) ve yanlılık ( $p < .05$ ) miktarları üzerinde ana etkisinin olduğu görülmüştür. Buna göre, yetenek parametresi ve eşik parametresi değerlerinin kullanılan kestirim yöntemine göre değişkenlik gösterdiği sonucu çıkmaktadır. EM kestirim yöntemi ölçekleme işlemi için bu parametrelerde daha kesin sonuçlar vermektedir. Ayrıca, İKM'lere ait b parametrelerine ait yanlılık değerleri için de ana etki bulunmaktadır. EM kestirim yöntemi bu parametre için de benzer şekilde daha az yanlı sonuçlar vermektedir.

OMY ve VD koşulları arasındaki ikişerli etkileşimler incelendiğinde anlamlı bir ana etkinin olmadığı görülmüştür ( $p < .05$ ). Elde edilen bu bulgu ortak maddelerin yapısına göre farklılık gösteren parametrelere ait hata ve yanlılık değerlerinin kestirim yöntemine göre değişkenlik göstermediği şeklinde yorumlanabilir.

OMY ve KY koşulları arasında hata indisleri için test edilen etkileşimlere bakıldığında ÇKM'lerin a parametreleri için elde edilen hata ( $p < .01$ ) ve yanlılık ( $p < .05$ ) değerlerinin bu iki koşula göre etkileşim içerisinde oldukları bulunmuştur. Bu bulguya göre ÇKM'lerin a parametrelerinden elde edilen hata ve yanlılık değerlerindeki OMY koşuluna göre gözlenen farklılaşmaların KY koşuluna göre değişkenlik gösterdiği görülmüştür. Ortak madde seti iki kategorili maddelerden oluştuğunda EM kestirim yöntemi MHRM kestirim yöntemine göre daha iyi performans gösterirken, ortak madde seti karma olduğu durumda ise MHRM kestirim yönteminin daha iyi performans gösterdiği bulunmuştur.

Diğer taraftan VD ve KY koşulları arasındaki ikişerli etkileşimler ve OM, VD ve KY koşulları arasındaki üçerli etkileşimler incelendiğinde hata değerlerinin bu koşullara

göre etkileşim içerisinde olmadıkları görülmüştür ( $p>0.05$ ). Tek boyutlu veriler için hesaplanan hata ve yanlılık değerlerinin koşullara göre değişimini karşılaştırmak için oluşturulan çizgi grafikleri Şekil 4.1’de yer almaktadır.



Şekil 4.1. Tek Boyutlu Veriler için Elde Edilen Hata ve Yanlılık Değerlerinin Her Bir Parametre için Koşullara Göre Karşılaştırması



## 4.2. II. Alt Probleme Ait Bulgular ve Yorumlar

*Karma format maddelerden oluşan iki boyutlu testler DOGOM deseni kullanılarak ölçeklendiğinde ölçme hatası ve yanlışlık; ortak madde setinin yapısına, yetenek daralmasına ve parametre kestirim yöntemlerine göre nasıl değişmektedir?*

İki boyutlu veriler için elde edilen hata ve yanlışlık değerleri koşullara göre Çizelge 4.3'de yer almaktadır.

**Çizelge 4.3. İki Boyutlu Veri Yapısı için Koşullara Göre Hata ve Yanlılık Değerleri**

		Eşit Varyans				Azalan Varyans			
		EM		MHRM		EM		MHRM	
		Hata	Yanlılık	Hata	Yanlılık	Hata	Yanlılık	Hata	Yanlılık
<i>İki Kategorili Ortak Maddeler</i>	<b>Birinci Boyut</b>								
	<i>Yetenek</i>	0.034	-2.024	0.038	-1.970	0.037	-1.952	0.038	-2.053
	a parametresi	0.015	0.082	0.025	0.125	0.022	0.100	0.015	0.076
	b parametresi (İKM)	0.554	3.164	0.578	3.303	0.548	3.129	0.576	3.293
	a parametresi	0.027	0.076	0.040	0.116	0.033	0.093	0.026	0.071
	b parametresi (ÇKM)	0.983	2.936	1.030	3.078	0.967	2.888	1.021	3.050
<i>Karma Format Ortak Maddeler</i>	<i>Yetenek</i>	0.034	-1.922	0.034	-1.897	0.035	-1.977	0.037	-2.011
	a parametresi	0.009	0.039	0.007	0.036	0.011	0.053	0.008	0.031
	b parametresi (İKM)	0.502	2.997	0.501	2.988	0.499	2.978	0.511	3.049
	a parametresi	0.018	0.036	0.017	0.033	0.022	0.049	0.015	0.029
	b parametresi (ÇKM)	1.135	2.769	1.152	2.810	1.133	2.765	1.161	2.832
<i>İki Kategorili Ortak Maddeler</i>	<b>İkinci Boyut</b>								
	<i>Yetenek</i>	0.032	-1.735	0.031	-1.776	0.028	-1.561	0.031	-1.693
	a parametresi	0.015	0.082	0.025	0.126	0.022	0.101	0.015	0.076
	b parametresi (İKM)	0.484	2.766	0.501	2.863	0.473	2.700	0.576	2.685
	a parametresi	0.027	0.076	0.041	0.118	0.033	0.092	0.026	0.072
	b parametresi (CKM)	0.860	2.936	0.893	3.078	0.834	2.888	1.021	3.050
<i>Karma Format Ortak Maddeler</i>	<i>Yetenek</i>	0.034	-1.770	0.032	-1.781	0.033	-1.808	0.033	-1.828
	a parametresi	0.009	0.039	0.007	0.036	0.011	0.053	0.008	0.031
	b parametresi (İKM)	0.473	2.821	0.473	2.824	0.466	2.781	0.474	2.825
	a parametresi	0.018	0.036	0.016	0.033	0.022	0.049	0.015	0.029
	b parametresi (ÇKM)	1.061	2.769	1.070	2.810	1.054	2.765	1.080	2.832

### 4.2.1. Ortak Madde Yapısının Etkisi

Çizelge 4.2'de yer alan iki boyutlu verilere ait hata ve yanlışlık değerleri ortak maddelerin yapısına göre incelendiğinde, birinci boyutta yer alan maddeler için EM kestirim yönteminin kullanıldığı kestirimlerde yetenek parametreleri için eşit varyans

koşulunda hata, azalan varyans koşulunda ise yanlılık değerlerinin karma format ortak madde setlerinin yer aldığı testlerde daha yüksek olduğu görülmüştür. Ayrıca, EM kestirim yönteminin kullanıldığı kestirimlerde ÇKM'lere ait eşik parametreleri için hata değerleri karma format ortak maddelerin yer aldığı testlerde daha yüksektir. MHRM kestirim yöntemi için bakıldığında ise eşit varyans koşulu için ÇKM'lere ait eşik parametrelerinin karma format ortak madde koşulunda daha fazla olduğu bulunmuştur. Genel olarak bakıldığında ise karma format ortak madde setleri kullanılarak oluşturulan testlerin daha kesin eşitleme sonuçları verdiği görülmektedir.

İkinci boyuta bakıldığında, varyansın daraldığı koşullar için yetenek parametresi ve İKM'lere ait  $b$  parametrelerinin hata ve yanlılık değerlerinin karma format ortak maddelerde daha yüksek olduğu görülmüştür. Ayrıca, ÇKM'lere ait eşik parametreleri için varyansın daraldığı ve MHRM kestirim yönteminin kullanıldığı koşul haricinde diğer koşullarda iki kategorili ortak maddelerin daha az ölçme hatasına yol açtığı bulunmuştur. Yetenek parametresine ait hatalar incelendiğinde, EM kestirim yönteminin kullanıldığı ve eşit varyansa sahip testlerde yine iki kategorili ortak maddelerin kullanılmasının daha fazla hataya yol açtığı bulunmuştur. Son olarak, ÇKM'lerin  $a$  parametresi için MHRM kestirim yönteminin kullanıldığı ve varyansın daraldığı koşullarda hata ve yanlılık değerlerinin karma format ortak madde durumu için daha fazla hata verdiği görülmüştür. Diğer taraftan, genel olarak, karma format ortak maddelerde hata ve yanlılık değerlerinin eşit varyans durumunda için daha düşük olduğu söylenebilir. Ayrıca EM kestirim yöntemi kullanıldığında İKM ve ÇKM'lerin  $a$  parametrelerine ait hata ve yanlılık değerlerinin eşit varyans durumunda için daha düşük olduğu bulunmuştur.

Yetenek parametresi için birinci boyutta karma format ortak madde seti olumsuz etkiye sahipken ikinci boyutta daha olumlu etkiye sahip olmuştur. Cao (2008) tarafından gerçekleştirilen çalışmada ortak madde setinin çok boyutlu veriler bağlamında etkilerini incelemiş ve karma format ortak madde setinin olumlu etkiye sahip olduğu sonucuna varmıştır. Bu açıdan, birinci boyut için elde edilen bulgular ilgili alan-yazın ile örtüşmemektedir.

#### 4.2.2. Yetenek Daralmasının Etkisi

Çizelge 4.'2 de hata ve yanlışlık değerleri varyans daralması koşuluna göre incelendiğinde, birinci boyut için karma ortak maddelerden oluşan testlerde yetenek parametrelerinden elde edilen hata ve yanlışlık değerlerinin varyansın eşit olduğu durumlarda daha düşük olduğu görülmüştür. İki kategorili ortak maddelerin yer aldığı testlerde ise EM kestirim yönteminden elde edilen hata ve MHRM kestirim yönteminden elde edilen yanlışlık değerleri eşit varyansın eşit olduğu durum için daha düşük olduğu bulunmuştur. Ortak madde yapısı karma format olduğunda ve MHRM kestirim yöntemi kullanıldığında da eşit varyans durumu için İKM ve ÇKM'lerin *a* parametreleri daha düşük hata ve yanlışlık değerleri vermektedir.

İkinci boyut için elde edilen bulgular karma format ortak madde kullanılan durumlar için birinci boyut ile benzer olduğu görülmüştür. Burada tek istisna EM kestirim yöntemi kullanıldığında yetenek parametresine ait hata miktarının azalan varyans durumunda daha az hata vermesidir. Ayrıca iki kategorili ortak maddeler kullanıldığında yetenek parametresi için hata ve yanlışlık değerlerinin varyansın azaldığı durumda daha az hata ve yanlışlık ortaya çıkardığı bulunmuştur. Varyans daralması ile ilgili alan-yazın incelendiğinde bunun sebeplerinin ele alan çalışmalar (Yen, 1985; Camilli, 1993) olsa da ölçekleme üzerindeki etkileri ile ilgili herhangi bir çalışmaya rastlanmamıştır.

#### 4.2.3. Kestirim Yönteminin Etkisi

İki boyutlu testlerde parametrelerin hata ve yanlışlık değerleri EM ve MHRM kestirim yöntemlerine göre incelendiğinde, birinci boyut için varyansın azaldığı durumda yetenek parametresi, İKM'lerin *b* parametreleri ve ÇKM'lerin eşik parametreleri için EM kestirim yöntemi kullanıldığında daha az hata ve yanlışlık olduğu görülmüştür. Varyansın eşit olduğu koşullarda ise iki kategorili ortak maddelerin yer aldığı testlerde parametrelerin tamamı için hata ya da yanlışlık değerlerinden en az birinin EM kestirim yönteminde daha düşük olduğu görülmüştür. Diğer taraftan karma format ortak maddelere sahip testler için yalnızca ÇKM'lerin eşik parametrelerinde EM döngülerinin daha az hata ve yanlışlığa sebep olduğu bulunmuştur. Diğer koşullarda ise MHRM döngülerinin kullanıldığı testlerde daha az hata ve yanlışlık gözlenmiştir.

Bulgular ikinci boyutta yer alan maddeler için incelendiğinde, ÇKM'lerin eşik parametresi için, koşulların tamamında EM kestirim yöntemi kullanıldığında daha az hata ve yanlılık değerleri elde edildiği sonucuna varılmıştır. Ayrıca, iki kategorili ortak maddelerin yer aldığı ve varyansın daraldığı durumlar için yetenek parametresine ait hata miktarları hariç diğer tüm parametrelerdeki hata ve yanlılık değerlerinin EM kestirim yönteminde daha az olduğu görülmüştür. Ayrıca, yetenek parametrelerine ait yanlılık değerleri de yine EM döngülerinin kullanıldığı veriler için daha düşük olduğu sonucuna varılmıştır. Alan-yazında kestirim yönteminin eşitleme sonuçları üzerindeki etkisini inceleyen bir çalışma bulunmamaktadır.

#### 4.2.4. İki Boyutlu Veriler için Koşulların Etkisinin ve Etkileşimlerin İncelenmesi

Gözlenen hata ve yanlılık değerlerinin koşullara göre etkileşimlerinin incelenmesi için gerçekleştirilen analiz sonuçlarına Çizelge 4.4'de yer verilmiştir.

**Çizelge 4.4. İki Boyutlu Veriler İçin ANOVA Sonuçları**

			İKM			ÇKM	
			Yetenek	a	b	a	b
I. Boyut	Ortak maddenin yapısı (OMY)	Hata	0.930	78.982**	70.157**	15.809**	103.728**
		Yanlılık	3.171	24.476**	27.511**	24.041**	23.309**
	Varyans daralması (VD)	Hata	0.265	0.020	0.002	0.175	0.088
		Yanlılık	2.825	0.219	0.000	0.189	0.130
	Kestirim yöntemi (KY)	Hata	1.900	0.180	4.73*	0.021	6.627**
		Yanlılık	0.267	0.012	4.778*	0.013	6.549*
	OMY*VD	Hata	0.027	2.049	0.274	0.502	0.315
		Yanlılık	2.134	0.829	0.277	0.784	0.343
	OMY*KY	Hata	1.466	1.729	2.126	0.940	0.966
		Yanlılık	0.114	0.987	2.064	0.877	1.476
	VD*KY	Hata	0.198	12.273**	0.410	3.541	0.112
		Yanlılık	3.981*	3.564	0.398	3.518	0.089
	OMY*VD*KY	Hata	1.757	13.284**	0.088	1.207	0.007
		Yanlılık	0.773	1.108	0.110	1.112	0.002
II. Boyut	Ortak maddenin yapısı (OMY)	Hata	5.881*	76.754**	2.577	16.309**	282.598**
		Yanlılık	345.484**	24.451**	2.482	24.277**	23.309**
	Varyans daralması (VD)	Hata	0.033	0.000	3.538	0.222	2.597
		Yanlılık	478.543**	0.191	3.518	0.236	0.130
	Kestirim yöntemi (KY)	Hata	1.248	0.096	0.728	0.011	1.834
		Yanlılık	328.951**	0.018	0.738	0.003	6.549*
	OMY*VD	Hata	0.000	1.650	1.947	0.586	3.021
		Yanlılık	556.353**	0.763	1.865	0.866	0.343
	OMY*KY	Hata	1.494	1.984	0.054	1.088	0.001
		Yanlılık	351.751**	0.982	0.057	1.026	1.476
	VD*KY	Hata	1.123	12.707**	0.237	3.432	0.090
		Yanlılık	179.019**	3.697	0.214	3.405	0.089
	OMY*VD*KY	Hata	0.118	13.789**	1.037	1.210	0.968
		Yanlılık	183.929**	1.150	1.029	1.080	0.002

\*p<0.05; \*\*p<0.01; F(1,396)

Koşullara ait ana etkiler ve bunlar arasındaki etkileşimler birinci boyut için OMY'nin ölçekleme sonucunda elde edilen hatalar üzerindeki ana etkisinin ÇKM ve İKM'lere ait madde parametrelerinin tamamı için ana etkisinin anlamlı olduğu bulunmuştur ( $p < .01$ ). Bu bulgu madde parametrelerine ait hata ve yanlılık değerlerinin ortak madde yapısına göre farklılaştığını göstermektedir. Gözlenen bu farklılaşmalar, ortak madde seti karma format maddelerden oluştuğunda ölçekleme sonuçlarının daha kesin olacağını göstermektedir. Elde edilen bu bulgu Cao (2008) tarafından elde edilen bulgularla kısmen örtüşmektedir. Cao tarafından yapılan bu çalışmada ölçekleme senaryosunun kullanıldığı ve çok boyutlu veriler ile yapılan ölçekleme işlemlerinde, boyutlar arası ilişkilere bağlı olarak bazı durumlarda format değişkeninin hata ve yanlılık değerleri üzerinde ana etkisinin olduğu bulunmuştur.

Varyans daralması için ana etkiler incelendiğinde ise bu koşula göre hata ve yanlılık değerlerinin anlamlı düzeyde farklılaşmadığı görülmüştür. Son olarak İKM'lerin  $b$  parametreleri ( $p < .05$ ) ile ÇKM'lerin eşik parametreleri ( $p < .01$ ) için yanlılık değerlerini üzerinde kestirim yönteminin anlamlı düzeyde ana etkisinin olduğu bulunmuştur. Buna göre güçlük ve eşik parametrelerine ait yanlılık değerlerinin EM kestirim yöntemi kullanıldığında daha az hatalı ve yanlı sonuçlar verdiği sonucu çıkmaktadır.

Birinci boyut için etkileşimler incelendiğinde ise OMY ve VD koşulları ile OMY ve KY koşulları arasındaki etkileşimlerin anlamlı düzeyde olmadığı, diğer taraftan VD ve KY değişkenleri için ise yetenek parametrelerine ait yanlılık ( $p < .05$ ) ve İKM'lerin  $a$  parametrelerine ( $p < .01$ ) ait hata değerleri açısından etkileşim olduğu görülmüştür. Buna göre EM kestirim yöntemi yetenek varyansının daraldığı senaryoda daha iyi performans gösterirken, varyansın eşit olduğu senaryoda MHRM kestirim yöntemi daha iyi performans göstermektedir.

Son olarak OMY, VD ve KY koşulları arasındaki üçerli etkileşimler incelendiğinde yetenek parametresi için elde edilen yanlılık değerleri ( $p < .01$ ) ve İKM'lerin  $a$  parametreleri için elde edilen hata değerleri için ( $p < .01$ ) bu üç koşula göre anlamlı düzeyde etkileşim içerisinde oldukları bulunmuştur. Varyans eşit olduğunda karma format ortak madde yapısı ve EM kestirim yöntemi daha iyi sonuçlar verirken, varyansın daraldığı durumlarda ortak madde yapısının bir etkisi olmazken MHRM kestirim yönteminin daha iyi sonuçlar verdiği görülmüştür.

İkinci boyut için elde edilen bulgular incelendiğinde yetenek parametresi ve İKM'lerin *b* parametrelerine ait hata ve yanlışlık değerleri haricinde diğer parametrelerin hata ve yanlışlık değerleri için parametreleri ortak madde yapısının ana etkiye sahip olduğu bulunmuştur. Buna göre belirtilen parametrelere ait hata ve yanlışlık değerlerinin ortak madde yapısı sadece İKM'lerden oluşan ve karma format maddelerden oluşan testler için farklılaştığı ve ortak madde yapısı karma format olan testlerin daha iyi performans gösterdikleri bulunmuştur.

VD koşulu için test edilen ana etkiler incelendiğinde yetenek parametresi için elde edilen yanlışlık değerleri üzerinde bu koşulun ana etkiye sahip olduğu bulunmuştur ( $p < .01$ ). Buna göre yetenek parametresi için elde edilen sistematik hata değerleri varyansın azaldığı senaryoda daha az değerler almakta ve ölçeklemenin daha kesin sonuçlar vermesini sağlamaktadır. Benzer şekilde, KY koşulunun da yetenek parametrelerine ait yanlışlık değerleri üzerinde anlamlı etkisinin olduğu bulunmuştur ( $p < .01$ ). Buradaki farklılaşma, EM kestirim yönteminin daha az hatalar verdiğini göstermektedir.

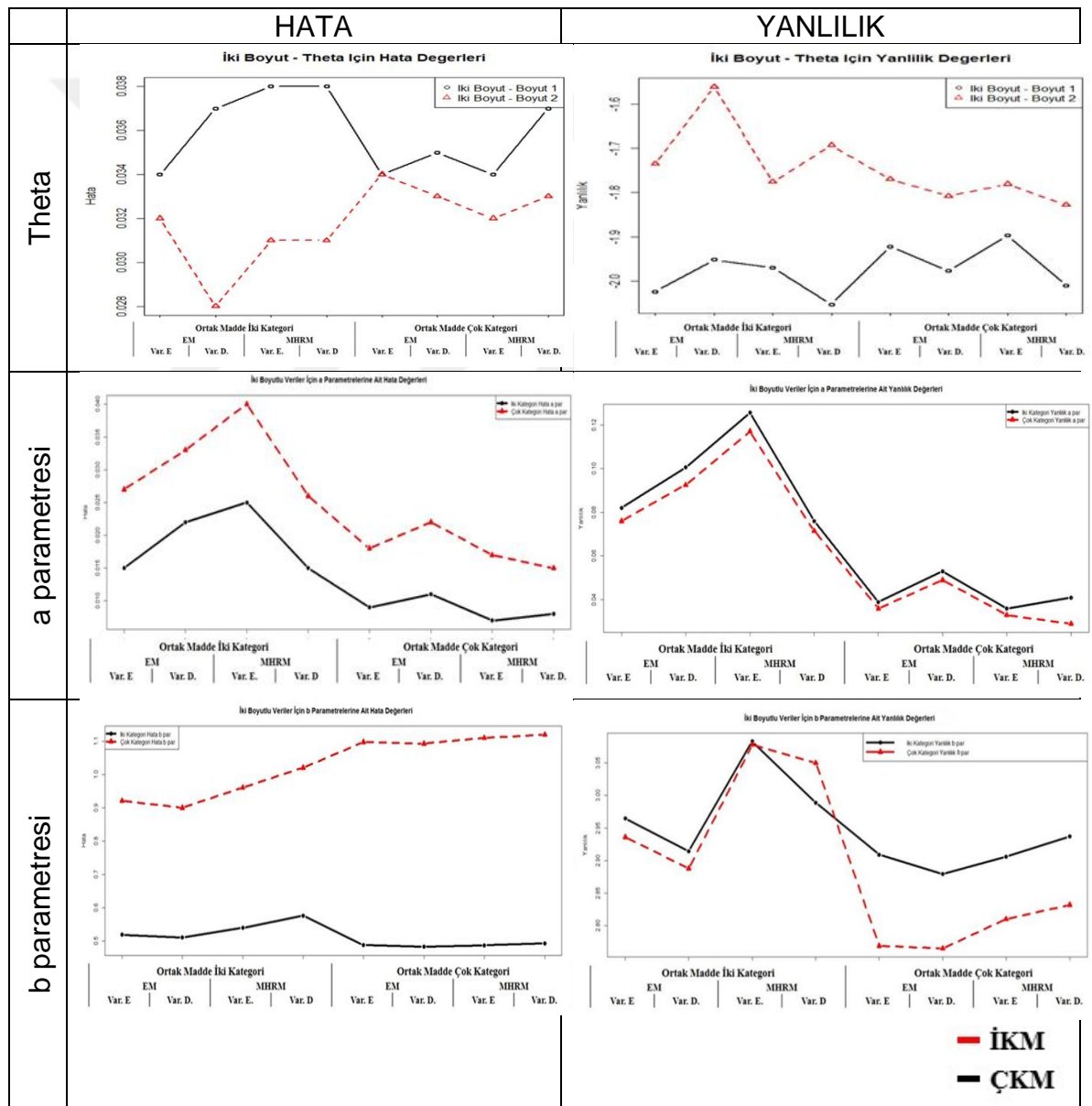
İkinci boyut için koşullar arasındaki etkileşimlerin incelenmesi sonucunda elde edilen bulgulara göre OMY ve VD koşulları ( $p < .01$ ) ile OMY ve KY koşullarının ( $p < .01$ ) yetenek parametresine ait yanlışlık değerlerine göre etkileşim içerisinde bulunduğu görülmüştür. Buna göre yanlışlık değerlerinin OMY koşuluna göre farklılaşması KY ve VD koşullarının farklı seviyelerine göre değişkenlik göstermektedir. İki kategorili ortak madde setinde varyans daralması daha iyi sonuçlar verirken, karma format ortak madde yapısında eşit varyansa sahip testlerin ölçeklenmesi daha iyi sonuçlar vermektedir. Ayrıca, iki kategorili ortak madde setinde MHRM kestirim yöntemi daha iyi sonuçlar verirken karma format ortak madde yapısında bu iki kestirim yöntemi yaklaşık olarak aynı sonucu vermiştir.

VD ve KY koşullarının ise yetenek parametresine ait yanlışlık değerleri ile İKM'lerin *a* parametresine ait yanlışlık değerleri için etkileşim gösterdikleri bulunmuştur ( $p < .01$ ). Genel olarak varyansın daralmasının ölçekleme sonuçlarını daha olumsuz etkilediği ayrıca MHRM kestirim yöntemi için bu etkinin daha da arttığı görülmüştür.

Son olarak, OMY, VD ve KY koşulları arasındaki üçlü etkileşim incelendiğinde yetenek parametresine ait yanlışlık değerleri ile İKM'lerin *a* parametresine ait hata değerleri için bu üç koşulun etkileşim içerisinde oldukları bulunmuştur. Yetenek

parametrelerine ait yanlışlık değerlerinin eşit varyans durumunda karma ortak maddeli testler için daha iyi sonuçlar verdiği görülürken, değişen varyans durumunda iki kategorili ortak maddeler için daha iyi sonuçlar verdiği görülmüştür. Ayrıca, eşit varyans durumunda EM kestirimleri daha iyi sonuç verirken, varyansın daraldığı senaryoda MHRM kestirim yöntemi daha az hata ve yanlışlık doğurduğu görülmüştür.

İki boyutlu veriler için hesaplanan hata ve yanlışlık değerlerinin koşullara göre değişimini karşılaştırmak için oluşturulan çizgi grafikleri Şekil 4.2’de yer almaktadır.



Şekil 4.2. İki Boyutlu Veriler için Elde Edilen Hata ve Yanlılık Değerlerinin Her Bir Parametre için Koşullara Göre Karşılaştırması

### 4.3. III. Alt Probleme Ait Bulgular ve Yorumlar

Üç boyutlu veriler için elde edilen değerler Çizelge 4.3'de yer almaktadır.

**Çizelge 4.5. Üç Boyutlu Veri Yapısı için Koşullara Göre Hata ve Yanlılık Değerleri**

		Esit Varvans				Azalan Varvans			
		EM		MHRM		EM		MHRM	
		Hata	Yanlılık	Hata	Yanlılık	Hata	Yanlılık	Hata	Yanlılık
<i>İki Kategorili Ortak Maddeler</i>	<b>Birinci Boyut</b>								
	<i>Yetenek</i>	0.068	-3.654	0.060	-3.294	0.060	-3.227	0.058	-3.431
	a parametresi (İKM)	0.074	-0.241	0.086	-0.340	0.069	-0.351	0.072	-0.337
	b parametresi (İKM)	0.624	2.911	0.641	2.992	0.604	2.817	0.669	3.119
	a parametresi (ÇKM)	0.132	-0.025	0.139	-0.036	0.136	-0.045	0.132	0.035
	b parametresi (ÇKM)	1.116	0.771	1.066	0.835	1.140	0.774	1.074	0.836
<i>Karma Format Ortak Maddeler</i>	<i>Yetenek</i>	0.034	-1,738	0.032	-1,736	0.032	-1,744	0.033	-1,964
	a parametresi (İKM)	0.050	0.158	0.050	0.104	0.045	0.158	0.052	0.101
	b parametresi (İKM)	0.457	2,226	0.475	2,315	0.433	2,111	0.489	2,384
	a parametresi (ÇKM)	0.123	0.034	0.129	0.023	0.121	0.035	0.130	0.019
	b parametresi (ÇKM)	1,308	0.512	1,189	0.498	1,246	0.547	1,233	0.496
<i>İki Kategorili Ortak Maddeler</i>	<b>İkinci Boyut</b>								
	<i>Yetenek</i>	0.053	-2.858	0.049	-2.543	0.047	-2.568	0.050	-2.449
	a parametresi (İKM)	0.074	-0.244	0.084	-0.347	0.070	-0.351	0.087	-0.339
	b parametresi (İKM)	0.556	2.595	0.590	2.754	0.578	2.696	0.590	2.755
	a parametresi (ÇKM)	0.128	0.023	0.130	0.022	0.122	0.027	0.122	0.021
	b parametresi (ÇKM)	1.206	0.771	1.175	0.835	1.258	0.774	1.171	0.836
<i>Karma Format Ortak Maddeler</i>	<i>Yetenek</i>	0.030	-1,751	0.033	-1,798	0.034	-1,874	0.032	-1,909
	a parametresi (İKM)	0.050	0.174	0.052	0.121	0.046	0.172	0.052	0.116
	b parametresi (İKM)	0.528	2,576	0.543	2,644	0.491	2,392	0.552	2,689
	a parametresi (ÇKM)	0.139	-0,074	0.139	-0,052	0.142	-0,073	0.133	-0,053
	b parametresi (ÇKM)	1,242	0.512	1,175	0.498	1,194	0.547	1,188	0.496
<i>İki Kategorili Ortak Maddeler</i>	<b>Üçüncü Boyut</b>								
	<i>Yetenek</i>	0.040	-2.344	0.039	-2.269	0.040	-2.168	0.034	-2.177
	a parametresi (İKM)	0.077	-0.250	0.087	-0.354	0.070	-0.361	0.080	-0.345
	b parametresi (İKM)	0.670	3.127	0.759	3.537	0.717	3.346	0.731	3.412
	a parametresi (ÇKM)	0.142	-0.037	0.150	-0.060	0.145	-0.059	0.140	-0.058
	b parametresi (ÇKM)	1.620	0.771	1.643	0.835	1.678	0.774	1.615	0.836
<i>Karma Format Ortak Maddeler</i>	<i>Yetenek</i>	0.016	-0.762	0.016	0.779	0.016	-0.701	0.016	-0.764
	a parametresi (İKM)	0.046	0.137	0.049	0.081	0.043	0.137	0.051	0.079
	b parametresi (İKM)	0.422	2,055	0.459	2,237	0.413	2,015	0.460	2,241
	a parametresi (ÇKM)	0.160	0.175	0.157	0.171	0.165	0.170	0.153	0.171
	b parametresi (ÇKM)	1,326	0.512	1,243	0.498	1,251	0.547	1,275	0.496

#### 4.3.1. Ortak Madde Yapısının Etkisi

Üç boyutlu testlerde koşullara göre hata ve yanlılık değerleri ortak maddelerin yapısına göre incelendiğinde, birinci boyut için ÇKM'lerin eşik parametrelerinde hata



değerlerinin karma format ortak madde setlerinin kullanıldığı durumlarda daha yüksek olduğu bulunmuştur. Ayrıca, yine ÇKM'lere ait  $a$  parametreleri için eşit varyans koşulunda EM döngüleri kullanıldığında hata değerlerinin, MHRM kestirim yöntemi kullanıldığında ise yanlılık değerlerinin karma format ortak madde durumunda daha fazla hataya yol açtığı bulunmuştur.

İkinci boyuta göre bakıldığında, ÇKM'lere ait  $a$  parametrelerinin her koşul altında, karma format ortak maddelerin kullanıldığı durumlarda daha fazla hataya sahip olduğu görülmüştür. Ayrıca, ÇKM'lerin eşik parametreleri için varyansın daraldığı ve MHRM kestirim yönteminin kullanıldığı durumlarda da yine karma format ortak madde yapısının daha fazla hataya sebep olduğu bulunmuştur. Diğer tüm bulgular ise karma format veri seti kullanılan ölçekleme işlemleri için daha az hata ve yanlılık değerleri verdiğini göstermektedir.

Son olarak, üçüncü boyut için bakıldığında ÇKM'lere ait  $a$  parametresi hariç, diğer parametrelerin tamamında karma format ortak maddelerin hata ve yanlılık miktarlarının daha düşük olduğu bulunmuştur. İlgili alan-yazında yer alan Meng (2007), Cao (2008), Walker ve Kim (2010) ve Kim ve Lee (2006) tarafından gerçekleştirilen çalışmalarda elde edilen bulgular, mevcut bulgularla örtüşmektedir.

#### **4.3.2. Yetenek Daralmasının Etkisi**

Hata ve yanlılık değerleri yetenek daralmasına göre incelendiğinde, EM kestirimlerinin ve iki kategorili ortak maddelerin kullanıldığı koşullarda ÇKM'lerdeki  $a$  ve eşik parametrelerinden elde edilen hata ve yanlılık değerlerinin varyansın eşit olduğu koşullar için daha düşük olduğu bulunmuştur. Ayrıca, iki kategorili ortak maddelerin kullanıldığı testlerde eşik parametresi için elde edilen hata ve yanlılık miktarının varyansın azaldığı durumlarda daha yüksek olduğu bulunmuştur. Karma format ortak maddeler ve MHRM kestirim yöntemi kullanıldığında varyansın eşit olduğu durumlarda hata miktarları madde parametrelerinin tamamı için daha düşük çıktığı görülmüştür.

İkinci boyut için bakıldığında, EM döngülerinin ve iki kategorili ortak maddelerin kullanıldığı koşullarda madde parametrelerine ait yanlılık miktarlarının eşit varyans koşulu için daha düşük olduğu bulunmuştur. Karma format ortak maddelerin kullanıldığı veri setleri için yetenek parametresine ait yanlılık değerlerinin de yine varyansın eşit olduğu durumda daha düşük olduğu sonucuna varılmıştır. Diğer

tarafından, gözlenen diğer farklılaşmaların sistematik bir düzene sahip olmadığı görülmüştür.

Son olarak, üçüncü boyuta ait bulgulara bakıldığında EM döngülerinin ve iki kategorili ortak maddelerin kullanıldığında madde parametrelerine ait hata ve yanlışlık miktarlarının eşit varyans durumunda daha düşük olduğu görülmüştür. Bunun tek istisnası ise İKM'lere ait  $a$  parametreleri için hesaplanan hata değerleridir. Bu değerler yeteneğin daraldığı durumlarda daha yüksektir. Yetenek parametrelerinin tamamı için yeteneğin daraldığı durumlarda daha düşük hata ve yanlışlık miktarları gözlenmiştir. Karma format ortak maddeler ve MHRM kestirimleri kullanıldığında İKM'lerin  $a$  ve  $b$  parametreleri ile ÇKM'lerin eşik parametreleri için varyansın eşit olduğu durumlarda daha az hataya yol açtığı görülmüştür.

#### **4.3.3. Kestirim Yönteminin Etkisi**

hata ve yanlışlık değerlerinin üç boyutlu veriler için kullanılan kestirim yöntemine göre nasıl etkilendiği incelendiğinde, birinci boyuttaki maddelerde İKM'lerin  $a$  ve  $b$  parametreleri için varyans eşit olduğunda hata ve yanlışlık değerlerinin, yetenek daraldığında ise sadece hata değerlerinin EM kestirim yönteminde daha düşük olduğu bulunmuştur. Ayrıca, iki kategorili ortak maddelerin yer aldığı veriler için ÇKM'lerin  $a$  parametrelerinden elde hata ve yanlışlık değerleri EM kestirim yöntemi için daha düşüktür. İkinci boyut için elde edilen bulgular incelendiğinde, bu değerlerin de birinci boyutta elde edilen değerler ile benzer değişimler gösterdiği bulunmuştur. Birinci boyuttan farklı olarak bu boyut için karma format ortak madde setinin yer aldığı ve varyansın eşit olduğu durumlarda yetenek parametrelerine ait yanlışlık değerlerinin daha düşük olduğu görülmüştür. Bir diğer farklılık ise, varyansın eşit olduğu ve karma format ortak maddelerin kullanıldığı testlerde İKM'lerin  $a$  parametresine ait yanlışlık değerinin MHRM kestirim yöntemi için daha düşük olmasıdır.

Son olarak, elde edilen bulgular üçüncü boyut için incelendiğinde, yeteneğin daralmadığı ve iki kategorili ortak maddelerin yer aldığı koşullarda, madde parametrelerinin tamamı için EM kestirim yöntemi kullanıldığında daha az hata ve yanlışlık değerleri elde edilmiştir. Ayrıca, İKM'lere ait  $a$  ve  $b$  parametrelerinde hata miktarları durumların tamamı için EM kestirim yöntemi kullanıldığında daha düşük olduğu bulunmuştur. Yetenek parametresi için varyansın değiştiği ve karma format

ortak madde kullanılan koşulda EM kestirim yönteminin kullanılması daha az yanlılığa yol açmış, diğer koşulların tamamında MHRM kestirim yöntemi daha iyi performans gösterdiği görüşmüştür.

#### 4.3.4. Üç Boyutlu Veriler İçin Koşulların Etkisinin ve Etkileşimlerin İncelenmesi

Üç boyutlu veriler için üç yönlü ANOVA sonuçları 4.6'da yer almaktadır.

**Çizelge 4.6. Üç Boyutlu Veriler için ANOVA Sonuçları**

		İKM			ÇKM		
		Yetenek	a	b	a	b	
I. Boyut	Ortak maddenin yapısı (OMY)	Hata	298.756**	99.88**	354.555**	9.947**	64.491**
		Yanlılık	790.483**	312.124**	267.519**	113.801**	150.291**
	Varyans daralması (VD)	Hata	3.963*	7.323**	0.001	0.103	0.039
		Yanlılık	0.265	1.214	0.004	0.959	0.157
	Kestirim yöntemi (KY)	Hata	1.620	2.783	18.726**	2.654	11.586**
		Yanlılık	0.001	3.800	18.940**	1.542	0.412
	OMY*VD	Hata	1.748	0.828	0.208	0.010	0.515
		Yanlılık	4.150*	1.068	0.211	0.487	0.090
	OMY*KY	Hata	2.633	1.714	0.056	1.046	0.049
		Yanlılık	1.953	0.121	0.016	1.170	4.051*
	VD*KY	Hata	0.841	0.054	5.577*	0.419	1.581
		Yanlılık	13.896**	1.167	5.617*	0.436	0.165
	OM*VD*KY	Hata	0.443	4.341*	0.076	1.407	2.811
		Yanlılık	1.567	1.320	0.049	1.097	0.135
II. Boyut	Ortak maddenin yapısı (OMY)	Hata	114.524**	98.007**	17.307**	11.491**	0.023
		Yanlılık	197.916**	333.092**	4.709*	254.163	150.291**
	Varyans daralması (VD)	Hata	0.955	1.476	0.019	1.244	0.032
		Yanlılık	0.836	1.101	0.025	0.039	0.157
	Kestirim yöntemi (KY)	Hata	3.629	4.941*	6.300*	0.140	6.532
		Yanlılık	5.209*	3.818	6.417*	2.802	0.412
	OMY*VD	Hata	0.178	0.666	1.066	0.503	1.256
		Yanlılık	6.971**	0.840	1.079	0.034	0.090
	OMY*KY	Hata	2.369	7.065**	0.345	0.492	0.366
		Yanlılık	3.020	0.027	0.401	4.888	4.051*
	VD*KY	Hata	3.555	3.778	0.272	0.377	0.007
		Yanlılık	2.282	1.206	0.315	0.041	0.165
	OMY*VD*KY	Hata	0.074	0.636	1.990	0.391	2.481
		Yanlılık	0.098	1.334	2.037	0.013	0.135
III. Boyut	Ortak maddenin yapısı (OMY)	Hata	321.824**	115.478**	584.166**	12.644**	312.86**
		Yanlılık	938.236**	302.728**	488.666**	2667.632**	150.291**
	Varyans daralması (VD)	Hata	7.121**	3.017	0.079	0.179	0.023
		Yanlılık	2.648	1.098	0.067	2.223	0.157
	Kestirim yöntemi (KY)	Hata	6.009*	3.205	16.128**	0.343	1.419
		Yanlılık	0.135	4.129*	16.029**	2.305	0.412
	OMY*VD	Hata	1.225	0.566	0.360	0.157	0.795
		Yanlılık	2.108	0.998	0.350	0.738	0.090
	OMY*KY	Hata	2.401	3.281	0.159	1.428	0.050
		Yanlılık	0.515	0.062	0.096	1.164	4.051*
	VD*KY	Hata	0.049	1.816	1.964	1.742	0.066
		Yanlılık	0.510	1.410	1.857	2.787	0.165
	OMY*VD*KY	Hata	1.243	2.002	3.193	0.055	5.479*
		Yanlılık	0.240	1.448	3.074	1.255	0.135

\*p<0.05; \*\*p<0.01; F(1,396)

Üç boyutlu testlerde incelenen koşullar arasında anlamlı düzeyde etkileşimler olup olmadığının belirlenmesi için gerçekleştirilen üç yönlü ANOVA sonuçları bu bölümde yer almaktadır. Birinci boyut için OMY'nin ana etkisinin yetenek parametresi, ÇKM ve İKM'lere ait *a* ve *b* parametrelerinden elde edilen hata ve yanlılık değerlerinin tamamı için anlamlı düzeyde olduğu görülmüştür ( $p < .01$ ). Buna göre bu hata indislerinin tamamında ölçekleme desenindeki ortak madde yapısına göre hata ve yanlılık değerlerinde anlamlı düzeyde farklılaşmalar olduğu söylenebilir. Gözlenen bu farklılaşmaların tamamında karma format ortak maddelere sahip testlerin daha az hata ve yanlılık sağladığı ve bu sayede daha kesin ölçekleme işlemlerinin yapılabileceği görülmüştür. Elde edilen bu bulgu Cao (2008) tarafından elde edilen bulgularla örtüşmektedir.

VD koşulu için ana etkiler incelendiğinde yetenek parametreleri ( $p < .05$ ) ve İKM'lerin *a* parametreleri ( $p < .01$ ) ait hata değerleri üzerinde anlamlı düzeyde etkileri olduğu görülmüştür. Buna göre varyansın azaldığı senaryoda her iki parametre için de daha kesin ölçekleme sonuçları elde edilmiştir.

Son olarak, KY koşulu için ana etkiler incelendiğinde ise İKM'lerin *b* parametrelerine ait hata ve yanlılık değerleri ile ÇKM'lerin eşik parametrelerine ait yanlılık değerleri için anlamlı düzeyde etkisi olduğu bulunmuştur ( $p < .01$ ). Elde edilen bulgular EM kestirim yönteminin kullanılmasının hata ve yanlılık düzeylerini anlamlı düzeyde düşürdüğünü göstermektedir.

Birinci boyut için etkileşimler incelendiğinde OMY ve VD koşulları arasında yetenek parametresine ait yanlılık değerleri için anlamlı düzeyde etkileşim olduğu görülmüştür ( $p < .05$ ). Buna göre yetenek parametresi için elde edilen sistematik hata değerleri için iki kategorili ortak madde yapısında eşit varyansın daha az sistematik hataya yol açtığı, diğer taraftan karma format yapısında ise varyansın azaldığı durumlarda daha az sistematik hataya yol açtığı gözlenmiştir.

OMY ve KY koşullarının etkileşim düzeylerinin anlamlı olup olmadığını test etmek için gerçekleştirilen analizlere göre ÇKM'lerin eşik parametrelerine ait yanlılık değerleri için bu iki koşulun etkileşim içerisinde olduğu bulunmuştur ( $p < .05$ ). Bu etkileşim sonucunda iki kategorili ortak maddeli testlerde EM kestirim yöntemi daha az sistematik hataya yol açarken, karma format ortak maddelere sahip testler için MHRM kestirim yönteminin daha az yanlı sonuçlara sebep olduğu bulunmuştur.

VD ve KY koşulları için anlamlı bir etkileşim olup olmadığının incelenmesi için gerçekleştirilen analiz sonuçlarına göre yetenek parametrelerinden elde edilen yanlılık ( $p < .01$ ) ve İKM'lerin  $b$  parametreleri için elde edilen hata ( $p < .05$ ) ve yanlılık değerleri ( $p < .05$ ) için bu iki koşul arasında anlamlı düzeyde etkileşim olduğu bulunmuştur. Yetenek parametrelerine ait yanlılık değerlerinin eşit varyans koşulunda EM kestirim yönteminin daha az yanlılığa sebep olduğu, azalan varyans koşulunda ise MHRM kestirim yönteminin daha etkili sonuçlar verdiği görülmüştür.

Son olarak, etkisi incelenen üç koşul arasındaki etkileşimler incelendiğinde İKM'lerin  $a$  parametrelerine ait hata değerleri için anlamlı düzeyde üçerli etkileşim olduğu görülmüştür ( $p < .05$ ). Buna göre en az hata değerleri ortak maddelerin kullanıldığı, karma format testlerde ve EM kestirim yönteminin kullanımıyla elde edilmiştir. Ayrıca, varyansın eşit olduğu durumlarda, kestirim yöntemlerinden hangisinin daha iyi sonuçlar vereceği ortak madde yapısına göre değişmektedir.

İkinci boyutta yer alan maddeler için ana etkiler ve etkileşimler incelendiğinde, OMY koşulu için yetenek parametreleri ve İKM'lerin  $a$  ve  $b$  parametreleri için hata ve yanlılık değerleri, ÇKM'lerin  $a$  parametresi için hata değerleri ve ÇKM'lerin eşik parametreleri için yanlılık değerlerinde anlamlı düzeyde ana etkinin olduğu görülmektedir. Bu bulgu, bu değerlerin OMY'ye göre farklılaştığını göstermektedir. Buna göre, karma format testler daha etkili eşitleme sonuçları vermektedir. VD koşulu için herhangi bir anlamlı ana etki bulunmamıştır. KY koşulu için ise yetenek parametrelerine ait yanlılık değerleri, İKM'lerin  $a$  parametrelerine ait hata değerleri ve  $b$  parametrelerine ait hata ve yanlılık değerleri için anlamlı düzeyde ana etkinin olduğu bulunmuştur. Bu bulguya göre, EM kestirim yöntemi belirtilen parametreler için için anlamlı düzeyde daha kesin ölçekleme sonuçları elde edilmesini sağlamaktadır.

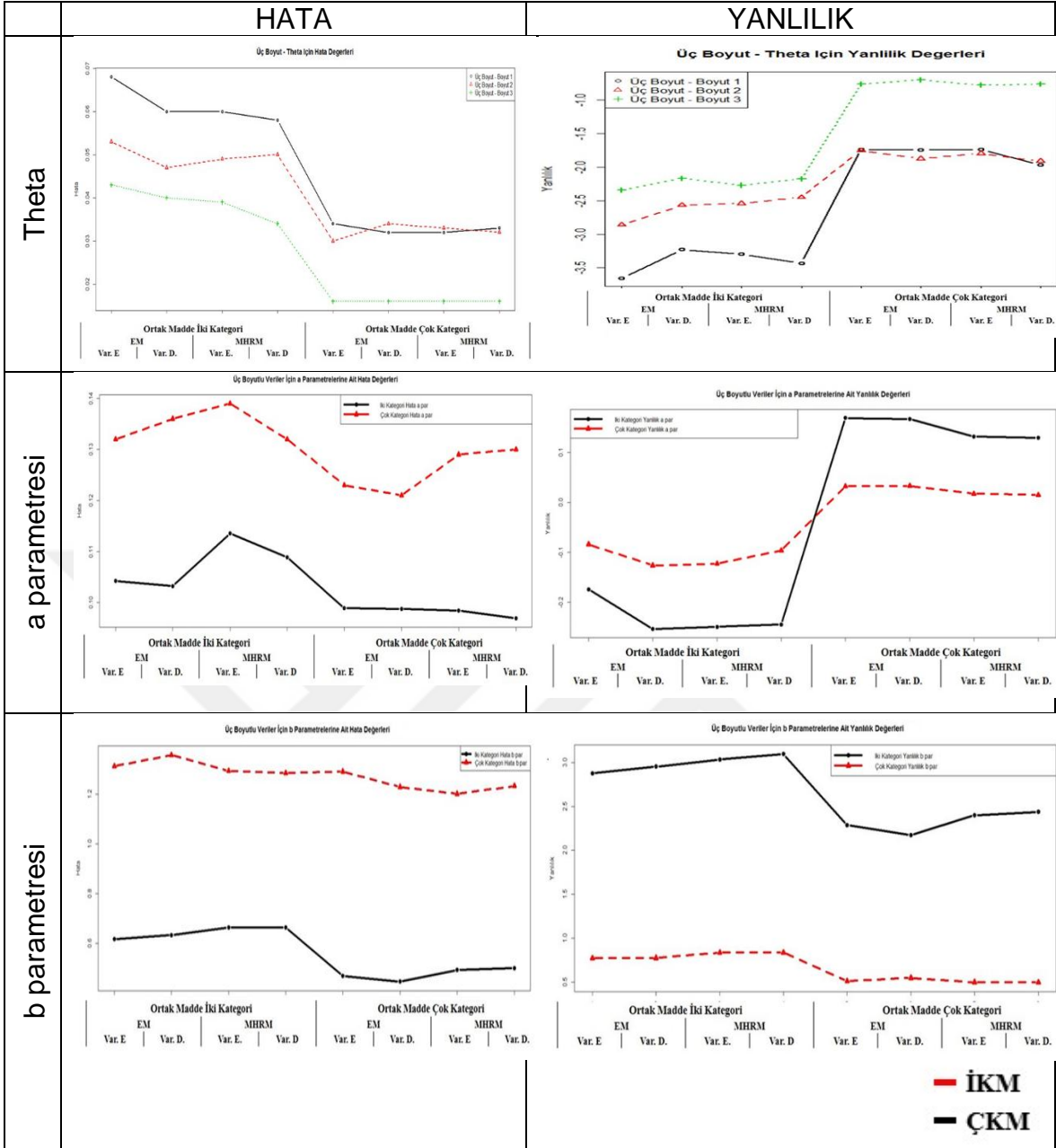
İkinci boyutta yer alan ikili etkileşimler incelendiğinde OMY ve VD koşulları için yetenek parametrelerinden elde edilen yanlılık değerlerinde anlamlı düzeyde etkileşim olduğu görülmüştür ( $p < .01$ ). Buna göre ortak madde iki kategorili olduğu senaryoda eşit varyanslı testler daha iyi sonuçlar verirken, karma format ortak madde setleri için azalan varyansa sahip testlerin ölçeklenmesi daha az yanlılığa yol açmaktadır. Ayrıca OMY ve KY koşullarının İKM'lerin  $a$  parametrelerine ait hata değerleri için etkileşim içerisinde oldukları bulunmuştur. Buna göre, iki kategorili ortak maddelere sahip testlerde EM kestirim yöntemi daha etkili olurken, karma

format ortak madde yapılarında MHRM kestirim yönteminin daha etkili olduğu sonucuna varılmıştır. Diğer taraftan KY ve VD koşulları arasındaki ikişerli etkileşimler ve OMY, VD ve KY koşulları arasındaki üçerli etkileşimler incelendiğinde parametrelere ait hata indisleri için anlamlı etkileşimler olmadığı bulunmuştur.

Son olarak üçüncü boyutta yer alan parametreler için bakıldığında OMY koşulu için yetenek parametreleri ve madde parametrelerinin tamamı için anlamlı düzeyde ana etkinin olduğu görülmektedir. Bu bulgu göstermiştir ki OMY koşuluna göre ortak madde yapısına göre karma format ortak maddelerin yer aldığı testlerde hata ve yanlışlık değerleri anlamlı düzeyde daha düşüktür ( $p < .01$ ). VD koşulu için ana etkiler incelendiğinde gözlenen tek anlamlı etkinin yetenek parametreleri için elde edilen hata değerleri için olduğu bulunmuştur ( $p < .01$ ). Son olarak, KY koşulu için yetenek parametrelerine ait hata ( $p < .05$ ), İKM'lerin *a* parametrelerine ait yanlışlık değerleri ( $p < .05$ ) ve *b* parametrelerine ait hata ( $p < .01$ ) ve yanlışlık ( $p < .01$ ) değerlerinde anlamlı düzeyde ana etki olduğu bulunmuştur. Güçlük parametresine ait gözlenen farklılaşmalar MHRM kestirim yönteminin daha etkili olduğunu göstermektedir.

Üçüncü boyut için elde edilen parametrelere ait hata ve yanlışlık değerlerinde koşullara göre etkileşim olup olmadığının belirlenmesi için gerçekleştirilen analiz sonuçlarına göre, VD koşulunun OMY ve KY koşulları ile arasındaki ikişerli etkileşimlerin anlamlı düzeylerde olmadığı görülmüştür ( $p > .05$ ). Diğer taraftan OMY ve KY koşulları arasında ÇKM'lerin eşik parametrelerinden elde edilen yanlışlık değerlerine göre anlamlı düzeyde etkileşim olduğu görülmüştür. Buna göre iki kategorili ortak madde setlerinin yer aldığı testler için EM kestirim yöntemi daha etkili iken, karma format ortak maddeli testler için MHRM kestirim yöntemi daha etkilidir. Ayrıca, OMY, VD ve KY koşulları arasındaki üçerli etkileşimlere bakıldığında elde edilen tek anlamlı sonucun eşik parametresine ait hata değerleri için olduğu bulunmuştur. Varyansın eşit olduğu senaryoda iki kategorili ortak madde seti için EM kestirim yöntemi daha etkili olurken, karma format ortak madde seti için MHRM kestirim yöntemi daha etkili olmuştur. Varyansın daraldığı senaryoda ise bu durumun tam tersi bir sonuç çıkmıştır.

Gözlenen değerlerin koşullar göre nasıl farklılaştığı çizgi grafikleri kullanılarak Şekil 4.3'de yer verilmiştir.



**Şekil 4.3. Üç Boyutlu Veriler için Elde Edilen Hata ve Yanlılık Değerlerinin Her Bir Parametre için Koşullara Göre Karşılaştırması**

## 5. SONUÇ ve ÖNERİLER

Araştırmanın bu bölümünde gerçekleştirilen analizlerden çıkan sonuçlara ve bu sonuçlardan yola çıkarak geliştirilen önerilere yer verilmiştir.

### 5.1. Sonuçlar

Gerçekleştirilen bu çalışmada DOGOM deseni kullanılarak çok boyutlu karma format testlerin ölçeklenmesinde ortak madde yapısının, kestirim yönteminin ve yetenek daralmasının eşitlemenin niteliği üzerindeki etkisine bakılmıştır. Ölçeklemenin niteliğini belirleyebilmek ve bu yolla farklı koşulların etkilerini karşılaştırabilmek için eşitleme hatası ve yanlılık değerleri kullanılmıştır. Bu amaç doğrultusunda veri setleri yöntem kısmında belirtildiği şekliyle türetilmiş ve önce veri setleri çok boyutlu MTK bağlamında kalibre edilmiş, daha sonrasında ise ölçekleme işlemi gerçekleştirilmiştir. Gerçekleştirilen her bir işlem 50 defa tekrarlanmıştır. Bu yolla ulaşılabilecek sonuçların geçerliliğinin artırılması amaçlanmıştır.

Bu doğrultuda gerçekleştirilen analizler sonucunda elde edilen hata ve yanlılık değerlerinin incelenmesi sonucunda elde edilen sonuçlar aşağıda yer almaktadır.

#### ***Ortak madde yapısının etkisinin incelenmesi***

- 1) Ölçekleme işlemi sonucunda *yetenek* parametresine ait hata değerleri incelendiğinde, bir ve üç boyutlu testlerde hata değerleri incelendiğinde bu değerlerin ortak madde setinin karma format olduğu veriler için anlamlı seviyede daha düşük olduğu görülmüştür. İki boyutlu testlerde ise birinci boyut için hata değerlerinde anlamlı bir farklılaşma gözlenmemişken, ikinci boyut için ortak madde seti iki kategorili olan testler lehine bir farklılaşma olduğu görülmüştür. Gözlenen bu farklılaşmalar yanlılık değerleri için de geçerlidir.
- 2) İKM'lerin *a* parametrelerine ait hataların testlerin boyut sayısından bağımsız olarak iki kategorili ortak maddelere sahip testler için daha yüksek olduğu görülmüştür. Bu farklılaşmalar yanlılık değerleri için de geçerlidir. Ayrıca, İki boyutlu testlerin her iki boyutu ve üç boyutlu verilerin birinci boyutu için *a* parametrelerine ait hata değerlerinin diğer koşullar ile üçerli etkileşim içerisinde. Son olarak üç boyutlu verilerin ikinci boyutunda hata değerleri açısından KY koşulu ile etkileşim içerisinde.



- 3) İKM'lerin  $b$  parametrelerine ait hata ve yanlılık değerlerinin ortak madde yapısına göre gösterdiği değişimler incelendiğinde hata ve yanlılık değerlerinin boyut sayısından bağımsız olarak testlerin tamamında karma format ortak madde seti için daha düşük olduğu bulunmuştur. Ayrıca, tek boyutlu veriler için hata ve yanlılık açısından OMY koşulunun diğer koşullarla üçerli etkileşim içerisinde olduğu görülmüştür.
- 4) ÇKM'lerin  $a$  parametrelerine ait hata ve yanlılık değerlerinin bir ve iki boyutlu testlerde ve üç boyutlu testlerin ilk iki boyutunda ortak madde seti karma format olan maddeler için anlamlı düzeyde daha düşük olduğu bulunmuştur. Üç boyutlu testlerin üçüncü boyutu için ise ortak madde seti İKM'lerden oluşan testlerin hata ve yanlılık değerlerinin daha düşük olduğu görülmüştür. Ayrıca tek boyutlu veriler için hata ve yanlılık değerleri açısından OMY koşulunun KY koşulu ile etkileşim içerisinde olduğu bulunmuştur.
- 5) ÇKM'lerin eşik parametresi için incelendiğinde OMY'nin bir ve iki boyutlu testlerde hata ve yanlılık değerleri üzerinde etkisinin olduğu görülmüştür. Buna göre, hata değerlerinin bir ve iki boyutlu testlerde karma format ortak madde içeren testler için daha yüksek olduğu görülmüştür. Yanlılık değerleri için bakıldığında ise tek boyutlu testler için karma format ortak maddelerin daha yüksek yanlılığa sebep olduğu, iki boyutlu testlerde ise iki kategorili maddelerden oluşan testlerin daha fazla yanlılığa sebep olduğu bulunmuştur. Üç boyutlu testler için bakıldığında ise her üç boyut için yanlılık değerlerinin ortak maddeleri karma format olan testler için daha düşük olduğu görülmüştür. Son olarak üç boyutlu testlerde OMY ve KY koşulları arasında yanlılık değerleri açısından etkileşim bulunmuştur.

Araştırma sonucunda elde edilen bulgular göstermektedir ki ortak madde yapısı ölçekleme işlemi sonucunda ortaya çıkan hata ve yanlılık miktarını önemli ölçüde etkilemektedir. Buna göre karma format testlerde ortak madde setinin sadece İKM'lerden oluşması ölçekleme hatasını bazı istisnalar haricinde arttırmaktadır. Elde edilen bu bulgu, araştırmada kullanılan testlerin boyut sayısından bağımsız olarak tutarlı bir şekilde kendisini göstermektedir. Kolen ve Brennan (2004)'ın da belirttiği üzere ortak madde setinin içerik ve istatistiksel özellikler açısından toplam testin "mini versiyonu" olması gerekmektedir. Pratikte ÇKM'lerin ortak madde setine konulması az rastlanır bir durum olmasına rağmen bu çalışmada elde edilen

bulgular göstermektedir ki ortak madde seti toplam teste ne ölçüde benzer ise; ölçekleme sonuçları o ölçüde olumlu yönde etkilenmektedir.

### ***Varyans daralmasının etkisinin incelenmesi***

- 1) Varyans daralmasının yetenek parametresine ait hata ve yanlışlık değerlerini nasıl etkilediğine bakıldığında iki boyutlu testlerde ikinci boyut için yanlışlık değerlerini etkilediği, üç boyutlu verilerde ise birinci ve üçüncü boyut için hata değerlerini etkilediği görülmüştür. Ayrıca iki ve üç boyutlu testlerde yanlışlık değerleri açısından varyans daralmasının diğer koşullar ile etkileşim içerisinde oldukları bulunmuştur. Belirtilen farklılıkların tamamı için eşit varyans durumunun olumsuz etkisinden söz etmek mümkündür.
- 2) İKM'lere ait  $a$  parametreleri için varyans daralmasının hata ve yanlışlık değerlerinin nasıl etkilediği incelendiğinde gözlenen tek farklılaşmanın üç boyutlu testlerde birinci boyutta yer alan maddelerden elde edilen hata değerleri için olduğu görülmektedir. Burada eşit varyansa sahip testlerle gerçekleştirilen ölçeklemelerin daha fazla hataya yol açtığı görülmüştür. Ayrıca, iki boyutlu veriler için varyans daralması koşulunun diğer koşullar ile etkileşim içerisinde olabileceğine dair bulgular edinilmiştir.
- 3) İKM'lere ait  $b$  parametreleri için hesaplanan hata ve yanlışlık değerlerinin varyans daralması koşuluna bağlı olarak anlamlı düzeyde farklılaşmadığı görülmüştür. Gözlenen bu bulgu ölçekleme işleminde kullanılan verilerin boyut sayısından bağımsızdır. Etkileşimler incelendiğinde ise üç boyutlu testlerde birinci boyut için hata ve yanlışlık değerlerinin kestirim yöntemi ile etkileşim içerisinde olduğu bulunmuştur.
- 4) ÇKM'lerin  $a$  parametresi için, varyans daralması koşulu için anlamlı bir farklılaşma gözlenmemiştir. Benzer şekilde varyans daralması diğer koşullarla herhangi bir etkileşim içerisine girmemiştir.
- 5) ÇKM'lerin eşik parametreleri için eşik parametrelere ait hata ve yanlışlık değerlerinin farklılaşmadığı ve diğer koşullar ile etkileşim içerisinde olmadığı görülmüştür.

### ***Kestirim yönteminin etkisinin incelenmesi***

1. Yetenek parametreleri açısından bakıldığında, tek boyutlu testler için hata ve yanlışlık değerlerinin kestirim yöntemine göre farklılaştığı görülmüştür. Buna

göre EM kestirim yöntemi daha iyi sonuçlar vermektedir. Ayrıca iki boyutlu testlerde ikinci boyutta ve üç boyutlu testlerde ikinci ve üçüncü boyutta yanlılık değerleri kestirim yöntemine göre farklılaştığı görülmüştür. İki ve üç boyutlu testlerde MHRM kestirim yönteminin daha az sistematik hata doğurduğu söylenebilir.

2. İKM'lerin  $a$  parametrelerinden elde edilen yanlılık değerleri için kestirim yönteminin iki boyutlu testlerin ikinci boyutunda ve üç boyutlu testlerin ikinci boyutunda yanlılık değerlerini etkilediği görülmüştür. Ayrıca üç boyutlu testlerde, üçüncü boyut için hata değerlerini etkilediği bulunmuştur. MHRM kestirim yöntemi gözlenen farklılıklarda daha iyi sonuçlar verdiğini söylemek mümkündür.
3. İKM'lerin  $b$  parametreleri için gözlenen farklılıklar incelendiğinde genel olarak EM kestirim yönteminin daha iyi sonuçlar verdiğini görülmüştür.
4. ÇKM'lerin  $a$  parametresi için anlamlı farklılıklar gözlemlenmemiştir.
5. ÇKM'lerin eşik parametrelerine için bakıldığında kestirim yöntemine göre tek boyutlu veriler ile iki boyutlu verilerin birinci boyutunda hata ve yanlılık değerlerinin farklılaştığı görülmüştür. Gözlenen bu farklılaşma incelendiğinde EM kestirim yönteminin daha iyi performans gösterdiği bulunmuştur. Benzer şekilde iki boyutlu verilerin ikinci boyutu için de EM yönteminin daha az yanlılık gösterdiği görülmüştür. Üç boyutlu veriler için ise herhangi bir farklılaşma gözlenmemiştir. Son olarak, yanlılık değerleri için kestirim yöntemi ortak madde yapısı koşulu ile etkileşim içerisindedir.

Görüldüğü üzere boyut sayısı arttıkça bazı durumlarda MHRM kestirim yönteminin ölçekleme sonuçları üzerindeki olumsuz etkisi azalmaktadır. Cai, (2008)'in de belirttiği üzere MHRM kestirim yöntemi yüksek boyutlu verilerin kestirimi için daha kullanışlıdır. Bu bulgular MHRM kestirim yönteminin bu özelliği ile açıklanabilir.

### ***Genel Değerlendirme***

MHRM ve EM döngüleri ile gerçekleştirilen parametre kestirimlerinin ölçekleme sonuçları üzerinde etkisinin birçok durumda olmaması, bu çalışmada tercih edilen en yüksek boyut sayısı ile açıklanabilir. Cai (2008) ve Cai (2010a) tarafından gerçekleştirilen çalışmalarda 5 ve daha üstündeki veri setleri için MHRM kestirimlerinin daha iyi sonuçlar vereceği, daha düşük boyut sayısına sahip veri setleri için ise bu iki kestirim yöntemlerinin oldukça benzer sonuçlar verdiğini

belirtilmiştir. Diğer taraftan, Cai (2010a) tarafından da belirtildiği gibi EM döngüleri kullanıldığında boyut sayısı doğrusal olarak artırıldığında, kestirimler için gerekli quadrat nokta sayısı çarpımsal olarak artar. MHRM kestirimleri için böyle bir gereksinim yoktur ve çok boyutlu verilerin parametre kestirimlerinde kullanılması daha avantajlı hale gelmektedir. Sonuç olarak, düşük boyut sayısına sahip verilerde bu iki kestirim yönteminin benzer sonuçlar verdiği söylenebilir.

Genel olarak yanlılık değerlerinin yüksek olduğu söylenebilir. Yanlılık değerlerinin yüksek olması ölçeklemede kullanılan test formları arasındaki  $\theta$  ortalama farkının 1 olması ve çok boyutluluk ile açıklanabilir. Özellikle ölçeklenen test formlarının boyutluluğunun artması ile birlikte yanlılık değerlerinin de arttığı kabul edilmektedir (Kim ve Lee, 2006). Kolen ve Brennan (2004, s. 286) ise yetenek ortalamaları arasındaki farkın 0.3'ü geçtiği durumlarda eşitleme sonuçlarının hata ve yanlılıklar içerebileceği ifade edilmiştir. Bu durumun olası bir sebebi gruplar arasındaki yetenek farkının artmasının formlar arası farklılıkları belirlemede daha fazla hataya yol açıyor olması olabilir.

Nitekim yetenek farkları arasındaki farklılığın yüksek olduğu durumlarda yanlılığın yüksek çıkabileceği farklı birçok çalışma tarafından doğrulanmıştır (Brennan ve Kolen, 2008; Kirkpatrick, 2005; Wang, Lee, Wu, Huang, Hu ve Harris, 2009). Diğer taraftan, Cao (2008) ve Kirkpatrick (2005) yetenek farklılıklarının ölçme sonuçları üzerindeki etkisinin yalnızca MTK bağlamında geçerli olduğunu belirtmişlerdir. Ayrıca, Hagge (2010) tarafından gerçekleştirilen çalışmada gruplar arasındaki yetenek farkının yüksek olduğu durumlarda yanlılığın arttığı fakat İKM ya da ÇKM'ler arasındaki ilişkinin çok yüksek olduğu durumlarda yanlılık değerinin göreceli olarak daha düşük olabildiği ifade edilmiştir.

Gerçekleştirilen ölçekleme işlemi sonuçlarına göre yanlılık değerlerinin yüksek çıkmasının olası sebeplerinden bir diğeri ise boyutlar arasındaki ilişki düzeylerinin sıfır (0) olması olarak gösterilebilir. Bolt, (1999) ve Camilli ve arkadaşları (1995) tarafından gerçekleştirilen çalışmalarda boyutlar arasındaki ilişkilerin ancak 0,7 düzeyine çıktığı durumlarda hata ve yanlılık düzeylerinin kabul edilebilir seviyelere düştüğü belirtilmiştir.

Kim, Walker ve McHale (2008) ise karma format ortak madde setinin etkili olabilmesi için ÇKM'lerde trend puanlama yöntemi kullanılmış olmasının gerekli olduğunu ifade

etmişlerdir. Hali hazırdaki bu çalışmada ise mevcut test edilen koşullar dikkate alınarak bir yorum getirmek gerekirse karma format ölçekleme yönteminin ölçekleme sonuçları üzerinde oldukça etkili olduğu söylenebilir.

## **5.2. Öneriler**

Bu çalışmada farklı koşulların karma format testlerle gerçekleştirilen ölçekleme üzerindeki etkileri ve eşitlemenin niteliği açısından etkileşim içerisinde olup olmadıkları incelenmiştir. Elde edilen bulguların bu koşulların etkileri üzerine çıkarımlar sağlamaya yardımcı olacağı düşünülmektedir. Bu çıkarımlar ışığında bundan sonra gerçekleştirilecek çalışmalar ve uygulayıcılar için getirilebilecek öneriler şu şekildedir.

## **5.3. Uygulayıcılara Yönelik Öneriler**

- 1) Karma format testlerle gerçekleştirilen bu çalışmada ortak madde setinin de karma format maddelerden oluşmasının gerçekleştirilen ölçekleme işlemini olumlu yönde etkilediği görülmüştür. Ortak madde seti ile toplam testin mümkün olduğu kadar gerek istatistiksel gerekse içerik olarak benzerliği ölçeklemenin sistematik ve tesadüfi hata miktarını azaltacağı açıktır. Bu sebeple test geliştiricilerinin maliyet ve ÇKM sayısının uygulamalarda az olması gibi pratikte gerçekleştirilmesi bir takım zorlukları beraberinde getirirse de denk olmayan gruplarla geliştirecekleri ölçekleme işleminde ortak madde setini oluştururken bu durumu dikkate almaları ve testi temsil edecek şekilde karma format maddelere yer vermeleri önerilmektedir.
- 2) Bu araştırmanın bir kısıtlılığı olarak veri türetme aşamasında kullanılan iki kademeli modelin gerçek test uygulamalarında görülen boyutluluk yapılarını tam olarak yansıtmamış olma durumu gösterilebilir. Yöntem kısmında da belirtildiği üzere bu çalışmada yanıt matrisleri türetilirken içeriğe dayalı boyutluluk için boyutlar arası ilişki olmayacak şekilde, madde formatına göre ise 0.75 düzeyinde ilişki olacak şekilde yanıt matrisleri türetilmiştir. Bu tercih iki kademeli modellerin gerekliliğidir. Gerçek test uygulamalarında ise iki kademeli modelleri mükemmel bir şekilde yansıtabilecek verilere karşılaşılmaması çoğunlukla mümkün değildir. Uzmanların elde edilen bulguları yorumlarken bu durumu dikkate almaları önerilmektedir.

- 3) Ölçeklemede kullanılan grupların yetenek farkları etkisi incelenen bir durum olmasa da gruplar arasındaki bu fark 1 olarak alınmıştır. Bu sebeple de ölçekleme sonucunda elde edilen hata ve yanlılık değerlerinin yüksek olmasının bir sebebinin bu olduğu düşünülmektedir. Buradan hareketle ölçekleme işleminin sık aralıklarla, örneğin ardışık sınıflarda gerçekleştirilmesinin gruplar arası yetenek farkının az olmasını ve ölçekleme işleminin daha az sistematik hata ile elde edilmesini sağlayacağı düşünülmektedir. Bu sebeple uygulayıcıların ölçekleme işlemi kullanılacak uygulamaları daha az zaman aralıkları ile gerçekleştirmeleri önerilebilir.
- 4) Toplam test içerisindeki ÇKM'lerin İKM'lere oranı ortak madde setinde var olan oranla birebir aynı değildir. Bu durum ortak madde setinin testi temsil gücünü düşürmektedir. Ortak maddelerin oranlarındaki bu farklılaşmanın ölçekleme sonucunu nasıl etkileyeceği bilinmemektedir. Uzmanların elde edilen sonuçları kendi uygulamalarına aktarırlarken bu durumu dikkate almaları önerilmektedir.

#### **5.4. Araştırmacılara Yönelik Öneriler**

- 1) Gerçekleştirilen bu çalışmanın en önemli kısıtlılığı test edilen koşulların yeterli değişkenliğe sahip olmamasıdır. Örneğin yetenek dağılımlarındaki yetenek daralması sadece %40 olarak alınmıştır. Fakat gerçek test uygulamalarında %70 ve üzerinde yetenek daralmalarının görülmesi olasıdır. Ayrıca, türetilen testler en fazla üç boyutlu yapı gösterecek şekilde tasarlanmıştır. Fakat gerçek test durumlarında daha yüksek sayıda boyuta sahip test uygulamaları olabilmektedir. Bu sebeple gerçekleştirilen bu çalışmanın koşulların değişkenliği artırılarak yinelenmesi önerilmektedir. Ayrıca MHRM kestirim yöntemlerinin özellikle yüksek boyutlu testlerde analizinde önemli avantajları olduğu düşünüldüğünde kestirim yönteminin ölçekleme sonuçları üzerindeki etkilerinin daha iyi görülebilmesi için daha fazla boyuttan oluşan yapıların incelenmesi önerilmektedir.
- 2) Gerçekleştirilen bu çalışmada test formlarının yetenek ortalamaları arasındaki fark 1 olarak alınmıştır. Bu değer gerçek test uygulamalarında gözlenebilecek en üst değerlerden birisidir. Ayrıca, gerçek test uygulamalarında test geliştiricileri test formları arasındaki güçlük farkını mümkün olduğu kadar düşük seviyede tutmaya gayret ettikleri bilinmektedir. Bu sebeple bundan sonra

gerçekleştirilecek olan çalışmalarda yetenek farkının daha düşük tutulması önerilmektedir.

- 3) Gerçekleştirilen bu çalışmada gerçek verilerden yararlanılmamıştır. Bu sebeple bu çalışmadan elde edilen sonuçlarından gerçek test durumlarına genelleme yapılırken dikkatli olunmalıdır. Bundan sonraki çalışmalarda gerçek verilerle test etkisi test edilen durumların benzer sonuçlar doğurup doğurmayacağı gerçek veriler kullanılarak test edilmesi önerilmektedir.
- 4) Bu çalışmada ortak madde seti oluşturulurken içerik ve madde formatına dayalı temsiliyet her veri seti için dikkate alınmıştır. Yani ortak veri setinde her üç boyutu temsil edecek şekilde İKM ve ÇKM'lere yer verilmiştir. Ortak madde setinden elde edilen bilginin önemi düşünüldüğünde bu maddelerin önemli olduğu gerçeği daha da anlaşılır olacaktır. Bundan sonra gerçekleştirilecek çalışmalarda içerik ve formata dayalı temsil edilme ihlal edilerek bu durumun etkileri incelenebilir.
- 5) Gerçekleştirilen bu çalışmada iki kategorili tepki modeli ve ATM kullanılmıştır. Karma format bir testin daha farklı madde türlerinden oluşabileceği düşünülürse, farklı tepki modelleri kullanılarak bu çalışma yinelenabilir. Ayrıca, karma format testlerin ikiden fazla madde formatından oluşması muhtemeldir. Bu çalışmada ise sadece farklı iki madde formatı kullanılmıştır. Bu sebeple ikiden fazla madde formatının yer aldığı karma format testlerle bu çalışma yinelenmesi önerilmektedir.
- 6) Gerçekleştirilen bu çalışmada boyutluluk kaynağı olarak içerik ve maddelerin formatı dikkate alınarak veriler türetilmiştir. Ayrıca içeriğe dayalı olarak testin basit boyutluluk yapısında olduğu varsayılmıştır. Gerçek test durumlarında ise testlerin boyutluluğunu etkileyen birden fazla faktör devreye girmektedir. Ayrıca testlerin basit yapıda olması özellikle sosyal bilimlerde güçlü olmayan bir varsayımdır. Bu durumun elde edilen bulguların genellenebilir olmasını olumsuz etkileyebileceği düşünülmektedir. Bu sebeple, bundan sonra gerçekleştirilecek olan çalışmalarda basit olmayan faktör yapıları ve boyutluluğa etki eden daha fazla etmenin dikkate alındığı boyutluluk yapıları ile bu çalışma yinelenabilir.
- 7) Gerçekleştirilen bu çalışmada DOGOM deseni kullanılmıştır. Ölçkleme desenleri farklı şartlarda farklı beklentileri karşılamak için gerçekleştirildiği düşünüldüğünde bundan sonra gerçekleştirilecek çalışmalarda farklı desenlerde test edilen koşulların nasıl bir etkide bulunacakları test geliştiricileri için önemli

bulgular sağlayacağı düşünülmektedir. Bu sebeple bundan sonra gerçekleştirilecek çalışmalarda farklı desenlerin kullanılması önerilmektedir.

- 8) Bu çalışmada madde parametrelerinin kestirilmesi MTK temelli modellerin kullanılması ile gerçekleştirilmiştir. MTK'nın model temelli bir kuram olması kestirimlerde modele bağlı bir yanlılık doğurabilir. Yapılabilecek tüm yorumlar da yine ancak bu model bağlamında geçerli olabilir. Bu sebeple bu çalışma klasik test kuramına dayalı puan hesaplamaları kullanılarak daha farklı bir anlayışla yenilenebilir.
- 9) Hali hazırdaki bu çalışmada ölçekleme işleminin gerçekleştirilebilmesi için Haebara yöntemi kullanılmıştır. Çok boyutlu karma format testlerin ölçeklenmesi ile ilgili çalışmalarda geleneksel eşitleme yöntemlerinin kullanıldığı çok az çalışma mevcuttur. Bundan sonra gerçekleştirilecek çalışmalarda farklı eşitleme yöntemlerin karşılaştırılması önerilmektedir.
- 10) Son olarak, bundan sonra gerçekleştirilecek çalışmalarda farklı istatistik programları (örneğin Multilog ya da Parscale) kullanılabilir.



## KAYNAKÇA

- Altun, A. (2013). *Dikey ölçeklemede madde tepki kuramına dayalı farklı kalibrasyon ve yetenek kestirim yöntemlerinin karşılaştırılması*. Doktora Tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü. Ankara.
- Baker, Frank B. (1992). *Item response theory: parameter estimation techniques*. New York: Marcel Dekker, Inc.
- Barnard, J. J. (1996). *In search for equity in educational measurement: traditional versus modern equating methods*. Paper presented at ASEESA's national conference at the HRSC Conference Centre, Pretoria, South Africa.
- Bastari, B. (2000). *Linking multiple-choice and constructed-response items to a common proficiency scale*. Unpublished doctorate dissertation, University of Massachusetts. Boston.
- Bennett, R. E., Rock, D. A., & Wang, M. (1991). Equivalence of free-response and multiple-choice items. *Journal of Educational Measurement*, 28, 77-92.
- Birnbaum A (1968). Some latent trait models and their use in inferring an examinee's ability. In FM Lord, MR Novick (eds.), *Statistical Theories of Mental Test Scores*, 397-479. Addison-Wesley, Reading, MA.
- Bolt, D. M. (1999). Evaluating the effects of multidimensionality on IRT true-score equating. *Applied Measurement in Education*, 12, 383-407.
- Cai, L. (2008). *A Metropolis-Hastings Robbins-Monro algorithm for maximum likelihood nonlinear latent structure analysis with a comprehensive measurement model*. Unpublished doctorate dissertation, Department of Psychology, University of North Carolina at Chapel Hill.
- Cai, L. (2010a). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika*, 75, 33-57.
- Cai, L. (2010b). Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, 35, 307-335.
- Camilli, G. (1999). Measurement error, multidimensionality, and scale shrinkage: a reply to Yen and Burket. *Journal of Educational Measurement*. 36(1), 73-78.
- Cao, L. (2008). *Mixed format test equating: effects of test dimensionality and common-item sets*. Unpublished doctorate dissertation, University of Maryland.
- Cao, Y., Yin, P., & Gao, X. (2007, April). *Comparison of IRT and classical equating methods for tests consisting of polytomously-scored items*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Cook, L. L., & Petersen N. S. (1987). Problems related to the use of conventional and item response theory equating methods in less than optimal circumstances. *Applied Psychological Measurement*, 11, 225-244.

- Çetin, E. (2009). *Dikey ölçeklemede klasik test ve madde tepki kuramına dayalı yöntemlerin karşılaştırılması*. Doktora Tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü. Ankara.
- Dodd, B. G. (1984). *Attitude scaling: A comparison of the graded response and partial credit latent trait models*. Unpublished doctorate dissertation, University of Texas at Austin. Texas.
- Gübeş, N.Ö. (2014). *Testlerin boyutluluğunun, ortak madde formatının, yetenek dağılımının ve ölçek dönüştürme yöntemlerinin karma testlerin eşitlenmesine etkisi*. Doktora tezi. Eğitim Bilimleri Enstitüsü. Eğitimde Ölçme ve Değerlendirme Bilim Dalı. Hacettepe Üniversitesi. Ankara.
- Haebara T (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, 22(3), 144–149.
- Haertel, E. (1991). *Report on TRP analyses of issues concerning within-age versus crossage scales for the national assessment of educational progress*. Washington, DC: National Center for Education Statistics.
- Hagge, S. L. (2010). The impact of equating method and format representation of common items on the adequacy of mixed-format test equating using nonequivalent groups. Doktora Tezi, University of Iowa, Iowa City, Iowa.
- Hagge, S. L. Lee, W., He, Y., Kolen, M. J., ve Wang, W. (2010). *Equating mixed-format tests using dichotomous anchor items*. Paper presented in the structured poster session Mixed-Format Tests: Addressing Test Design and Psychometric Issues in Tests with Multiple-Choice and Constructed-Response Items at the Annual Meeting of the American Educational Research Association, Denver, CO.
- Hardy, M.A., Young, M. J., Yi, Q., Sudweeks, R. R., & Bahr, D. L. (2011). *Investigating content and construct representation of a common-item design when creating a vertically scaled test*. Paper presented at the annual meeting of the National Council on Measurement in Education New Orleans, LA.
- Harris, D. J. (1991). A comparison of Angoff's Design I and Design II for vertical equating using traditional and IRT methodology. *Journal of Educational Measurement*, 28(3), 221-235.
- Houts, CR, & Cai, L. (2013). *flexMIRT user's manual version 2.0: flexible multilevel multidimensional item analysis and test scoring*. Chapel Hill, NC: Vector Psychometric Group.
- Jonathan P. Weeks (2010). plink: An R package for linking mixed-format tests using IRT-based methods. *Journal of Statistical Software*, 35(12), 1-33.
- Kamata, A., & Tate, R. (2005). The performance of a method for the long-term equating of mixed-format assessment. *Journal of Educational Measurement*. 42, 193-213.
- Karasar, N. (2000). *Bilimsel araştırma yöntemi*. Ankara: Nobel Yayın Dağıtım.
- Kelecioğlu, H. (1994). *Öğrenci seçme sınavı puanlarının eşitlenmesi üzerine bir çalışma*. Doktora tezi. Sosyal Bilimler Enstitüsü Eğitimde Ölçme ve Değerlendirme Bilim Dalı. Hacettepe Üniversitesi. Ankara.

- Kim S ve Lee W. (2006). An extension of four IRT linking methods for mixed-format tests. *Journal of Educational Measurement*, 43(1), 53-76.
- Kim, J.S. ve Hanson, B.A. (2002). Test equating under the multiple-choice model. *Applied Psychological Measurement*, 26, 255-270.
- Kim, S. ve Lee, W. (2004). *IRT scale linking methods for mixed-format tests* (ACT Research Report 2004-5). Iowa City, IA: ACT, Inc.
- Kim, S., Walker, M. E., & McHale, F. (2008). *Equating of mixed-format tests in large-scale assessments. Technical Report (RR-08-26)*. Princeton, NJ: Educational Testing Service.
- Kim, S.H., ve Cohen, A.S. (2002). A comparison of linking and concurrent calibration under the graded response model. *Applied Psychological Measurement*, 26 (1), 25–41.
- Kirkpatrick, R. K. (2005). *The effects of item format in common item equating. Unpublished doctorate dissertation*. University of Iowa. Iowa.
- Klein, L. W., & Jarjoura, D. (1985). The importance of content representation for common-item equating with nonrandom groups. *Journal of Educational Measurement*, 22, 197-206.
- Kolen, M. J., ve Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices*. NY: Springer.
- Koretz, D.M., & Hamilton, L.S. (2006). Testing for accountability in K-12. In R.L. Brennan (Ed.), *Educational measurement* (4th ed. 531-578). Westport, CT: American Council on Education and Praeger Publishers.
- Lee, W., He, Y., Hagge, S. L., Wang, W., & Kolen, M. J. (2012). Equating mixed-format tests using dichotomous common items. In M. J. Kolen & W. Lee (Eds.), *Mixed-format tests: Psychometric properties with a primary focus on equating* (volume 2). (CASMA Monograph Number 2.2,13-44). Iowa City, IA: CASMA, The University of Iowa.
- Li, Y.H., Lissitz, R. W., & Yang, Y.-N. (1999). *Estimating IRT equating coefficients for tests with polytomously and dichotomously scored items*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Montreal, Canada.
- Livingston, S. A. (2004). *Equating test scores (without IRT)*. Princeton, NJ: Educational Testing Service.
- Livingston, S. A. (2009). *Constructed-response test questions: why we use them; how we use them*. NJ: Educational Testing Service.
- Loyd, B. H., & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, 17, 179-193.
- Lukhele, R., Thissen, D., & Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement*, 31, 234-250.
- Marco, G. L., (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, 14, 139-160.

- Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34, 207-218.
- Meng, H. (2007). *A comparison study of IRT calibration methods for mixed-format tests in vertical scaling*. Unpublished doctorate dissertation. University of Iowa. Iowa.
- Muraki, E. ve Carlson, E. B. (1995). Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement*, 19, 73-90.
- OSYM (2016). *Akademik becerilerin izlenmesi ve değerlendirilmesi (abide) projesi*. [Çevrimiçi: <http://abide.meb.gov.tr/proje-hakkinda.asp>, Erişim tarihi: 02.06.2016.]
- Patz, R. J., & Hanson, B. A. (2002). *Psychometric issues in vertical scaling*. Paper presented at the annual meetings of the National Council on Measurement in Education, New Orleans, LA.
- Pollock, J. M., & Rock, D. A. (1997). *Constructed-response tests in the NELS:88 high school effectiveness study*. Washington, DC: National Center for Education Statistics, U.S. Department of Education, Office of Educational Research and Improvement.
- R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- R. Philip Chalmers (2012). mirt: A Multidimensional item response theory package for the R Environment. *Journal of Statistical Software*, 48(6), 1-29.
- Reckase, Mark D. (2009). *Multidimensional item response theory*. New York: Springer.
- Ricker, K. L., ve von Davier, A. A. (2007). *The impact of anchor test length on equating results in a nonequivalent groups design*. Technical Report (RR-07-44).
- Rosa, K., Swygert, K., Nelson, L. & Thissen, D. (2001). Item response theory applied to combinations of multiple-choice and constructed response items: scale scores for patterns of summed scores. D. Thissen and H. Wainer (Eds.), *Test scoring* (253-292). Hillsdale, NJ: Lawrence Erlbaum.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17.
- Samejima, F. (1972). A general model for free-response data. *Psychometric Monograph*, No. 18.
- Schilling, S., & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, 70(3), 533-555.
- Smith, R. L., & Yen, W. M. (2006). Models for evaluating grade-to-grade growth. In R. W. Lissitz (Ed.), *Longitudinal and value added modeling of student performance* (82-94). Maple Grove, MN: JAM Press.
- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201-210.
- Suksuwana, S., Junpengb, P., Ngudgratokec S. ve Guayjarernpanishk, P. (2012). The effect of proportion common item's with mixed format test on multidimensional item response theory linking. *Procedia - Social and Behavioral Sciences*. 69, 1505-1511.

- Sykes, R. C., Hou, L-L., Hanson, B., ve Wang, Z. (2002). *Multidimensionality and the equating of a mixed-format math examination*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.
- 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, 329-346.
- Thissen, D., Wainer, H., ve Wang, X. (1994). Are tests comprising both multiple-choice and free-response items necessarily less unidimensional than multiple-choice tests? An analysis of two tests. *Journal of Educational Measurement*. 31, 113-123.
- Tong, Y. (2005). Comparison of methodologies and results in vertical scaling for educational achievement tests. *Unpublished doctorate dissertation*, The University of Iowa, Iowa City.
- Tong, Y., and Kolen, M.J. (2007). Comparisons of methodologies and results in vertical scaling for educational achievement tests. *Applied Measurement in Education*, 20(2), 227-253.
- Trautmann, H., Steuer, D., Mersmann, O. ve Bornkamp, B. (2014). *truncnorm: Truncated normal distribution*. R package version 1.0-7.
- Uysal, İ. (2014). *Madde Tepki Kuramı'na dayalı test eşitleme yöntemlerinin karma modeller üzerinde karşılaştırılması*. Yüksek lisans tezi. Eğitim Bilimleri Enstitüsü. Eğitimde Ölçme ve Değerlendirme Bilim Dalı. Abant İzzet Baysal Üniversitesi. Bolu.
- von Davier, A., Holland, P., Thayer, D. T. (2004). *The kernel method of test equating*. New York: Springer.
- Wainer, H. & Thissen D. (1993). Combining multiple-choice and constructed response test scores: Toward a marxist theory of test construction, *Applied Measurement in Education*, 6(2), 103-118.
- Wainer, H., Wang, X.B., & Thissen, D. (1994). How well can we compare scores on test forms that are constructed by examinees choice? *Journal of Educational Measurement*, 31, 183–199.
- Walker, M. E. & Kim, S. (2010). *Examining two strategies to link mixed-format tests using multiple-choice anchors*. *Technical Report (RR-10-18)*. Princeton, NJ: Educational Testing Service.
- Walker, M. E., & Kim, S. (2009). *Linking mixed-format tests using multiple choice anchors*. Paper presented at the annual conference of the National Council on Measurement in Education, San Diego, CA.
- Wang, T., Lee, W., Brennan, R. L., & Kolen, M. J. (2008). A comparison of the frequency estimation and chained equipercentile methods under the common-item non-equivalent groups design. *Applied Psychological Measurement*, 32(8), 632-651.
- Wu, N., Huang, C., Huh, N., & Harris, D. (2009). *Robustness in using multiple-choice items as an external anchor for constructed-response test equating*. Paper presented at the 2009 annual meeting of the National Council on Measurement in Education, San Diego, CA.

- Yao, L. ve Boughton, K. (2009). Multidimensional linking for tests with mixed item types. *Journal of Educational Measurement*, 46 (2), 177–197.
- Yen, W. M. (1985). Increasing item complexity: a possible cause of scale shrinkage for unidimensional item response theory. *Psychometrika*, 50, 399-410.
- Yu, Chong Ho & Sharon E. Osborn Popp (2005). *Test equating by common items and common subjects: concepts and applications. Practical Assessment Research & Evaluation*, 10(4). [Çevrim-içi: <http://pareonline.net/getvn.asp?v=10&n=4>, Erişim tarihi: 02.05.2016.]





# EK 1. ETİK KURUL İZİN MUAFİYET FORMU

Form: 40

## Tez Çalışması Etik Komisyon İzin Muafiyeti Formu

29 / 06 / 2016

Hacettepe Üniversitesi  
Eğitim Bilimleri Enstitüsü  
Eğitim Bilimleri Anabilim Dalı Başkanlığı'na

**Tez Başlığı / Konusu:** Çok Boyutlu Karma-Format Testlerin Ölçeklenmesini Etkileyen Faktörlerin İncelenmesi

Yukarıda başlığı/konusu gösterilen tez çalışmam:

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır.
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne müdahale içermemektedir.
4. Gözlemsel ve betimsel araştırma (anket, ölçek/skala çalışmaları, dosya taramaları, veri kaynakları taraması, sistem-model geliştirme çalışmaları) niteliğinde değildir.

Hacettepe Üniversitesi Etik Kurullar ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Akif AVCU  
(Öğrencinin Adı, Soyadı, İmzası)

### Öğrenci Bilgileri

<b>Adı Soyadı</b>	Akif AVCU
<b>Öğrenci No</b>	N11240303
<b>Anabilim Dalı</b>	Eğitim Bilimleri
<b>Programı</b>	Eğitimde Ölçme ve Değerlendirme
<b>Statüsü</b>	<input type="checkbox"/> Yüksek Lisans <input checked="" type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr.

### Danışman Görüşü ve Onayı

Yapılan bu tez çalışması için etik kurul izni almaya gerek duyulmamıştır.

Prof. Dr. Hülya KELECİOĞLU  
(İmza)  
(Danışmanın Ünvanı, Adı ve Soyadı)



## EK 2. ORJİNALLİK RAPORU

Turnitin Doküman Görüntüleyici - Google Chrome  
https://turnitin.com/div?lang=tr&o=681790439&s=3&u=1051458858  
Deneme ÇOK BOYUTLU KARMA-FORMAT TESTLERİNİN İNCELENMESİ

Originality GradeMark PeerMark  
1/1

turnitin.com

AKIF AVCI TARAFINDAN

ÇOK BOYUTLU KARMA-FORMAT TESTLERİNİN ÖLÇEKLENMESİNİN ETKİLEYEN FAKTÖRLERİNİN İNCELENMESİ

AN INVESTIGATION OF THE FACTORS AFFECTING THE SCALING OF MULTIDIMENSIONAL MIXED FORMAT TESTS

AKIF AVCI

Hacettepe Üniversitesi  
Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin  
Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı İçin  
Ongörülüğü  
Yüksek Lisans/Doçtora Tezi  
olarak hazırlanmıştır.

2016

Eşleşmeyi Gözden Geçir

1	www.egitimbilimlendern... İnternet kaynağı	%1
2	www.openaccess.hace... İnternet kaynağı	%1
3	blog.is İnternet kaynağı	%<1
4	katalog.hacettepe.edu.tr İnternet kaynağı	%<1
5	blog.remzihoca.com İnternet kaynağı	%<1
6	content.onlinejacc.org İnternet kaynağı	%<1
7	dergipark.ulakbim.gov.tr İnternet kaynağı	%<1
8	Leung, A.W. "Existenc... Yayın	%<1

SAYFA: 1 / 08

Masaüstü TR

Sel-Meln Raporu

1604  
06.06.2016

## ÖZGEÇMİŞ

### Kişisel Bilgiler

<i>Adı Soyadı</i>	Akif AVCU
<i>Doğum Yeri</i>	Hekimhan/Malatya
<i>Doğum Tarihi</i>	26/09/1983

### Eğitim Durumu

<i>Lise</i>	Çukurova Lisesi/Osmaniye	2001.
<i>Lisans</i>	Boğaziçi Üniversitesi/Psikolojik Danışmanlık ve Rehberlik	2007
<i>Yüksek Lisans</i>	Minnesota Üniversitesi/Eğitimde Nicel Yöntemler-Ölçme	2011
<i>Yabancı Dil</i>	İngilizce: Okuma (iyi), Yazma (iyi), Konuşma (iyi)	

### İş Deneyimi

<i>Stajlar</i>		
<i>Projeler</i>		
<i>Çalıştığı Kurumlar</i>	-Ümraniye Rehberlik Araştırma Merkezi -Marmara Üniversitesi	2007-2009 2011-

### İletişim

<i>e-Posta Adresi</i>	avcuakif@yahoo.com
	avcuakif@gmail.com
<i>Jüri Tarihi</i>	31.06.2016