

**KERNEL EŐİTLEME VE MADDE TEPKİ KURAMINA  
DAYALI EŐİTLEME YÖNTEMLERİNİN  
KARŐİLAŐTIRILMASI**

**COMPARISON OF KERNEL EQUATING AND ITEM  
RESPONSE THEORY EQUATING METHODS**

**Çiğdem AKIN ARIKAN**

Hacettepe Üniversitesi

Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Doktora Tezi

olarak hazırlanmıştır.

2017

## KABUL ve ONAY

Eđitim Bilimleri Enstitüsü M¼d¼rl¼ę¼'ne,

Çiđdem AKIN ARIKAN'ın hazırladıđı "Kernel Eđitleme ve Madde Tepki Kuramı Eđitleme Y¼ntemlerinin Karşılařtırılması" bařlıklı bu çalıřma j¼rimiz tarafından **Eđitim Bilimleri Anabilim Dalı, Eđitimde Olçme ve Deđerlendirme Bilim Dalı'nda Doktora Tezi** olarak kabul edilmiřtir.

Başkan Prof. Dr. H¼lya KELECİOđLU



¼ye (Danıřman) Prof. Dr. Selahattin GELBAL



¼ye Prof. Dr. Nuri DOđAN



¼ye Doç. Dr. Neře G¼LER



¼ye Doç. Dr. Hakan Yavuz ATAR



### ONAY

Bu tez Hacettepe ¼niversitesi Lisans¼st¼ Eđitim-¼đretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri ¼yeleri tarafından...../...../..... tarihinde uygun g¼r¼lm¼ř ve Enstit¼ Y¼netim Kurulunca...../...../.....tarihinde kabul edilmiřtir.

Prof. Dr. Ali Ekber řAHİN  
Eđitim Bilimleri Enstitüsü M¼d¼r¼

## YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

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

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

**Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.**

(Bu seçenekle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etseniz ve kütüphane bu talebinizi yerine getirirse bile, teziniz arama motorlarının önbelleklerinde kalmaya devam edebilecektir)

**Tezimin/Raporumun ..... tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.**

(Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisi alınabilir).

**Tezimin/Raporumun ..... tarihine kadar erişime açılmasını istemiyorum ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.**

**Serbest Seçenek/Yazarın Seçimi:** .....

.....

01/08/2017

Çiğdem AKIN ARIKAN

## 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.

Çiğdem AKIN ARIKAN

## TEŞEKKÜR

Doktora öğrenimime başladığım günden bu güne her türlü desteği ile bilgisini esirgemeyen sürekli yanımda olan ve çalışmamın tamamlanmasında büyük katkıları olan değerli hocam ve danışmanım Prof. Dr. Selahattin GELBAL'a,

Doktora öğrenimim boyunca kendimi geliştirmemde büyük katkıları olan, bilgilerinden faydalandığım değerli hocalarım Prof. Dr. Hülya KELECİOĞLU'na, Prof. Dr. Nuri DOĞAN'a ve Doç. Dr. Burcu ATAR'a,

Tez jürimde yer alarak değerli görüş ve önerilerini esirgemeyerek katkıda bulunan Doç. Dr. Hakan Yavuz ATAR'a ve Doç. Dr. Neşe GÜLER'e,

Desteklerini esirgemeyen sevgili arkadaşlarım Araş. Gör. Sümeyra SOYSAL'a, Dr. Sakine GÖÇER ŞAHİN'e ve Araş. Gör. Ceylan GÜNDEĞER'e,

Beni her zaman destekleyen, "kız çocuğu okutmasan da olur" diyen herkesin karşısında duran ve her tökezlememde beni ayağa kaldıran canım annem ve babama;

Tez yazma sürecinde omuzlarımdaki yükü hafifleten kayınvalidem Hanım ARIKAN'a,

Bu süreçte sürekli yanımda olan, varlığıyla bana güç veren ve bütün kararlarımdaki tek destekçim, yol arkadaşım, dostum, sevgili eşim İnan ARIKAN'a,

Ve bu günlere gelmemde emeği geçen herkese sonsuz **TEŞEKKÜRLER...**

Canım Ođlum Rodin Ali'ye,

# KERNEL EŐİTLEME VE MADDE TEPKİ KURAMI EŐİTLEME YÖNTEMLERİNİN KARŐİLAŐTIRILMASI

Çiğdem AKIN ARIKAN

## ÖZ

Bu araŐtırmada denk olmayan gruplarda ortak madde deseninde Madde Tepki Kuramı gerçek puan eŐİtleme (Haebara) ile Kernel son tabakalama eŐİt yüzdelikli, Kernel son tabakalama dođrusal, Kernel zincirleme eŐİt yüzdelikli ve Kernel zincirleme dođrusal eŐİtleme yöntemlerinin performansları, eŐİtleme hatası (RMSD) ve eŐİtlemenin standart hatasına (SEE) göre deđerlendirilmiŐtir. Bu amaçla örneklem büyüklüğü, yetenek dağılımı, ortak madde tipi, ortak madde oranı ve ortak madde güçlük dağılımı koŐulları ele alınmıŐtır.

Bu araŐtırmada, test uzunluđu 50 olarak sabit tutularak, 3 örneklem büyüklüğü, 2 yetenek dağılımı, 2 ortak test tipi (iç ve dıŐ), 3 ortak madde oranı ve 2 ortak madde güçlük dağılımı (mini ve midi) olmak üzere 72 farklı simülasyon koŐuluyla beŐ eŐİtleme yöntemini karşılaŐtıran bir desen oluşturulmuŐtur. AraŐtırma kapsamında gerçekteŐirilen analizler R yazılımı ile gerçekteŐirilmiş ve her bir koŐul için 100'er tekrar yapılmıŐtır. MTK gerçek puan eŐİtlemede madde parametrelerinin kestirimi için R yazılımında yer alan "lrm" paketi (Rizopoulos, 2015), elde edilen madde parametrelerini aynı ölçeđe yerleŐtirmek ve test puanlarının eŐİtlenmesi için "plink" paketi (Weeks, 2010), Kernel eŐİtleme yöntemleri için ise "kequate" paketi (Andersson, Branberg & Wiberg, 2013) kullanılmıŐtır. 3PLM'ye uygun olarak "irtoys" paketi (Partchev, 2016) kullanılarak iki kategorili puanlanan test formları (ortak test ve ana test formları) üretilmiŐtir. ÇalıŐmada yer alan koŐulların eŐİtleme hatası ve standart hatası üzerindeki ortak etkisi için "lattice" paketi (Sarkar, 2017) kullanılmıŐtır.

AraŐtırmadan elde edilen sonuçlar incelendiđinde, örneklem büyüklüğü, yetenek dağılımı, ortak madde tipi, ortak madde oranı ve ortak madde güçlük dağılımı koŐullarının ve bu koŐulların etkileŐimlerinin eŐİtleme yöntemlerinin performansını etkilediđi görülmüŐtür. Örneklem büyüklüğü arttıkça eŐİtleme hatasının ve eŐİtlemenin standart hatasının azaldıđı; ancak örneklem büyüklüğünün standart hata üzerindeki etkisinin toplam hata üzerindeki etkisine göre daha fazla olduđu görülmüŐtür. Gruplar arası yetenek dağılımı farklı olduđunda bütün eŐİtleme

yöntemlerinde hatanın arttığı, ancak bu durumun eşitleme yöntemlerine göre farklılık gösterdiği sonucuna ulaşılmıştır. Gruplar arası yetenek dağılımı farklılaşmasından zincirleme eşitleme yöntemlerinin daha az etkilendiği görülmüştür. Ayrıca Kernel eşitleme yöntemlerinde gruplar arası yetenek dağılımı farklı olduğunda uç puanların, MTK gerçek puan eşitlemede ise orta yüksek puanların daha yüksek standart hata verdiği sonucuna ulaşılmıştır. Dış ortak testte, iç ortak teste göre daha düşük standart ve toplam hata elde edilmiş ve ortak madde oranı arttıkça standart ve toplam hatanın azaldığı bulunmuştur. İç ortak testte, gruplar arası yetenek dağılımı benzer olduğunda mini ve midi ortak test benzer sonuçlar verirken, gruplar arası yetenek dağılımı farklı olduğunda midi ortak test; dış ortak testte ise midi ortak test daha iyi sonuçlar vermiştir. Kernel eşitlemeye dayalı yöntemler karşılaştırıldığında, doğrusal eşitleme yöntemleri, eşit yüzdeli eşitleme yöntemlerine göre daha düşük standart hata vermiş; eşitleme hatası açısından ise Kernel eşit yüzdeli eşitleme yöntemleri daha iyi sonuçlar vermiştir. Bütün koşullarda Kernel eşitleme yöntemlerinin MTK gerçek puan eşitleme yöntemine göre orta puan ölçeğinde daha düşük standart hata verdiği, puan frekansının daha az olduğu uç puanlarda ise daha yüksek standart hata verdiği bulunmuştur. Uç puanlarda MTK gerçek puan eşitleme yöntemden Kernel eşitleme yöntemlerine göre daha düşük hata elde edilmiştir. Bu nedenle, özellikle bireyler hakkında önemli kararlar verilecek durumlarda Kernel eşitleme yöntemleri yerine MTK gerçek puan eşitleme yöntemleri kullanılabilir.

Genel olarak, yeni bir yaklaşım olan Kernel eşitleme yöntemlerinin MTK gerçek puan eşitleme yöntemi kadar iyi sonuçlar verdiği sonucuna ulaşılmıştır. Gerçek puan eşitleme yöntemlerinin kullanılmasının uygun olmadığı durumlarda Kernel eşitleme yöntemlerinin kullanılması önerilebilir. Test formlarının eşitlenmesi sonucunda elde edilen eşitlenmiş puanlar eşitleme yöntemlerine göre farklılık gösterdiğinden; testin amacı doğrultusunda, eşitleme yöntemlerinin güçlü ve zayıf yönleri dikkate alınarak, eşitleme yöntemine karar verilmelidir.

**Anahtar sözcükler:** Test eşitleme, Kernel eşitleme, madde tepki kuramı, eşitleme hatası

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



# COMPARISON OF KERNEL EQUATING AND ITEM RESPONSE THEORY EQUATING METHODS

Çiğdem AKIN ARIKAN

## ABSTRACT

This study aimed to evaluate the performances of Item Response Theory (IRT) true-score equating (Haebara) and the Kernel post-stratification equipercentile, Kernel post-stratification linear, Kernel chained equipercentile and Kernel chained linear equating methods based on equating errors (the root-mean-square difference-RMSD) and standard error of equating (SEE) using the common-item nonequivalent groups design. To this purpose; the sample size, group ability differences, common item type, common item rate and the difficulty distribution of common items were examined.

The current study was designed as a comparison of five equating methods on 72 simulation conditions consisting of three sample sizes, two group ability distribution, two common item type (internal and external), three rates of common items, and two item difficulty distribution of common items (mini and midi common test) by fixing the test length at 50 items. The analysis of the study was conducted on R software, and 100 replication were performed for each condition. The “itm” package (Rizopoulos, 2015) of the R software was used in item parameter estimation in the IRT true-score equating; the “plink” package (Weeks, 2010) was used for scaling item parameters, and for the equating test scores, and the “kequate” package (Andersson, Branberg & Wiberg, 2013) was used for Kernel equating methods. Test forms scored in two categories (common test and main test forms) were generated using the “irtoys” package (Partchev, 2016) in accordance with the 3PLM model. For the common effect of equating error and standard error on the conditions of the present study, the ‘lattice’ package (Sarkar, 2017) was utilized.

The results showed that the conditions, including the sample size, the distribution of ability, the type of common item, the rate of common items and the power distribution of common items, as well as the interaction of these conditions, effected the performance of the equation methods. However, the performances of

the methods were found to be different based on these conditions. The larger the sample size, the less the equating error and the standard error of equating. However, the effect of the sample size on the standard error was found to be greater than on the total error. As the group ability distributions varied, the errors in all equating methods increased. However the increase in the errors differs by equating methods. The chained equating methods were observed to be less affected by the difference on ability distributions between groups. Furthermore, when ability distributions differ, the extreme scores have the greatest error in Kernel methods, and middle and high scores have the greatest error in IRT true score equating. The external common test revealed a lower standard and total error than the internal common test. As the rate of common items increased, the standard errors and the total errors decreased. In the internal common test, the mini and midi common tests concluded similar results when the group ability differences between groups were similar. When the group ability distribution were different, while the midi common test concluded better results, than midi common test in the external common test, the mini common test gave better results than mini common test in internal common test.

When Kernel equating methods were compared, linear equating method performed better with respect to standard errors while Kernel equipercentile equating method performed better with respect to equating error. In every condition, the Kernel equating methods were found to reveal lower standard errors in the medium score scale, and higher standard errors in extreme scores where score frequency was lower, compared to the IRT true-score equating. Regarding the extreme scores, less errors were obtained through the IRT true-score equating method than the Kernel equating methods. Therefore, the IRT true-score equating methods could be used instead of the Kernel equating methods, especially when important decisions are to be made about individuals.

In general, it was concluded that the results obtained from Kernel equating methods as a new approach were as appropriate as those of the IRT true-score equating method. Use of the Kernel equating methods can be recommended when the true-score equating methods are not available. Since the equated scores resulting from the equation of test forms vary by the equating methods, the decision regarding which equating method to use should be made after

considering the strengths and weaknesses of each method in accordance with the test purpose.

**Keywords:** Test equating, Kernel equating, item response theory, equating error

**Advisor:** Prof. Dr. Selahattin GELBAL, Hacettepe University, Department of Educational Science, Program in Measurement and Evaluation

## İÇİNDEKİLER

KABUL ve ONAY.....	ii
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI .....	iii
ETİK BEYANNAMESİ .....	iv
TEŞEKKÜR.....	v
ÖZ .....	vii
ABSTRACT .....	ix
İÇİNDEKİLER.....	xii
TABLolar DİZİNİ .....	xv
ŞEKİLLER DİZİNİ.....	xvi
SİMGELER VE KISALTMALAR DİZİNİ .....	xviii
1. GİRİŞ.....	1
1.1. Problem Durumu.....	1
1.2. Araştırmanın Amacı ve Önemi.....	4
1.3. Problem Cümlesi .....	5
1.3.1. Alt Problemler.....	6
1.4. Sayıtlar.....	7
1.5. Sınırlılıklar.....	7
1.6. Tanımlar.....	7
1.7. Araştırmanın Kuramsal Temeli .....	7
1.7.1. Test Eşitleme .....	8
1.7.1.1. Eşitleme Koşulları.....	9
1.7.1.2. Eşitleme Desenleri .....	12
1.7.1.2.1. Tek grup deseni .....	12
1.7.1.2.2. Random grup deseni .....	13
1.7.1.2.3. Denk olmayan gruplarda ortak madde deseni .....	13
1.7.2. Eşitleme Metotları.....	14
1.7.2.1. Kernel Eşitleme.....	15
1.7.2.2. Madde Tepki Kuramına Dayalı Eşitleme.....	27
1.7.2.2.1. Rasch Model ve Bir Parametrelili Lojistik Model (1PLM) .....	28
1.7.2.2.2. İki Parametrelili Lojistik Model (2PLM).....	29
1.7.2.2.3. Üç Parametrelili Lojistik Model (3PLM) .....	29
1.7.2.2.4. Ayır Kestirim (Seperate Estimation) .....	30
2. İLGİLİ ARAŞTIRMALAR.....	33
2.1. İlgili Araştırmalar Özet.....	43
3. YÖNTEM.....	44
3.1. Araştırmanın Türü.....	44
3.2. Araştırmada Kullanılan Eşitleme Deseni.....	44
3.3. Simülasyon Koşulları .....	44
3.3.1. Örneklem Büyüklüğü .....	45
3.3.2. Gruplararası Ortalama Yetenek Farkı .....	46
3.3.3. Ortak Madde Tipi ve Oranı .....	46
3.3.4. Ortak Madde Güçlük Dağılımı .....	46

3.4. Verilerin Üretilmesi.....	47
3.5. Eşitleme Süreci ve Verilerin Analizi .....	48
3.6. Değerlendirme Kriterleri .....	50
4. BULGULAR VE TARTIŞMA .....	53
4.1. Alt Problem 1'e İlişkin Bulgular ve Yorumlar .....	53
4.1.a. Eşitleme hatası çalışmada ele alınan koşullara göre nasıl değişmektedir?.....	53
4.1.b. Eşitlemenin standart hatası ve her bir puan için elde edilen standart hata çalışmada ele alınan koşullara göre nasıl değişmektedir? .....	58
4.2. Alt Problem 2'ye İlişkin Bulgular ve Yorumlar .....	64
4.2.a. Eşitleme hatası çalışmada ele alınan koşullara göre nasıl değişmektedir? .....	64
4.2.b. Eşitlemenin standart hatası ve her bir puan için elde edilen standart hata çalışmada ele alınan koşullara göre nasıl değişmektedir? .....	70
4.3. Alt Problem 3'e İlişkin Bulgular ve Yorumlar .....	76
4.3.a. Eşitleme hatası çalışmada ele alınan koşullara göre nasıl değişmektedir? .....	76
4.3.b. Eşitlemenin standart hatası ve her bir puan için elde edilen standart hata çalışmada ele alınan koşullara göre nasıl değişmektedir? .....	81
4.4. Alt Problem 4'e İlişkin Bulgular ve Yorumlar .....	87
4.5. Alt Problem 5'e İlişkin Bulgular ve Yorumlar .....	89
4.6. Tartışma.....	91
5. SONUÇ ve ÖNERİLER .....	99
5.1. Sonuçlar.....	99
5.2. Öneriler.....	101
5.2.1. Uygulayıcılara Yönelik Öneriler .....	101
5.2.2. İleri Araştırmalara Yönelik Öneriler.....	102
EKLER DİZİNİ .....	112
EK 1. ETİK KOMİSYONU ONAY BİLDİRİMİ .....	113
EK 2. ORJİNALLİK RAPORU.....	114
EK 3. VERİ SETLERİNİN ÜRETİLMESİNDE KULLANILAN R KODU .....	116
EK 4. ORTAK MADDELERE İLİŞKİN BETİMSSEL İSTATİSTİKLER.....	118
EK 5. 500 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN RMSD DEĞERLERİ.....	119
EK 6. 1500 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN RMSD DEĞERLERİ.....	120
EK 7. 5000 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN RMSD DEĞERLERİ.....	121
EK 8. 500 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN SEE DEĞERLERİ.....	122
EK 9. 1500 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN SEE DEĞERLERİ.....	123

EK 10. 5000 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN SEE DEĞERLERİ.....	124
EK 11. PUAN ÖLÇEĞİ BOYUNCA 500 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN SEE DEĞERLERİNE İLİŞKİN GRAFİKLER .....	125
EK 12. PUAN ÖLÇEĞİ BOYUNCA 1500 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN SEE DEĞERLERİNE İLİŞKİN GRAFİKLER .....	130
EK 13. PUAN ÖLÇEĞİ BOYUNCA 5000 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN SEE DEĞERLERİNE İLİŞKİN GRAFİKLER .....	135
ÖZGEÇMİŞ .....	140

## TABLULAR DİZİNİ

Tablo 1.1: Tek Grup Deseni .....	12
Tablo 1.2: Random Grup Deseni.....	13
Tablo 1.3: Denk Olmayan Gruplarda Ortak Madde Deseni .....	14
Tablo 3.1: Araştırmanın Eşitleme Deseni .....	44
Tablo 3.2: Araştırmanın Simülasyon Koşulları.....	45

## ŞEKİLLER DİZİNİ

Şekil 1.1. Kestirim ve Eşitlemede Simetri Özelliğinin Farklılığı .....	10
Şekil 1.2. Zincirleme Eşitlemenin Grafikselsel Gösterimi .....	24
Şekil 3.1. İç ve Dış Ortak Testin Test Formlarında Gösterimi.....	48
Şekil 4.1. 500 Örneklem için İç Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen RMSD Değerleri .....	54
Şekil 4.2. 500 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen RMSD Değerleri.....	54
Şekil 4.3. 500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen RMSD Değerleri .....	56
Şekil 4.4. 500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen RMSD Değerleri.....	56
Şekil 4.5. 500 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen SEE Değerleri.....	59
Şekil 4.6. 500 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Farklı Olduğunda Elde Edilen SEE Değerleri.....	59
Şekil 4.7. 500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen SEE Değerleri.....	61
Şekil 4.8. 500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen SEE Değerleri .....	61
Şekil 4.9. 500 Örneklem için ortak madde oranının %20 olduğu koşulda elde edilen SEE değerleri.....	63
Şekil 4.10. 1500 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen RMSD Değerleri .....	65
Şekil 4.11. 1500 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen RMSD Değerleri .....	66
Şekil 4.12. 1500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen RMSD Değerleri .....	67
Şekil 4.13. 1500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen RMSD Değerleri .....	68
Şekil 4.14. 1500 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen SEE Değerleri .....	70
Şekil 4.15. 1500 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen SEE Değerleri.....	71
Şekil 4.16. 1500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen SEE Değerleri .....	72
Şekil 4.17. 1500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen SEE değerleri .....	73
Şekil 4.18. 1500 örneklem için ortak madde oranının %20 olduğu koşulda elde edilen SEE değerleri.....	75



Şekil 4.19. 5000 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen RMSD Değerleri .....	77
Şekil 4.20. 5000 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen RMSD Değerleri .....	77
Şekil 4.21. 5000 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen RMSD Değerleri .....	79
Şekil 4.22. 5000 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen RMSD Değerleri .....	79
Şekil 4.23. 5000 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen SEE Değerleri .....	82
Şekil 4.24. 5000 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen SEE Değerleri .....	82
Şekil 4.25. 5000 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen SEE Değerleri .....	84
Şekil 4.26. 5000 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen SEE Değerleri .....	84
Şekil 4.27. 5000 Örneklem İçin Ortak Madde Oranının %20 olduğu Koşulda Elde Edilen SEE Değerleri .....	86
Şekil 4.28. İç ortak testte çalışmada ele alınan koşulların eşitleme hatasına (RMSD) ortak etkisi .....	88
Şekil 4.29. Dış ortak testte çalışmada ele alınan koşulların eşitleme hatasına (RMSD) ortak etkisi .....	88
Şekil 4.30. İç ortak testte çalışmada ele alınan koşulların eşitlemenin standart hatasına (SEE) ortak etkisi .....	90
Şekil 4.31. Dış ortak testte çalışmada ele alınan koşulların eşitlemenin standart hatasına (SEE) ortak etkisi .....	90

## SİMGELER VE KISALTMALAR DİZİNİ

**MTK:** Madde Tepki Kuramı

**KE:** Kernel Eşitleme

**1PLM:** Bir parametrelili lojistik model

**2PLM:** İki parametrelili lojistik model

**3PLM:** Üç parametrelili lojistik model

**RMSD:** Ortalama Hatanın Kare Kökü Farkı (Hata)

**SEE:** Eşitlemenin Standart Hatası

**Z1:** Kernel Zincirleme Eşit Yüzdellikli Eşitleme

**ST1:** Kernel Son Tabakalama Eşit Yüzdellikli Eşitleme

**Z2:** Kernel Zincirleme Doğrusal Eşitleme

**ST2:** Kernel Son Tabakalama Doğrusal Eşitleme

**YB:** Gruplar arası yetenek dağılımının benzer olduğu koşul

**YF:** Gruplar arası yetenek dağılımının farklı olduğu koşul

# 1. GİRİŞ

Bu bölümde problem durumu, araştırmanın amacı ve önemi, problem cümlesi, alt problemler, sayılılar, tanımlar ve sınırlılıklar yer almaktadır.

## 1.1. Problem Durumu

Bireyler hakkında bilgi toplamak, ilgi, yetenek, tutum gibi özelliklerini ölçmek, bireyleri seçmek ve yerleştirmek gibi birçok farklı amaç için kullanılan testler, bireyler hakkında verilen önemli kararların bir parçasıdır. Birçok test uygulamasında, özellikle geniş ölçekli ve yüksek riskli testlerde, test güvenliği ve farklı günlerde bireylerin test edilebilmesi için aynı testin farklı formlarının geliştirilmesi önemli bir gereklilik olarak uygulayıcıların karşısına çıkmaktadır. Testlerin farklı formlarının kullanılmasıyla, bireylerin test soruları ile ilgili ön bilgisinin olmaması sağlanmış ve bir gruba karşı avantaj ortadan kaldırılmış olur. Ancak testin farklı formlarını kullanmak, bu formlardan elde edilen puanların karşılaştırılabilir veya birbiri yerine kullanılabilir olup olmadığı sorusunu akla getirir. Uygulamada, farklı sorularla paralel test formları geliştirmek (aynı özelliği ölçen, aynı güvenilirlikte ve aynı güçlükte vb.) oldukça güçtür. Kapsam ve psikometrik özellikler açısından her ne kadar benzer test formlar geliştirilmeye çalışılırsa da, bu formlar güçlük bakımından farklılık gösterebilir. Test formlarının güçlük düzeyleri arasındaki bu farkı açıklayabilmek ve farklı formlardan alınan puanları birbiri yerine kullanabilmek için, formlardan alınan puanlar arasında bir ilişki kurulması gereklidir (Cook, 2007). Test puanları arasında kurulan bu ilişkiye genel olarak bağlama (linking) denir. Holland'a (2007) göre bağlama iki test formuna ait puanlar arasındaki genel bir dönüştürmedir ve kestirim (predicting), ölçekleme (scale aligning veya scaling) ve eşitleme (equating) olarak üç kategoriye ayrılır. Puanları bağlamanın en eski yolu olan kestirim, birey hakkındaki mevcut bilgilerin (başka bir test puanı, demografik bilgiler vb.) doğrusal regresyon modeli kullanılarak test puanının tahmin edilmesidir ve kestirim hem gözlenen puanlara hem de gerçek puanlara dayalı olabilir.

Diğer bir yöntem olan ölçekleme ise ikinci en eski bağlama metodudur. Ölçeklemenin amacı, iki farklı test formuna ait puanların aynı ölçeğe yerleştirilmesidir ve ölçekleme için kullanılan işlemler eşitleme için de kullanılabilir

(Dorans & Walker, 2007). Holland (2007), iki test puanı arasındaki en güçlü bağlama metodunun eşitleme olduğunu belirtmektedir. İki veya daha fazla test arasındaki ilişkiyi açıklayan istatistiksel bir süreç olan eşitleme, birden fazla test formu varsa ve bu farklı formları alan bireyler birbiriyle karşılaştırılacaksa kullanılır. Eğer test formları başarılı bir şekilde eşitlenmiş ise iki test formu arasında gözlenen fark, kolay testi alan gruba bir avantaj veya zor testi alan gruba dezavantaj sağlamamış olur (Kolen, 1988; Kolen & Brennan, 2004). Test eşitlemede olduğu gibi bağlamada da, iki farklı test formundan alınan puanlar arasında ilişki kurarak, puanları aynı ölçek üzerine yerleştirme işlemi yapar. Bu açıdan iki terim de birbirine benzerdir. Fakat bağlamada kapsam ve güçlük açısından test formlarının aynı olması gibi bir şart aranmaz iken, eşitlemede ise birbirleriyle karşılaştırılmak ya da birbiri yerine kullanılmak istenilen farklı test formlarının kapsam ve güçlük açısından benzer olması gerekmektedir (Kolen & Brennan, 2004).

Eşitleme için gerekli olan ilk adım uygun eşitleme yönteminin seçilmesidir. Eşitleme yöntemleri dayandıkları kuram ve varsayımlar açısından farklılaştığından, hem test geliştiriciler hem de bireyler için eşitleme yönteminin seçimi büyük önem taşımaktadır. Test eşitleme yöntemleri, gerçek puan ve gözlenen puan eşitleme olarak sınıflandırılabilir. Gerçek puan eşitlemede, gerçek puanın gözlenen puan ve hatadan oluştuğu kabul edilir ve temel varsayımı, aynı yetenek düzeyindeki bireylerin hangi test formunu aldığına bakılmaksızın aynı gerçek puana sahip olmasıdır. Gerçek puan eşitleme yöntemleri eş zamanlı kestirim (Lord,1980) ve ayrı kestirim metotlarını (Stocking and Lord, 1983; Haebara, Ortalama-Ortalama ve Ortalama-Standart Sapma) kapsamaktadır. Gözlenen puan eşitlemede, belli bir grup için puan dağılımlarının özelliklerinin denk olması sağlanır ve puan dağılımları ortak bir ölçeğe yerleştirilir. Sıklıkla kullanılan gözlenen puan eşitleme yöntemleri ise doğrusal ve eşit yüzdelli eşitleme yöntemleridir (von Davier, 2008). Bu yöntemler eşit yüzdelli eşitlemeye dayalı yöntemler, doğrusal eşitleme yöntemleri, MTK gözlenen puan eşitleme, van der Linden yerel eşitleme, Levine doğrusal olmayan metot ve yeni bir yaklaşım olan Kernel eşitlemeyi kapsar (von Davier, 2013).

Kernel eşitleme, ilk olarak Holland ve Thayer (1981) tarafından tanımlanmış ve daha sonra von Davier, Holland ve Thayer (2004) tarafından geliştirilmiştir. Kernel

eşitleme, kesikli puan dağılımlarının sürekli dağılımlara dönüştürülerek puan dağılımlarının eşitlendiği bir eşit yüzdelikli gözlenen puan eşitleme yöntemidir (von Davier, Holland & Thayer, 2004). Eşit yüzdelikli eşitleme yapılabilmesi için, başlangıçta kesikli puan dağılım fonksiyonlarının süreklileştirilmesi gerekmektedir. Geleneksel eşitleme yöntemleri puan dağılımlarını süreklileştirmek için doğrusal kestirimler kullanırken, Kernel eşitleme Gauss Kernel yaklaşımını kullanır (Lee & von Davier, 2011; von Davier ve diğerleri, 2004; von Davier, Holland, Livingston, Casabianca, Grant & Martin, 2006). Kernel eşitleme, eşit yüzdelikli eşitleme yöntemi olmasına karşın doğrusal eşitleme yöntemlerini de içerir (Andersson & von Davier, 2014; von Davier, 2008). Kernel eşitlemede kullanılan bant genişliğinin (h parametresi) seçimi, eşitleme yönteminin eşit yüzdelikli ve doğrusal eşitlemeyi belirlemektedir. Kernel eşitleme beş aşamadan oluşmaktadır. Bunlar: 1) Öndüzgünleştirme: puan dağılımlarının düzgünleştirilmesi için log lineer istatistiksel modelin kullanılması; 2) Eşitleme desenine göre puan olasılıkları kestirilmesi; 3) Süreklileştirme aşamasında kesikli puan dağılımlarını süreklileştirmek için standartlaştırılmış Gauss Kernel yaklaşımını kullanılması; 4) Bir önceki aşamada elde edilen sürekli dağılımların kullanılarak eşitleme yapılması ve 5) Eşitlemenin standart hatası ve eşitlemenin standart hatasının farkının hesaplanmasıdır.

Kernel eşitleme tek grup, eşdeğer grup, dengelenmiş grup ve denk olmayan gruplarda ortak test desenlerinde (Non-Equivalent Groups Anchor Test-NEAT) kullanılabilir (Livingston, 1993; von Davier ve diğerleri, 2004). NEAT deseninde sıklıkla Kernel zincirleme ve Kernel son tabakalama eşitleme olmak üzere iki yöntem kullanılmaktadır (von Davier ve diğerleri, 2006). NEAT deseni için Kernel son tabakalama eşitlemede, ideal bant genişliği seçilirse eşitleme sonuçları frekans kestirimi eşit yüzdelikli eşitlemeye; geniş bant genişliği seçilirse, Braun-Holland doğrusal eşitlemeye yaklaşacaktır. Kernel zincirleme eşitlemede ise, ideal bant genişliği kullanılırsa eşitleme sonuçları zincirleme eşit yüzdelikli eşitlemeye; geniş bant genişliği kullanılırsa zincirleme doğrusal eşitleme metoduna yaklaşır (Ricker & von Davier, 2007; von Davier ve diğerleri, 2006).

van der Linden (2006), gözlenen puan eşitlemedeki problemlerin yaklaşık 100 yıldır test kuramcılarının dikkatini çektiğini belirtmiştir. Gözlenen puan eşitleme yöntemleri diğer yöntemlerden daha önce geliştirilmiş ve kullanımlarının diğer yöntemlere göre basit olmasından dolayı alanyazında daha çok kullanılmaktadır

(Kolen & Brennan; 2014). Kolen (2006), Kernel eşitlemenin geliştirilmesinin, gözlenen puan eşitleme kuramında önemli bir gelişme olduğunu belirtmiştir. Yeni bir yaklaşım olan Kernel eşitlemenin test uygulayıcıları tarafından test puanlarının eşitlenmesinde kullanılmadan önce, uygulamalarda sıklıkla kullanılan geleneksel eşitleme yöntemleri sonuçlarıyla karşılaştırılması; sonuçların benzerlik ve farklılıklarının olup olmadıklarının gösterilmesi önemlidir (Mao, von Davier & Rupp, 2006). Eşitleme yöntemlerinin güçlü ve zayıf yönlerinin bilinmesi, test programlarının gereksinimine göre uygun eşitleme yönteminin seçimini kolaylaştırır. Ayrıca geniş ölçekli ve yüksek riskli testlerde elde edilen eşitleme sonuçlarına göre bireyler hakkında önemli kararlar verildiği için, bir yöntemin diğer yöntemlere göre tercih edilmesinin verilecek kararı nasıl etkilediğinin bilinmesi önemlidir (Kim & Cohen, 2002).

Kernel eşitleme yöntemlerinin birbirleriyle veya diğer geleneksel eşitleme yöntemleri ile karşılaştırıldığı birçok çalışma bulunmaktadır (Choi, 2009; Grant, Zhang & Damiano, 2009; Holland, von Davier, Sinharay & Han, 2006; Kim, 2014; Livingston, 1993; Mao, 2006; Mao, von Davier & Rupp, 2006; Moses & Holland, 2007; Qu, 2007; Ricker & von Davier, 2007; von Davier ve diğerleri, 2006). Ancak Kernel eşitleme yöntemleri ile MTK'ya dayalı eşitleme yöntemleriyle karşılaştırıldığı çok az sayıda çalışmaya rastlanılmıştır (Godfrey, 2007; Meng, 2012; Norman Dvorak, 2009). MTK ile Kernel eşitlemenin karşılaştırıldığı çalışmalarda ise Kernel son tabakalama eşit yüzdelikli eşitleme ile zincirleme eşit yüzdelikli eşitleme metotları karşılaştırılmış, ancak Kernel son tabakalama doğrusal eşitleme (geniş bant genişliği) ile Kernel zincirleme doğrusal eşitleme (geniş bant genişliği) yöntemlerinin karşılaştırıldığı bir çalışmaya rastlanılmamıştır.

Bu nedenle bu çalışmada denk olmayan gruplarda ortak madde deseninde MTK gerçek puan eşitleme ile Kernel son tabakalama eşit yüzdelikli eşitleme, Kernel son tabakalama doğrusal eşitleme, Kernel zincirleme eşit yüzdelikli eşitleme ve Kernel zincirleme doğrusal eşitleme yöntemlerinin karşılaştırılmasına ihtiyaç duyulmuştur.

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

Bu araştırmanın amacı, alanyazında sıklıkla kullanılan MTK gerçek puan eşitleme (Haebara) ile yeni bir yaklaşım olan Kernel eşitleme yöntemlerinden Kernel son

tabakalama eşit yüzdelikli (ideal bant genişliği), Kernel son tabakalama doğrusal (geniş bant genişliği), Kernel zincirleme eşit yüzdelikli (ideal bant genişliği) ve Kernel zincirleme doğrusal (geniş bant genişliği) eşitleme yöntemlerinin performanslarının karşılaştırılarak incelenmesidir. Bu amaçla örneklem büyüklüğü, yetenek dağılımı, ortak madde oranı, ortak madde tipi ve ortak madde güçlük dağılımı koşullarında eşitlenmiş puanlara ilişkin eşitleme hatası (hataların ortalama karekökü farkı -RMSD) ve eşitlemenin standart hatası (SEE) hesaplanmıştır.

Ülkemizde yapılan çalışmalar incelendiğinde (Aksekiöglü, 2017; Aşiret, 2014; Atalay Kabasakal, 2014; Bozdağ, 2007; Demirus, 2015; Gök, 2012; Gündüz, 2015; İnci, 2014; Kahraman, 2012; Kan, 2010; Kelecioğlu, 1994; Kilmen, 2010; Mutluer, 2013; Öztürk, 2010; Şahhüseyinoğlu, 2005; Uçar, 2016; Uysal, 2014) Kernel eşitlemenin ele alındığı bir çalışmaya rastlanmamıştır. Yeni bir yaklaşım olan Kernel eşitlemenin simülasyon çalışmaları ile uygulamada sıklıkla kullanılan MTK gerçek puan eşitleme yöntemi karşılaştırılarak, Kernel eşitlemenin avantaj ve dezavantajlarının ortaya çıkarılması ile hangi durumlarda kullanılmasının daha uygun olduğunun belirlenmesinin önemli olduğu düşünülmektedir. MTK gerçek puan eşitleme ile Kernel eşitlemenin örneklem büyüklüğü, yetenek dağılımı, ortak madde tipi, ortak madde oranı ve ortak madde güçlük dağılımı koşulları altında performanslarının karşılaştırıldığı bu çalışmayla ilgili alanyazındaki boşluğun doldurulacağı düşünülmektedir.

Ülkemizdeki geniş ölçekli testlerin yılın farklı dönemlerinde/aylarında (Yabancı Dil Sınavı ve ALES gibi) uygulanan sınavlarından elde edilen puanlar, akademik personel alımında ve lisansüstü programlara geçişte birbiri yerine kullanılmaktadır. Bu sınavlardan elde edilen puanların karşılaştırılabilmesi ve birbirinin yerine kullanılabilmesi için eşitleme yöntemlerinin uygulanması, bazı bireyler için avantaj ya da dezavantaj oluşturmaması için gerekli görülmektedir. Ancak uygulamada, farklı dönemlerde yapılan sınavlara ait puanların eşitlenmesinin yapılmadığı görülmektedir. Bu araştırmanın yapılacak eşitleme çalışmalarına kaynak olma niteliği taşıyacağı söylenebilir.

### **1.3. Problem Cümlesi**

Testler, Kernel son tabakalama doğrusal, Kernel zincirleme doğrusal, Kernel son tabakalama eşit yüzdelikli, Kernel zincirleme eşit yüzdelikli ve MTK gerçek puan

eşitleme yöntemlerine göre eşitlendiğinde eşitlemenin hatası ve eşitlemenin standart hatası örneklem büyüklüğü, yetenek dağılımı, ortak madde tipi, ortak madde oranı ve ortak madde güçlük dağılımına göre nasıl değişmektedir?

### **1.3.1. Alt Problemler**

Problem cümlesine ilişkin oluşturulan alt problemler aşağıdaki gibidir;

**1)** Testler; Kernel son tabakalama doğrusal, Kernel zincirleme doğrusal, Kernel son tabakalama eşit yüzdelikli, Kernel zincirleme eşit yüzdelikli ve MTK gerçek puan eşitleme yöntemleriyle küçük örneklerde (500) eşitlendiğinde eşitlenmiş puanlara ilişkin;

**1.a.** Eşitleme hatası çalışmada ele alınan koşullara göre nasıl değişmektedir?

**1.b.** Eşitlemenin standart hatası ve her bir puan için elde edilen standart hata çalışmada ele alınan koşullara göre nasıl değişmektedir?

**2)** Testler, Kernel son tabakalama doğrusal, Kernel zincirleme doğrusal, Kernel son tabakalama eşit yüzdelikli, Kernel zincirleme eşit yüzdelikli ve MTK gerçek puan eşitleme yöntemleriyle küçük örneklerde (1500) eşitlendiğinde eşitlenmiş puanlara ilişkin;

**2.a.** Eşitleme hatası çalışmada ele alınan koşullara göre nasıl değişmektedir?

**2.b.** Eşitlemenin standart hatası ve her bir puan için elde edilen standart hata çalışmada ele alınan koşullara göre nasıl değişmektedir?

**3)** Testler, Kernel son tabakalama doğrusal, Kernel zincirleme doğrusal, Kernel son tabakalama eşit yüzdelikli, Kernel zincirleme eşit yüzdelikli ve MTK gerçek puan eşitleme yöntemleriyle büyük örneklerde (5000) eşitlendiğinde eşitlenmiş puanlara ilişkin;

**3.a.** Eşitleme hatası çalışmada ele alınan koşullara göre nasıl değişmektedir?



**3.b.** Eşitlemenin standart hatası ve her bir puan için elde edilen standart hata çalışmada ele alınan koşullara göre nasıl değişmektedir?

**4)** Çalışmada yer alan koşulların Kernel son tabakalama doğrusal, Kernel zincirleme doğrusal, Kernel son tabakalama eşit yüzdelikli, Kernel zincirleme eşit yüzdelikli ve MTK gerçek puan eşitleme yöntemleriyle elde edilen eşitlemenin hatasına ortak etkisi nasıldır?

**5)** Çalışmada yer alan koşulların Kernel son tabakalama doğrusal, Kernel zincirleme doğrusal, Kernel son tabakalama eşit yüzdelikli, Kernel zincirleme eşit yüzdelikli ve MTK gerçek puan eşitleme yöntemleriyle elde edilen eşitlemenin standart hatasına ortak etkisi nasıldır?

#### **1.4. Sayıtlılar**

Araştırma kapsamında kullanılan verilerin gerçek durumları yansıttığı varsayılmaktadır.

#### **1.5. Sınırlılıklar**

1. Araştırma verileri iki kategorili simülasyon verisi ile sınırlıdır.
2. Araştırmada kullanılan eşitleme deseni olarak denk olmayan gruplarda ortak madde deseni ile sınırlıdır.
3. Araştırma, kullanılan eşitleme yöntemleriyle sınırlıdır.

#### **1.6. Tanımlar**

**Mini Ortak Test:** Ortak testin toplam test ile aynı kapsama sahip ve istatistiksel açıdan benzer olduğu durum (Sinharay ve Holland, 2006).

**Midi Ortak Test:** Ortak testin kapsam açısından toplam test ile aynı, ancak madde güçlükleri açısından farklılaştığı ve madde güçlüklerinin orta güçlükte olduğu durum (Sinharay ve Holland, 2006).

#### **1.7. Araştırmanın Kuramsal Temeli**

Bu araştırmanın kuramsal temeli, "Test Eşitleme" ve "Eşitleme Metotları" olmak üzere iki alt başlık çerçevesinde oluşturulmuştur.

### 1.7.1. Test Eşitleme

Eşitleme bir testin alternatif formu olduğu durumlarda kullanılır (Kolen & Brennan, 2014). Test eşitlemeye olan ihtiyaç, test programları için aynı kapsam ve formatta fakat farklı sorulardan oluşan test formlarının geliştirmesi ve uygulanmasından doğmuştur. Ancak formlardaki sorular farklı olduğunda, bu formlardaki maddeler güçlükleri açısından da farklılaşabilmektedirler. Bu durumda kolay test formunu alan birey yüksek puan alırken, zor testi alan birey daha düşük puan alabilir. Farklı formları alan bireyleri karşılaştırmaya duyulan gereksinimden, test güçlükleri arasındaki farkı ayarlamak için test eşitleme teknikleri geliştirilmiştir (Kolen & Brennan, 2014; von Davier ve diğerleri, 2004). Test puanlarının diğer test formundan elde edilen puanlarla ilişkilendirilmesine veya dönüştürülmesine eşitleme denir (Hambleton & Swaminathan, 1985). Lord (1980), her bir yetenek düzeyindeki bireyin X veya Y formunu alması fark yaratmıyorsa, X ve Y formlarının eşit olduğunu belirtmiştir.

Angoff (1971) eşitlemeyi, bir test formunun birim sisteminin diğer formun birim sistemine çevrilmesi olarak tanımlamıştır. Eşitleme, formlar arası güçlük farkını ayarlayarak, bu formlardan elde edilen puanlar ile ölçek puanlarını ilişkilendirmek için istatistiksel metotları kullanır (Tong & Kolen, 2005). Kolen ve Brennan (2004) ise eşitlemeyi benzer test özelliklerine sahip test formları arasındaki farklılıkları düzenleyerek, bu test formlarından elde edilen puanların birbiri yerine kullanılmasını sağlayan istatistiksel bir süreç olarak tanımlamaktadır. Eşitlemenin amacı, farklı test formlarından elde edilen puanların birbiri yerine kullanılabilirliğini sağlamaktır (Angoff, 1971; Tong & Kolen, 2005; von Davier ve diğerleri, 2004).

Test eşitleme genel olarak yatay ve dikey olmak üzere iki sınıfta incelenmektedir (Cook & Eignor, 1983). Yatay eşitleme, test güvenliğini sağlamak için bir testin birden fazla formuna ihtiyaç duyulduğunda kullanılır. Bu formlar aynı olmasa bile, kapsam ve güçlük bakımından benzer olmaları gerekmektedir. Ayrıca test formlarını cevaplayan bireylerin yetenek dağılımlarının da birbirine benzer olması beklenir. Eğer bireylerin yetenek dağılımları birbirinden çok farklı ise, geleneksel eşitleme yöntemleri uygun olmayacaktır. Test formları güçlük, kapsam ve güvenilirlikleri arasında büyük farklılıklar varsa eşitleme yöntemleri işlevini yerine getiremez. Dikey ölçekleme, genellikle sınıflar arası ilerlemeyi görmek için yapılır, böylece farklı sınıf düzeylerinde elde edilen puanlar aynı ölçek üzerine

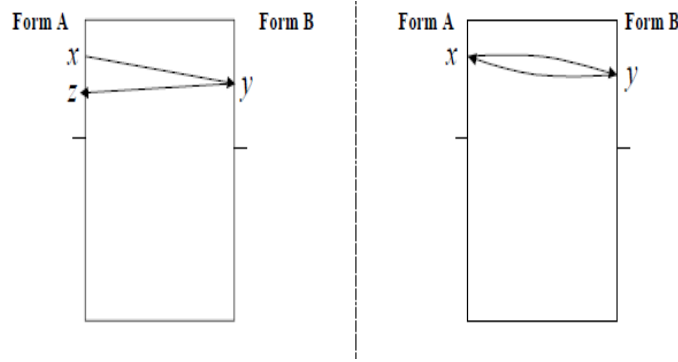
yerleştirilmiş olur. Böylece farklı seviyelerde elde edilen puanlar karşılaştırılabilir. Yatay eşitlemede eşitlenmiş puanlar birbiri yerine kullanılabilirken, dikey ölçeklemede puanlar karşılaştırılabilir, ancak birbirleri yerine kullanılamazlar (Kolen, 1988; Kolen & Brennan, 2004). Dikey ölçekleme ile bireylerin farklı düzeylerde almış oldukları puanlar karşılaştırılabilir ve bu da bireyin zaman içindeki gelişimini görmemizi sağlar. Farklı düzeylerdeki testlerin kapsamaları farklı olduğu için, aynı ölçek üzerine yerleştirilseler bile, bu test formlarından elde edilen puanlar birbiri yerine kullanılamaz. Çünkü eşitleme test formlarının kapsamaları arasındaki farkı değil, güçlükleri arasındaki farkı ayarlar (Kolen,1998). Dikey ölçekleme farklı yetenek düzeylerini karşılaştırdığı için yatay eşitlemeden daha karmaşıktır. Hem yatay eşitleme hem de dikey ölçeklemede birden fazla form gereklidir.

#### 1.7.1.1. Eşitleme Koşulları

Testlerin eşitlenmesi için yerine getirilmesi gerekli koşullarla ilgili farklı görüşler ileri sürülmüştür (Angoff, 1971; Dorans&Holland,2000; Hambleton & Swaminathan,1985; Kolen, 1988; Kolen & Brennan, 2014; Lord, 1980; Petersen, Kolen & Hoover, 1989; Zhu, 1998). Hambleton ve Swaminathan (1985) eşitleme koşullarını eşitlik, simetri, gruplar arası değişmezlik ve tek boyutluluk olarak; Lord (1980) ile Holland ve Dorans (2000) simetriklik, gruplar arası değişmezlik, aynı güvenilirliğe sahip olma, eşitlik ve aynı yapıyı ölçme; Kolen ve Brennan (2004), Petersen, Kolen ve Hoover (1989) ve Kolen (1988) eşitleme koşullarını simetriklik, gruplar arası değişmezlik, eşitlik ve aynı yapıyı ölçme olarak belirtmiştir. Bu çalışmada Petersen, Kolen ve Hoover'ın (1989) belirttiği koşullar dikkate alınmıştır.

**Simetriklik:** Eşitlenecek testlerde hangi testin referans veya temel test olduğuna bakılmaksızın eşitlemenin yapılabilmesidir (Hambleton, Swaminathan & Rogers, 1991; Lord, 1980). Simetriklik, eşitlemede X formunun Y formuna dönüştürülebilir olmasının aynı zamanda Y formunun da X formuna dönüştürülebiliyor olmasını belirtir. Örneğin, bu özelliğe göre X Formundaki 26 ham puanı, Y Formundaki 27 ham puanına karşılık geliyorsa, Y Formundaki 27 ham puanı X formundaki 26 ham puanına karşılık gelir. Eğer bu eşitleme ilişkileri grafik üzerinde gösterilirse, aynı olduğu görülür (Kolen & Brennan, 2014). Simetriklik özelliği, istatistiksel kestirim (statistical prediction) ile eşitlemeyi ayıran bir özelliktir. Kestirimde simetriklik

özelliği yokken, eşitlemede vardır (Livingston, 2004). Şekil 1.1'de kestirim ve eşitlemede simetri özelliğinin farklılığının nasıl olduğu yer almaktadır.



**Şekil 1.1. Kestirim ve Eşitlemede Simetri Özelliğinin Farklılığı** (Livingston, 2004)

**Aynı yapıyı ölçme:** Bu özellik, eşitlenecek testlerin aynı özelliği, yeteneği ve beceriyi ölçmesidir. Testlerin içerikleri ve ölçülen özellikleri farklı ise eşitlenemez. Testlerin eşitlenebilmesi için bu testlerin aynı yeteneği, özelliği ölçmesi gereklidir (Zhu, 1998). Örneğin mekanik testinden elde edilen puanlar ile dinamik testinden elde edilen puanlar eşitlenemez. Daha önce belirtildiği gibi eğer eşitlenmesi düşünülüyorsa test formları aynı içerik ve istatistiksel özelliklerle inşa edilmelidir. Aksi takdirde, hangi istatistiksel yöntem kullanılırsa kullanılsın puanlar birbiri yerine kullanılamaz (Kolen & Brennan, 2004).

**Gruplar arası değişmezlik:** Eşitleme ilişkisinin eşitleme sürecinde yer alan örneklerden bağımsız olmasına gruplar arası değişmezlik denir (Hambleton ve diğerleri, 1991). Yani eşitleme ilişkisinin tüm gruplar için aynı olmasıdır. Ayrıca eşitlemeden elde edilen dönüşümün benzer koşullarda uygulanabilmesidir (Lord, 1980; Zhu, 1998). İki test için eşitleme yapılmış ise o testi alan alt gruplar (kızlar ve erkekler) için de aynı eşitleme ilişkisi bulunacağını varsayar. Grup değişmezliği özelliği altında, eşitleme ilişkisi eşitlemenin yapıldığı sınanan grup ne olursa olsun aynıdır (Kolen & Brennan, 2004).

**Eşitlik:** Eşitlik özelliği, iki test formu eşitlendiğinde, bireylerin A veya B test formunu almasının bir fark oluşturmamasıdır (Kolen & Brennan, 2004; Lord, 1980). Her bir  $\theta$  yetenek düzeyinde, X puanlarının koşullu frekans dağılımının, Y testinin

dönüştürülmüş  $x(y)$  puanlarının koşullu frekans dağılımı ile aynı olmalıdır. X veya dönüştürülmüş  $x(y)$  puanlarının koşullu olasılık dağılımlarının varyansı sıfırdan farklı ise, bu X veya Y ölçümlerinin güvenilir olmadığını gösterir. X veya dönüştürülmüş  $x(y)$  puanlarının koşullu olasılık dağılımlarının ortalamalarının aynı olması eşitlik için yeterli değildir. Eğer test formlarının varyansı eşit değilse, bireylerin X veya Y test formunu alıp almaması bir fark oluşturmamaktadır (Lord, 1980).

Lord (1980) eşitlik koşulunun aşağıdaki özelliklere sahip olması gerektiğini belirtmiştir (akt: Hambleton & Swaminathan, 1985);

- Testler farklı yetenek veya özelliği ölçüyorsa eşitlenemez.
- Eşit olmayan güvenilirlikteki testlere ilişkin ham puanlar eşitlenemez.
- Farklı güçlük düzeylerindeki testlere ait ham puanlar eşitlenemez. Bu durumda gerçek puanlar doğrusal olmayan ilişkiye sahiptir ve bu nedenle farklı yetenek düzeylerindeki testler güvenle eşitlenemez.
- X testi puanlarının  $\theta$  yetenek düzeyindeki koşullu frekans dağılımları, Y testi puanlarının dönüştürülmüş koşullu frekans dağılımları ile aynı olmalıdır.
- X ve Y testi tam anlamıyla paralel olmadıkça eşitlenemez. Çünkü koşullu frekans dağılımının aynı olması, iki dağılımın aynı olması anlamına gelir.
- Güvenilir testler eşitlenebilir.

Bu koşulun gereklilikleri katı gibi gözükse de test puanlarının eşitlenmesi için gerekli koşullardan yalnızca biridir (Hambleton & Swaminathan, 1985). Morris (1982), Lord'un eşitlik özelliğinin daha az kısıtlayıcı olan I. sıra ve II. sıra eşitlik özelliğini önermiştir (akt: Kolen & Brennan, 2014).

Eşitlemeyle ilgili bu gerekliliklerin özellikle eşitlik koşulunun sağlanması klasik eşitleme metotlarında oldukça zordur. Teorik olarak MTK, geleneksel eşitleme metotlarındaki bu problemlerin üstesinden gelmektedir (Hambleton ve diğerleri, 1991). Eşitleme için gerekli olan koşullar, eşitlemenin farklı aşamalarına işaret etmektedir. Eşitlik, sınava giren adayları bireysel olarak ifade ederken, eşit güvenilirlik ve gruplar arası değişmezlik ise örneklem büyüklüğü açısından ifade edilmektedir. Ayrıca simetri özelliği, matematiksel bir özellik iken, aynı yapıyı ölçme ise testlerin doğası ve olası kullanımları hakkındadır (von Davier ve diğerleri, 2004).

Kolen ve Brennan'a (2014) göre test eşitleme, test geliştirme, uygulama ve istatistiksel süreçlerinin koordineli bir biçimde olmasını gerektirir. Uygulamada aşamalar farklılaşsa da eşitleme basamakları aşağıdaki gibidir.

- Eşitlemenin amacına karar vermek
- Alternatif formlar oluşturmak (Aynı kapsam ve istatistiksel özelliklere sahip)
- Veri toplama desenini seçmek
- Veri toplama desenini uygulamak
- Eşitlemenin bir ya da daha fazla operasyonel tanımını yapmak
- Bir ya da daha fazla istatistiksel kestirim metodunu seçmek
- Eşitleme sonucunu değerlendirmek

#### 1.7.1.2. Eşitleme Desenleri

Eşitleme için farklı veri toplama desenleri oluşturulabilir. Eşitleme için kullanılacak desen hem pratiklik hem de istatistiksel açıdan uygun olmalıdır (Kolen & Brennan, 2014). Bu desenlerden alanyazında sıklıkla kullanılan tek grup, random grup ve denk olmayan gruplarda ortak test deseni (Kolen & Brennan, 2004; Tsai, 1997) hakkında bilgi verilmiştir. Her bir desenin avantajları ve sınırlılıkları vardır.

##### 1.7.1.2.1. Tek grup deseni

Testleri ilişkilendirme ve eşitleme için kullanılan en eski ve en basit eşitleme deseni olan tek grup deseninde aynı bireyler iki veya daha fazla test formunu alırlar. Bu desende testi alan gruptan elde edilen eşitleme ilişkisinin hedef evrene genellenebileceğini ve bu grubun hedef evreni temsil etmesine gerek olmadığı belirtilmektedir. Tek grup deseninde teste yeni soru eklendiğinde veya çıkarıldığında oluşturulan yeni test eşitleme için kullanılabilir (Livingston, 2004). Tek grup deseni Tablo 1.1'de gösterilmiştir.

**Tablo 1.1: Tek Grup Deseni**

Grup	Form A	Form B
X	✓	✓

Tek grup deseninde her iki test de aynı bireylere uygulandığından, testi alan bireylerin değişen yeterlilik düzeyi kontrol altına alınmış olur. Aynı bireyler iki test formunu aldıkları için yorgunluk ve pratik etkisi eşitleme sürecini etkileyebilir. Bunun için test formlarının sırası dengelenir. Bu problemi ortadan kaldırmak için

grup ikiye ayrılır, her iki grup eşitlenmek istenen iki formu da alır. Birinci grup ilk önce I. formu, daha sonra II. formu alırken; ikinci grup ilk olarak II. formu daha sonra I. formu alır. Aynı gruba iki test formu da uygulandığından küçük örneklem ile eşitleme kesinliği sağlanabilir. Aynı bireylere iki test formunun uygulanması ve tek oturumda birden fazla formun uygulanması da her zaman olası değildir (Kolen, 1988; Kolen & Brennan, 2004).

#### 1.7.1.2.2. Random grup deseni

Random grup deseninde bireyler test formlarına random olarak atanır ve iki grup farklı test formlarını alır (Cook & Eignor, 1991). Bu desende sıklıkla kullanılan döngüsel süreç X ve Y formları alternatifleri olacak şekilde ayrı ayrı paketlenmesidir. Böylece ilk öğrenci Form X'i, ikinci öğrenci Form Y'yi ve üçüncü öğrenci tekrardan Form X'i alır ve süreç böyle devam eder. Bu süreç random eşit gruplar oluşmasına olanak verir. Random grup deseni Tablo 1.2'de gösterilmiştir.

**Tablo 1.2: Random Grup Deseni**

Grup	Form A1	Form A2	FormB1	FormB2
X	✓			✓
Y		✓	✓	

Bu desende bireyler sadece tek bir form aldıkları için daha fazla tercih edilir ve farklı bireyler iki veya daha fazla formu aldıkları için, tek grup desenine göre daha geniş örneklem gerektirir (Kolen & Brennan, 2004; Kolen, 2007). Ancak gruplar aynı olmadığı için bireylerin yetenek dağılımları arasındaki farklılık eşitlemeye yanlılık olarak karışır (Hambleton & Swaminathan, 1985; Petersen ve diğerleri, 1989).

#### 1.7.1.2.3. Denk olmayan gruplarda ortak madde deseni

Denk olmayan gruplarda ortak madde deseni, test güvenliğinden dolayı aynı gün birden fazla formun uygulandığı durumlarda kullanılır (Tsai, 1997). Bu desen denk olmayan iki gruba uygulandığı için denk olmayan gruplarda ortak test veya ortak test deseni (NEAT) olarak adlandırılır (von Davier ve diğerleri, 2004). Alanyazında en çok kullanılan desendir. Bunun en önemli nedeni her bir gruba sadece tek bir

formun bir kez uygulanmasıdır (Kolen, 1988). NEAT deseni Tablo 1.3'de gösterilmiştir.

**Tablo 1.3: Denk Olmayan Gruplarda Ortak Madde Deseni**

Grup	Form A1	Form A2	Ortak Test
X	✓		✓
Y		✓	✓

Denk olmayan iki gruba uygulanan test formları eşitlenirse, test formlarının ortalamaları arasında gözlenen farkın test formlarının güçlükleri arasındaki farktan mı, grupların yetenek düzeyleri arasındaki farktan mı yoksa bu iki faktörün kombinasyonundan mı kaynaklandığı bilinemez. Bunun için denk olmayan gruplar test formlarından birini ve ortak madde setini alırlar. Test formlarındaki bu ortak maddeler üzerinden de eşitleme ilişkisi kurulur. Bu desende ortak maddeler testi iyi yansıtmalı ve maddenin bulunduğu yer maddenin güçlüğünü etkileyebileceğinden her iki formda da aynı yerde yer almalıdır (Kolen, 1988; Kolen & Brennan, 2004; Tsai, 1997). Ortak test formu iç ortak ve dış ortak olmak üzere ikiye ayrılır. Eğer ortak maddelerden elde edilen puanlar bireyin toplam test puanına eklenecekse iç ortak, toplam puanına eklenmeyecekse dış ortak olur. Dış ortak test bireye ayrı bir form halinde verilirken, iç ortakta testin içinde yer alır (Cook & Eignor, 1991; Kolen, 1988; Zhu, 1998). NEAT deseninde; geleneksel eşitleme metotları iki test formundaki ortak maddeler iki grubun yetenek dağılımları arasındaki farkı ayarlar. MTK gerçek puan eşitleme metotlarında ise ortak maddeler madde parametre kestirimlerini aynı ölçek üzerine yerleştirmek için kullanılır (Cook & Eignor, 1991).

### **1.7.2. Eşitleme Metotları**

Eşitleme metotları gerçek ve gözlenen puan eşitleme olmak üzere ikiye ayrılır. Hem gözlenen hem de gerçek puan eşitleme metotları test formlarının güçlükleri arasındaki küçük farkları dikkate alarak eşitleme yapar. Bu çalışmada Kernel eşitleme ile MTK gerçek puan eşitleme yöntemleri karşılaştırıldığından, bu iki yaklaşım hakkında bilgi verilmiştir.



### 1.7.2.1. Kernel Eşitleme

Kernel eşitleme, kesikli puan dağılımlarının sürekli dağılımlara dönüştürerek puan dağılımlarının eşitlendiği bir eşit yüzdellikli gözlenen puan eşitleme yöntemidir (von Davier ve diğerleri, 2006). Eşit yüzdellikli eşitlemede, aynı yüzdellik sırasına denk gelen puanların eşit olduğu kabul edilir. Bunun için ilk olarak her bir formun yığılmalı frekansı hesaplanarak tablollaştırılır ve bu yığılmalı frekanslara göre aynı yüzdellik puanlara karşılık gelen puanlar eşitlenir. Eşit yüzdellikli eşitlemede, aynı yüzdellik sırasındaki eşitlenmiş puanlara sahip bireylerin aynı yetenek düzeyinde olduğu kabul edilir (Kolen, 1988). Ancak gerçek uygulamalarda bu bireylerin aynı yetenek düzeyinde olması oldukça güçtür. Kernel eşitleme, eşit yüzdellikli eşitlemede ortaya çıkan bu problemi çözmek için geliştirilmiştir. Bu problemin ortaya çıkmasının nedeni ise puan dağılımlarının kesikli olmasıdır. Holland ve Thayer (1981) bu probleme, kesikli dağılımları Kernel eşitleme ile süreklileştirerek çözüm getirmiştir. Kernel eşitlemede, kesikli dağılımlar sürekli hale getirilerek, sürekli dağılımlar üzerinden puanlar eşitlenir (Livingston, 1993; Ricker & von Davier, 2007).

Kernel eşitleme, geleneksel eşitleme metotlarının gelişmiş bir şekli olarak görülebilir. Bunun iki temel nedeni bulunmaktadır. İlki, ön düzgünleştirmenin kullanılması verileri tutarlı hale getirmesinin yanında verilerin dönüşümünü de düzgünleştirerek diğer yöntemlere göre daha küçük standart hata vermesi ve örneklem değişkenliğine daha az bağımlı olması; ikincisi ise, Kernel eşitlemenin tüm desenlerde ve eşitleme fonksiyonlarına uygulanabilir olmasıdır (von Davier ve diğerleri, 2004).

Kernel eşitleme 5 basamaktan oluşur. Bu basamaklar hem genel olarak Kernel eşitleme için hem de Kernel eşitleme yöntemlerinden zincirleme ve son tabakalama eşitleme yöntemleri için tanımlanmıştır.

**1) Ön düzgünleştirme:** Bu aşamada puan olasılıkları eşitleme desenine göre, ham veriye uygun istatistiksel modellerle elde edilir. İstatistiksel modeller, kullanılan eşitleme desenine göre farklılık göstermektedir. Öndüzgünleştirme, tek değişkenli ve iki değişkenli dağılımlarda farklılık göstermektedir. NEAT deseninde iki test kullanıldığı için iki değişkenli bir dağılımdır. Holland ve Thayer (2000) iki değişkenli puan dağılımların öndüzgünleştirmesi için log lineer modelleri önermiştir. Log lineer modeller, birçok farklı puan dağılımlarının özelliklerine uyum

sağlayabilmektedir (von Davier ve diğerleri, 2006). Log lineer modeller, etkili iteratif teknikleri kullanarak en çok olabilirlik (Maximum Likelihood) yöntemiyle kestirilebilirler (von Davier ve diğerleri, 2004). Kernel eşitlemede öndüzgünleştirmenin yapılmasının iki amacı vardır. İki eşitleme süreci için gerekli olan düzgünleştirilmiş puan dağılımının elde edilmesi, ikincisi ise eşitlemenin standart hatasının hesaplanabilmesi için matris üretmesidir (von Davier ve diğerleri, 2006).

NEAT deseninde yer alan iki gruptan P grubu X test formu ile A ortak test formunu alırken, diğer grup olan Q grubu Y test formu ile A ortak test formunu alırlar. Log lineer modeller, iki değişkenli dağılıma (P grubu için (X+A, A)) uygulandığında aşağıdaki eşitlik elde edilir. Aynı eşitlik Q grubu için de oluşturulur.

$$\log_e(p_{jl}) = \alpha + \sum_{i=1}^{T_{XP}} \beta_{XP_i} (x_j)^i + \sum_{i=1}^{T_{AP}} \beta_{AP_i} (a_l)^i + \sum_{i=1}^{I_{XP}} \sum_{i'=1}^{I_{AP}} \beta_{XAP_{i i'}} (x_j)^i (a_l)^{i'} \quad (\text{Eşitlik 1.1})$$

$x_j$ ; X+A testinden alınan puan ve  $a_l$  A testinden alınan puan olmak üzere  $p_{jl}$  ise  $x_j$ ,  $a_l$ 'nin puanının joint (bileşik) puan olasılığıdır.

$T_{XP}$ ; X+A'nın marjinal (tek değişkenli) dağılımının momenti

$T_{AP}$ ; A'nın marjinal (tek değişkenli) dağılımının momenti

$I_{XP}, I_{AP}$  ; ise (X+A, A)'nın iki değişkenli dağılımda kesişen momenti

von Davier ve diğerleri (2004) puan olasılıklarının kestirilmesinin seçiminde 4 istatistiksel özelliğın olduğunu belirtmişlerdir. Bunlar;

- a) **Tutarlılık (consistency):** Tutarlılık, örneklem büyüklüğü arttıkça, kestirilen değerlerin evren parametresine yaklaşmasıdır.
- b) **Yeterlik (efficiency):** Evren değerlerinden kestirilen puan olasılıklarının sapmasının mümkün olduğu kadar küçük olmasıdır. Yani, örnekleme ait verinin mümkün olduğu kadar evren verisine yakın olmalıdır.
- c) **Pozitiflik:** Her bir puan için kestirilen puan olasılıklarının pozitif olmasıdır. Ya da her bir puanın puan olasılığının sıfırdan farklı ve pozitif olmasıdır.

olmasıdır. Ancak NEAT deseninde iç ortak testin olduğu ve bazı tek grup deseni uygulamalarında istisnai durumlar mevcuttur.

- d) **Bütünlük (integrity):** Düzgünleştirilmiş puan dağılımlarının gözlenen puan dağılımlarıyla eşleşmesidir. Bu özellik doğrusal eşitlemede ortalama ve varyansın eşleştirilmesiyle yakından ilişkilidir ve ayrıca eşit yüzdeli metotlarda da istenilen bir özelliktir. Von Davier ve diğerleri (2004), tek değişkenli dağılımlarda iyi uyum elde etmek için test formlarının beş ya da altı momentinin kullanılmasını önermişlerdir.

Kernel eşitlemede NEAT deseninde iç ortak ve dış ortak teste göre diğer aşamalar aynıyken, öndüzgünleştirme aşaması farklılaşmaktadır. Bu durumu tek bir test formu üzerinde bir örnekle açıklayacak olursak, toplam test  $X$ , ortak test formu  $A$  ve  $X^*$  ise ortak olmayan test formunun bölümünü göstermek üzere,  $X=X^*+A$  olur.  $A$  formundan 5 puan alındığında,  $X^*$  hiçbir zaman negatif olamayacağından toplam puan her zaman 5'e eşit veya 5'den büyük ( $X \geq 5$ ) olur. Yani iç ortak test kullanıldığında, toplam test puanı iç ortak testten elde edilen puandan asla daha küçük olamayacağından yapısal sıfırlar (structural zeroes) durumunun ortaya çıkmasına neden olur (Holland & Thayer, 2000; von Davier, Holland ve Thayer, 2004; s. 41). Log lineer model, yapısal sıfırların olduğu iki değişkenli dağılımlarda ( $X+A, A$ ) da uygundur (Holland & Thayer, 2000). Dış ortak test formunda bu durum ortaya çıkmaz.

Diğer aşamalar zincirleme ve son tabakalama eşitleme için farklılaşırken, öndüzgünleştirme aşaması her iki yöntem için de aynıdır (von Davier ve diğerleri, 2004).

**2) Puan Dağılımlarının Kestirilmesi:** İkinci aşamada, birinci aşamada elde edilen düzgünleştirilmiş puan dağılımları, sentetik evren üzerinde puan olasılık dağılımlarını kestirmek için kullanılır.

Puan olasılık dağılımları  $r$  ve  $s$  olmak üzere,

$$r_j = P\{X = x_j|T\}, \quad s_k = P\{Y = y_k|T\}$$

(Eşitlik 1.2)

$X$  ham puanlarla ilişkili  $X$  puan olasılık dağılımı  $\{x_j\}$  ve  $Y$  ham puanlara dayalı  $Y$  için puan olasılık dağılımı  $\{y_k\}$

$$r = (r_1, \dots, r_J), \text{ and } s = (s_1, \dots, s_K)$$

(Eşitlik 1.3)

Sentetik evren için kestirilen puan dağılımları kullanılan eşitleme desenine göre farklılık göstermektedir.

**Zincirleme eşitleme için;** öncelikle iki ayrı tek grup deseni oluşturulur. Zincirleme eşitlemede, X formundan ortak test formuna (A) bağlanılır ve sonra (A) ortak test formundan Y formuna bağlantı kurulur. Zincirleme eşitlemede ortak test formu ile test formları arasında eşitleme yerine bağlanma olur. Bunun nedeni ise, eşitlemenin varsayımlarından olan formların güvenilirliklerinin eşit olma varsayımının genellikle sağlanmamasıdır. Daha sonra bu bağlanma fonksiyonları eşitleme için kullanılır (von Davier ve diğerleri, 2004).

X ve Y'ye ait puan dağılımları r ve s ile gösterilmek üzere ve desen fonksiyonu da DF ile gösterilirse,

P: P popülasyonu için iki değişkenli puan olasılık matrisi

Q: Q popülasyonu için iki değişkenli puan olasılık matrisi

M ve N: sıfır matrisi

V(P) ve V(Q): P ve Q nun vektör versiyonu ise,

$$\begin{pmatrix} r_P \\ t_P \end{pmatrix} = DF_P(P) = \begin{pmatrix} M_P \\ N_P \end{pmatrix} v(P)$$

ve

$$\begin{pmatrix} r_Q \\ t_Q \end{pmatrix} = DF_Q(Q) = \begin{pmatrix} M_Q \\ N_Q \end{pmatrix} v(Q)$$

Bu iki tek grup desenine ait fonksiyonlar bir araya getirilirse,

$$\begin{pmatrix} r_Q \\ t_Q \\ r_P \\ t_P \end{pmatrix} = DF(Q, P) = \begin{pmatrix} DF_P(P) \\ DF_Q(Q) \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} M_P \\ N_P \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} M_Q \\ N_Q \end{pmatrix} \end{pmatrix} \begin{pmatrix} v(P) \\ v(Q) \end{pmatrix}$$

(Eşitlik 1.4)

Zincirleme eşitlemede 4 adet yığılmalı dağılım fonksiyonu (cdfs) vardır. Bunların iki tanesi X ve Y formuna, diğer iki tanesi de her bir grup için ortak test formuna aittir.

$F_P$  X formuna ait cdfs ve  $H_P$  ise P grubu için ortak teste ait cdfs

$G_Q$ ,  $X$  formuna ait cdfs ve  $H_Q$   $Q$  grubu için ortak teste ait cdfs

Zincirleme eşitleme fonksiyonlarının ( $F_P, H_P, G_Q$  ve  $H_Q$ ) hesaplanabilmesi için, yığılmalı dağılım fonksiyonlarına ait puan olasılıklarının vektörlerinin kestirilmesi gereklidir. Puan dağılımlarına ilişkin vektörler,

$$\begin{aligned} r_P &= (r_{Pj}) \\ t_P &= (t_{Pl}) \\ t_Q &= (t_{Ql}) \\ s_Q &= (s_{Qk}) \end{aligned}$$

(Eşitlik 1.5)

$j=1, \dots, J, k=1, \dots, K, l=1, \dots, L$  ve  $J, K, L$  test formlarından alınan olası puanları göstermek üzere; eşitlikler yardımıyla her bir test için iki gruba ait marjinal olasılıklar hesaplanır.

$$\begin{aligned} r_{Pj} &= Prob\{X = x_j | P\} = \sum_j p_{jl} \\ t_{Pl} &= Prop\{A = a_l | P\} = \sum_j p_{jl} \\ t_{Ql} &= Prob\{A = x_j | Q\} = \sum_j q_{kl} \\ s_{Qk} &= Prob\{Y = y_k | Q\} = \sum_j q_{kl} \end{aligned}$$

(Eşitlik 1.6)

**Son Tabakalama eşitleme için;** son tabakalama eşitleme de, öndüzgünleştirme yapıldıktan sonra  $P$  ve  $Q$  grupları ağırlıklandırılması olan sentetik evren ( $T$ ) üzerinde  $X$  ve  $Y$  formu için marjinal dağılımlar ( $r$  ve  $s$ ) kestirilir.

$$T = w_1 P + (1 - w_1) Q$$

(Eşitlik 1.7)

Bu ağırlıklandırma katsayısı 0-1 arasında değer alır ve toplamı  $w_2 + w_1 = 1$ 'e eşittir. Zincirleme eşitlemede ağırlıklandırmanın seçimi eşitleme sonuçlarını etkilemezken, son tabakalama eşitlemede ağırlıklandırma katsayılarının seçimi eşitleme fonksiyonlarını etkileyebilmektedir.

$p_l$  ve  $q_l$  vektörleri  $l = 1 \dots L$  için sırasıyla  $P$  ve  $Q$ 'nun  $l^{th}$  sütunları olmak üzere; marjinal dağılımlar ( $r$  ve  $s$ ) aşağıdaki eşitlikler yardımıyla elde edilir.

$$\begin{aligned} \begin{pmatrix} r \\ s \end{pmatrix} &= DF(P, Q; w) = \begin{pmatrix} r(P, Q, w) \\ s(P, Q, w) \end{pmatrix} v(P) \\ r(v(P), v(Q), w) &= \sum_l \left[ w + \frac{(1-w)(t_{Ql})}{t_{Pl}} \right] p_l \\ s(v(P), v(Q), w) &= \sum_l \left[ (1-w) + \frac{w(t_{Pl})}{t_{Ql}} \right] q_l \end{aligned}$$

(Eşitlik 1.8)

**3) Süreklileştirme:** Üçüncü aşama olan süreklileştirme, sadece Kernel eşitlemede değil, bütün eşit yüzdelli eşitleme metotlarında kullanılmaktadır. Süreklileştirme, sentetik evren üzerindeki X ve Y'nin kesikli puan dağılımlarının tüm puan ölçeği boyunca sürekli puan dağılıma dönüştürülmesidir. Kesikli puan dağılım olasılıklarının sürekli normal dağılım olasılığına dönüştürülmesi yaklaşımına benzerdir. Diğer eşit yüzdelli eşitleme metotları yüzdellik sıralara dayalı olarak parçalı doğrusal yığılmalı (kümülatif) frekansları kullanır (von Davier ve diğerleri, 2006). Kernel eşitleme ise kesikli dağılımları süreklileştirmek için adını aldığı Gauss Kernel düzgünleştirme kullanılır. Kernel eşitlemede X ve Y için kesikli yığılmalı dağılım fonksiyonları,

$$\begin{aligned} F(x) &= Prob(X \leq x) = \sum_{j, x_j \leq x} r_j \\ G(y) &= Prob(Y \leq y) = \sum_{k, y_k \leq y} s_k \end{aligned}$$

(Eşitlik 1.9)

X kesikli dağılıma sahip olduğu için, F(x)'in grafiği her bir x değeri için düz (flat) olacak ve aynı durum Y içinde geçerlidir. Bu durumda u'nun aynı (0,1) değeri için; F(x) =u ve G(y) =u durumu pratikte meydana gelmez. Ancak eğer F(x) ve G(y) sürekli olursa, bütün x ve y değerler için bu durum oluşmaz.

$$Y = G^{-1}(F(x))$$

(Eşitlik 1.10)

y=G(y) eşit olduğu durumda, G<sup>-1</sup> (G'nin tersi olmak üzere) y= G<sup>-1</sup> (u) olur.

Phi(z) normal veya gauss dağılımı olmak üzere, X dağılımı için Gauss Kernel düzgünleştirme yığılmalı dağılım fonksiyonu (cdf's)

$$F_{hx}(x) = \sum r_j \varphi(R_{jX}(x))$$

ve  $R_{jX}$  ise,

$$R_{jX}(x) = \frac{x - (1 - a_x)\mu_x - a_x x_j}{a_x h_x}$$

$$a_x^2 = \frac{\sigma_x^2}{h_x^2 + \sigma_x^2}$$

(Eşitlik 1.11)

$\mu_x$  ve  $\sigma_x^2$  (ortalama ve varyansı)  $X$  in ortalama ve varyansı gösterir.

$h_x$  sabiti, X formu için  $F_{hx}$  için bant genişliğidir (von Davier ve diğerleri, 2004) ve düzgünleştirme parametresi olarak düşünülebilir (Lee, 2007). Pozitif her  $h_x$  değeri için,  $F_{hx}$  sürekli yığılmalı dağılım fonksiyonu (cdf's) kabul edilir.  $F_{hx}(x)$ , yoğunluk fonksiyonu, normal yığılmalı dağılım fonksiyonunun ortalaması ile ağırlıklandırılarak elde edilir.

$$f_{hx}(x) = \sum_j r_j \varphi(R_{jX}(x)) \frac{1}{a_x h_x}$$

(Eşitlik 1.12)

Aynı yolla,  $G_{hy}$  kullanılarak Y de süreklileştirilir.

Kernel süreklileştirmenin asıl özelliği ise, bant genişliğinin (h parametresi) seçimidir. Çünkü h parametresi Kernel eşitleme sonuçlarını etkileyebilmektedir. Kernel eşitlemede h parametresinin seçiminde birçok farklı yöntem bulunmaktadır. Bu yöntemlerin temel amacı ortalama hatalar karesini (MSE mean squared error) azaltmaktır (Ferrando, 2004; Lee, 2007). h parametresi pozitif değer almalıdır (h > 0). NEAT deseninde iki test formu olduğu için iki h parametresi vardır (Ferrando, 2004).

von Davier ve diğerleri (2004; 2006), ideal bant genişliğinin seçiminde ceza (penalty) fonksiyonunun kullanılmasını önermişlerdir.

Ceza fonksiyonu aşağıdaki gibidir.

$$\begin{aligned} \text{Ceza}_1(h) &= \sum_j \left[ \left( \frac{r_j}{d_j} \right) - f_T(x_j; h) \right]^2 \\ \text{Ceza}_2(h) &= \sum_j A_j (1 - B_j) \end{aligned}$$

(Eşitlik 1.13)

$d_j$ ;  $x_j$  puanı ile ilişkilendirilmiş aralık genişliğidir ve genellikle 1 kabul edilir.

$x$ 'e göre  $f_T(x; h)$ 'un türevi sıfırdan küçük,  $x_j$  nin solundaysa  $A_j=1$  ve  $x_j$ 'nin sağındaysa  $f_T(x; h)$ 'un türevi sıfırdan büyük ve  $B_j=0$  dir.

Ceza1 ve Ceza2, D gibi bir katsayıyla ağırlıklandırılarak birleştirilerek aşağıdaki eşitlik elde edilir.

$$\text{Ceza}_1(h_x) + D * \text{Ceza}_2(h_x)$$

(Eşitlik 1.14)

Eğer dağılımda düzgünleştirilmesi gereken boşluklar varsa, D ağırlıklandırma katsayısının 1 olarak kullanılması önerilmektedir (von Davier ve diğerleri, 2006). D değerleri büyüdükçe, yoğunluk fonksiyonundan elde edilen modlar azalmış olur (von Davier ve diğerleri, 2004). X formu için hesaplanan ceza fonksiyonları Y formu için de aynı şekilde hesaplanmaktadır.

**Zincirleme eşitleme için;** X, Y ve ortak test formlarına ait yığılmalı dağılım fonksiyonlarının da süreklileştirilmesi gereklidir. Süreklileştirilen her bir yığılmalı dağılım fonksiyonu aynı yolla elde edildiğinden sadece X formu burada tanımlanmaktadır.

$$\begin{aligned} \hat{F}_{Ph_{XP}}(x) &= \sum_j \hat{r}_j \varphi \left( R_{jXP}(x) \right) \\ \hat{R}_{jXP}(x) &= \frac{x - \hat{a}_{XP}x_j - (1 - \hat{a}_{XP}) - \hat{\mu}_{XP}}{\hat{a}_{XP}h_{XP}} \\ \hat{a}_{XP}^2 &= \frac{\hat{\sigma}_{XP}^2}{\hat{\sigma}_{XP}^2 + h_{XP}^2} \end{aligned}$$

(Eşitlik 1.15)

Bant genişliği için ceza fonksiyonları hesaplanır.



$$Ceza_1(h_{XP}) + D \times Ceza_2(h_{XP}),$$

(Eşitlik 1.16)

Aynı şekilde  $H_P$ ,  $H_Q$  ve  $G_Q$  yığılmalı dağılım fonksiyonları hesaplanır. Bu fonksiyonlar hesaplandıktan sonra eşitleme aşamasına geçilir.

**Son Tabakalama eşitleme için;** Yığılmalı dağılım fonksiyonları X ve Y formlarının puan olasılıkları ile ilişkilendirilir.

$$\hat{F}_w(x) = Prob\{X \leq x_j\} = \sum_{x_j \leq x} \hat{r}_{wj}$$

$$\hat{G}_w(Y) = Prob\{Y \leq y_k\} = \sum_{y_j \leq y} \hat{s}_{wk}$$

(Eşitlik 1.17)

**4) Eşitleme:** İlk üç aşamadaki adımlar tamamlandıktan sonra, Kernel eşitleme için eşitleme fonksiyonları hesaplanmaktadır.

X formunun Y formuna eşitlenmesi için Kernel eşitleme fonksiyonu,

$$\begin{aligned} \hat{e}_Y(x) &= G_{h_Y}^{-1}(F_{h_X}(x; \hat{r}); \hat{s}) \\ &= G_{h_Y}^{-1}(F_{h_X}(x_j)) \end{aligned}$$

(Eşitlik 1.18)

Y formunun X formuna eşitlenmesi için Kernel eşitleme fonksiyonu,

$$\begin{aligned} \hat{e}_X(y) &= F_{h_X}^{-1}(G_{h_Y}(y; \hat{s}); \hat{r}) \\ &= F_{h_X}^{-1}(G_{h_Y}(y_k)) \end{aligned}$$

(Eşitlik 1.19)

**Zincirleme eşitleme için;**  $e_{Y(CE)}$  zincirleme eşitleme fonksiyonu yukarıda tanımlanan dört süreklileştirilmiş yığılmalı dağılım fonksiyonu uygulandığında eşitleme fonksiyonu aşağıdaki eşitlikteki gibi yazılır.

$$\hat{e}_{Y(CE)}(x) = \hat{G}_{QhYQ}^{-1} \left( \hat{H}_{QhAQ} \left( \hat{H}_{PhAP}^{-1} \left( \hat{F}_{PhXP}(x) \right) \right) \right)$$

(Eşitlik 1.20)

$e_Y(CE)$  zincirleme eşitleme fonksiyonu P ve Q grupları için ayrı ayrı tanımlandığında;

$$\hat{e}_A(x) = \hat{H}_{PhAP}^{-1} \left( \hat{F}_{PhXP}(x) \right)$$

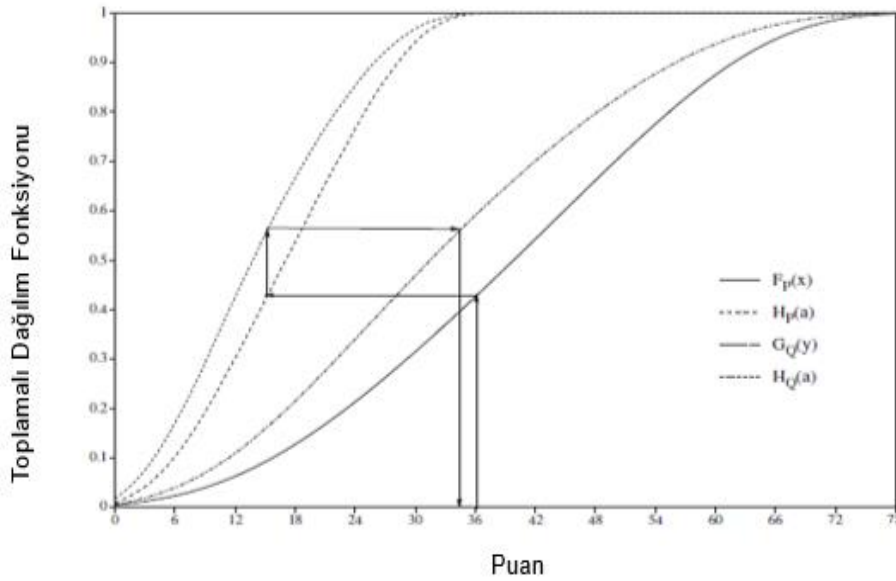
$$\hat{e}_Y(a) = \hat{G}_{QhYQ}^{-1} \left( \hat{H}_{QhAQ}(a) \right)$$

(Eşitlik 1.21)

Zincirleme eşitleme fonksiyonunu elde edilir.

$$\hat{e}_{Y(CE)}(x) = \hat{e}_Y(\hat{e}_A(x))$$

(Eşitlik 1.22)



**Şekil 1.2. Zincirleme Eşitlemenin Grafiks gösterimi** (von Davier ve diğerleri, 2004)

Yukarıdaki şekilde zincirleme eşitleme sürecinin nasıl gerçekleştiği grafiks olarak görülmektedir. Şekil incelendiğinde, belirli bir X puanı için önce  $F_P(x)$  de karşılık geldiği nokta bulunur, sonra bu noktaya karşılık  $H_P(a)$  noktası belirlenir. Belirlenen  $H_P(a)$  noktasından dikey olarak  $H_Q(a)$ 'da karşılık gelen nokta belirlenir ve buradan yatay olarak  $G_Q(y)$  eğrisine gidilir.  $G_Q(y)$  noktasından puan çizelgesine denk gelen

Y puanı elde edilmiş olur. Yani, X puanının zincirleme eşitleme sonucunda denk gelen Y puanı belirlenmiş olur (von Davier ve diğerleri, 2004).

Bu eşitleme fonksiyonlarında, bant genişlikleri her iki form için de yeteri kadar geniş olduğunda, Kernel eşitleme doğrusal eşitleme metotlarının özel bir hali olarak düşünülebilir.

## 5) Eşitlemenin Standart Hatasının Hesaplanması

Kernel eşitlemede hem eşitlemenin standart hatası (SEE) hem de eşitlemenin standart hatası farkı (SEED) hesaplanabilmektedir. Eşitlemenin standart hatası 3 faktöre bağlıdır. İlki, ön düzgünleştirme, ikincisi düzgünleştirilmiş veriden r ve s'nin hesaplanması ve üçüncüsü ise süreklileştirme ve eşitleme fonksiyonunun matematiksel formunun kombinasyonudur (von Davier ve diğerleri, 2006).

Kernel eşitlemede, eşitleme desenine bağlı olarak standart hata hesaplanması farklılaşmaktadır. Standart hata hesaplanırken Taylor açılımı veya delta metodu kullanılmaktadır (Mao, van Davier ve Rupp, 2006). Eşitlemenin standart hata ve standart hatası farkı, Kernel eşitlemenin etkililiğini belirlemek için kullanılır. Eşitlemenin standart hatası, geniş örneklemelere dayalıdır ve örneklem varyansını ölçmektedir (von Davier ve diğerleri, 2004). X'in Y ye eşitlenmesindeki eşitlemenin standart hatası;

$$SEE_Y(x) = \hat{\sigma}_y(x) = \sqrt{Var(\hat{e}_Y(x))}$$

(Eşitlik 1.23)

Aynı şekilde Y'nin X'e eşitlenmesinde eşitlemenin standart hatası

$$SEE_X(y) = \hat{\sigma}_x(y) = \sqrt{Var(\hat{e}_X(y))}$$

(Eşitlik 1.24)

Eşitlemenin standart hatası farkı (SEED), Kernel eşitleme yöntemlerinin karşılaştırılmasında kullanılır. İki Kernel eşitleme yöntemine ait eşitlemenin standart hataları arasındaki farktır (von Davier ve diğerleri, 2004).

$$SEED_Y(x) = \sqrt{Var(\hat{e}_1(x) - \hat{e}_2(x))}$$

(Eşitlik 1.25)

**Zincirleme Eşitleme için;** X formundan ortak test formuna (A) ve ortak test formundan (A), Y formuna olmak üzere iki tane bağlama vardır ve standart hatanın hesaplanması için iki ayrı standart hatanın hesaplanması gerekmektedir. X formundan A ortak test formuna bağlama için standart hata  $SEE_A(x)$  ve A ortak test formundan Y formuna bağlama için standart hata  $SEE_B(x)$  olarak tanımlanırsa, Zincirleme eşitleme için standart hata aşağıdaki eşitlikle hesaplanır.

$$[SEE_{Y(CE)}(x)]^2 = [SEE_{Y(CE)}(e_A(x))]^2 + [e'_Y(e_A(x))SEE_A(x)]^2$$

(Eşitlik 1.26)

$e_A(x)$  : X formundan A ortak test formuna bağlama

$e_Y(a)$ : A ortak test formundan Y formuna bağlama

**Son tabakalama eşitleme için;** son tabakalama eşitlemede standart hata aşağıdaki eşitlikle elde edilir.

$$U_r = wU_p + (1-w)U_p^* - (1-w) \sum_l \left( t_{Pl}^{Ql} / t_{Pl} \right) t_{Pl}^{-1} p_l v_{Pl}^t$$

$$V_r = w \sum_l t_{Ql}^{-1} q_l v_{Pl}^t$$

$$U_s = (1-w) \sum_l t_{Pl}^{-1} q_{Pl} v_{Ql}^t$$

$$V_s = (1-w)U_Q + wU_Q^* - w \sum_l \left( t_{Pl}^{Pl} / t_{Ql} \right) t_{Ql}^{-1} q_l v_{Ql}^t$$

olmak üzere;

$$SEE_Y(x) = \frac{1}{G'} \left[ \left\| \frac{\partial F}{\partial r} U_R - \frac{\partial G}{\partial s} V_R \right\|^2 + \left\| \frac{\partial F}{\partial r} U_S - \frac{\partial G}{\partial s} V_S \right\|^2 \right]^{1/2}$$

(Eşitlik 1.27)

Son tabakalama ve zincirleme eşitleme için elde edilen standart hata değerleri birbirine oldukça benzer ve aynı şekildedir. Son tabakalama eşitlemenin bazı puan aralıklarında zincirleme eşitlemeye göre daha az düşük standart değerlere sahip olsa da, bu durum birçok puan aralığında ihmal edilebilir düzeydedir (von Davier ve diğerleri, 2004).

### 1.7.2.2. Madde Tepki Kuramına Dayalı Eşitleme

Madde Tepki Kuramı (MTK) bireylerin yeteneklerini matematiksel modellerle açıklar. Lord (1953), gerçek ve gözlenen puanın yetenek puanıyla aynı anlama gelmediğini, yetenek puanının testten bağımsız iken, gerçek ve gözlenen puanının teste bağımlı olduğunu belirtmiştir (akt: Hambleton & Jones, 1993). MTK, iki temel esasa dayanmaktadır. Bunlardan ilki bireyin test maddesi üzerindeki performansının özellik, örtük özellik veya yetenek gibi faktörlerle kestirildiği, diğeri ise bireyin madde performansı ile madde performansının altında yatan özellik arasında ilişki olduğunu ve bu ilişkinin madde karakteristik fonksiyonu ya da madde karakteristik eğrisi adı verilen monoton artan bir fonksiyonla ifade edildiğidir. Madde karakteristik eğrisi madde puanının yetenek üzerindeki regresyonudur. Bu fonksiyona göre, bireyin yetenek düzeyi arttıkça bireyin maddeye doğru cevap verme olasılığı da artmaktadır (Hambleton & Swaminathan, 1985; Hambleton ve diğerleri, 1991).

MTK'da bireyin yeteneği madde parametreleri ve diğer bireylerin performansından, madde parametreleri ise örneklemden bağımsızdır ve bu özelliğe değişmezlik denir (Crocker & Algina, 1986; Zhu, 1998). Ancak değişmezlik özelliği teoride var olduğu kabul edilse de, uygulamada sağlanması güç bir varsayımdır (Zhu, 1998). Madde parametreleri bilinen farklı test formlarını cevaplayan iki bireyin yetenek kestirimleri aynı ölçektir. Bu durumda ölçekleme de eşitleme de gerekli değildir. Ancak madde ve yetenek parametreleri bilinmediğinde durum farklılaşır (Hambleton ve diğerleri, 1991).

MTK gerçek puan eşitleme üç aşamadan oluşur. Geleneksel eşitleme metotlarında olduğu gibi ilk aşama veri toplama deseninin seçilmesidir, ikinci aşama uygun MTK modeline karar verilmesi ve son aşama ise uygun model ile kestirilen madde parametrelerinin ortak ölçeğe yerleştirilmesidir (Cook & Eignor, 1991; Zhu, 1998). Eğer toplam puan kullanılacaksa, puanlar aynı ölçeğe yerleştirildikten sonra toplam puanların eşdeğerleri bulunur. Parametre kestirimleri aynı ölçek üzerinde ise, MTK gerçek puan ve MTK gözlenen puan eşitleme metotları X formuna ait toplam puanlar ile Y formuna ait toplam puanları ilişkilendirmek için kullanılır (Kolen, 2007). MTK gerçek puan eşitlemede, bir forma ait belirli bir  $\theta$  değeriyle ilişkilendirilen gerçek puan ile diğer formdaki aynı  $\theta$  değeriyle ilişkilendirilen gerçek puan arasında ilişki olduğunu varsayar ve bu ilişkiyi eşitleme için kullanır. MTK

gözlenen puan eşitlemesi her form için madde parametre kestirimleri ile toplam puan dağılımları için kestirilmiş yetenek dağılımını kullanır. MTK gözlenen puan eşitleme, her iki test formuna ait ham puan dağılımlarını düzgünleştirerek kullanır (Kolen, 2007; Tong & Kolen, 2005). Güçlükleri bakımından farklı olan testler yetenek dağılımları farklı gruplara uygulanacaksa en iyi yol MTK gerçek puan eşitleme metotlarıdır. Uç puanlarda MTK gerçek puan eşitleme KTK ile eşitlemeden daha iyi eşitleme sonuçları verir (Cook & Eignor, 1991). Parametre kestirimlerinin aynı ölçeğe yerleştirmede ölçekleme sabitinin belirlenmesi önemli bir süreçtir. Bunun için farklı kalibrasyon yöntemleri bulunmaktadır. Bu yöntemleri birbirinden ayıran ise hangi parametre veya bilgiyi kullandığıdır (Zhu, 1998).

Alanyazında birçok MTK modeli bulunmaktadır. Bunlardan iki kategorili puanlanan veriler için 1, 2 ve 3 PLM (Rasch, 1966; Birnbaum, 1957, 1958, 1968); çok kategorili puanlanan veriler için ise adlandırılmalı tepki modeli (Bock, 1922; Samejima, 1972), aşamalı tepki modeli (Samejima, 1969) ve genelleştirilmiş kısmi puan modeli (Muraki, 1992; 1993) sıklıkla kullanılan modellerdir (Embretson & Reise, 2000; Hambleton & Swaminathan, 1985; Kolen & Brennan, 2004). Bu çalışmada iki kategorili maddeler kullanıldığı için bu modeller hakkında bilgi verilmiştir.

#### **1.7.2.2.1 Rasch Model ve Bir Parametrelili Lojistik Model (1PLM)**

MTK modelleri arasında en çok kullanılan modeldir (Hambleton ve diğerleri, 1991). 1 parametrelili lojistik model (PLM) ve Rasch modeli, 3 PLM ve 2 PLM 'in özel bir halidir (Crocker & Algina, 1986; Embretson & Reise, 2000). 1 PLM de madde ayırt edicilik indeksi (a parametresi) tüm maddeler için aynı iken, madde güçlük parametresi (b parametresi) farklılaşmaktadır. Rasch modelin 1 PLM den farkı ise, madde ayırt edicilik değerinin tüm maddeler için 1'e eşit olmasıdır (Embretson & Reise, 2000; De Ayala, 2009). b parametresi yetenek ölçeği üzerinde madde karakteristik eğrisinin konumunu belirler. Eğer madde zor ise yetenek ölçeğinin sağında, kolay ise yetenek ölçeğinin solunda yer alır. b parametresi -3 ile +3 arasında değer alır. 1 PLM'de bireyin maddeye doğru cevap verme olasılığı, madde güçlük indeksi ve bireyin yeteneğinin bir fonksiyonudur. Bir maddenin güçlük indeksi, doğru cevap verme olasılığının 0.5 olduğu noktanın  $\theta$  yetenek ölçeği üzerindeki değeridir. Bu modelde bireyin performansını etkileyen tek madde

parametresinin madde güçlük indeksi olduğu varsayılır (Hambleton ve diğerleri, 1991). Modele ilişkin matematiksel gösterim aşağıdaki gibidir.

$$P(\theta) = \frac{e^{D(\theta-b)}}{1 + e^{D(\theta-b)}}$$

(Eşitlik 1.28)

#### 1.7.2.2.2. İki Parametrelili Lojistik Model (2PLM)

1 PLM bütün maddelerin eşit ayırt ediciliğe sahip olduğu varsayılır; ancak bu durum pratikte pek mümkün değildir. İlk olarak Lord (1952) tarafından 2 parametrelili normal ogive model, daha sonra Birnbaum (1968) tarafından 2 PLM fonksiyonu geliştirilmiştir. Lojistik modeller, ogive modellere göre matematiksel olarak kolay çözümlenebilir olduğu için daha pratiktirler. Lojistik modelde  $D=1,7$  olduğunda, 2 parametrelili lojistik ve ogive modellerdeki  $P(\theta)$  değerleri arasındaki fark 0,01 den daha küçüktür (Hambleton & Swaminathan, 1985; Hambleton ve diğerleri, 1991). 2 PLM de b parametresinin yanında a parametresi de eklenmiştir ve a parametresi maddeler arasında farklılık gösterir. a parametresi, madde karakteristik eğrisinin eğimini belirler. Teorik olarak  $+\infty$  ile  $-\infty$  arasında değer alsada, negatif değerler testten çıkarılır. Ayrıca a parametresinin 2'den büyük olmasında yaygın karşılaşılan bir durum değildir. Bu yüzden 0 - 2 değer aralığı kabul edilir. a değeri arttıkça madde karakteristik eğrisi dikleşir. 2 PLM için madde karakteristik eğrisine ait eşitlik aşağıdaki gibidir.

$$P(\theta) = \frac{e^{Da(\theta-b)}}{1 + e^{Da(\theta-b)}}$$

(Eşitlik 1.29)

2 PLM, örtük özellikle eşit seviyede ilişkili olmayan maddeler için uygundur. Bu modelde, madde ayırt edicilik indeksi arttıkça, madde karakteristik eğrisi dikleşir ve maddeden elde edilen bilgi artar (De Ayala, 2009).

#### 1.7.2.2.3. Üç Parametrelili Lojistik Model (3PLM)

2 PLM'e c parametresinin eklenmesiyle 3PLM oluşur. c parametresi minimum yeteneğe sahip bireyin maddeyi doğru cevaplama olasılığıdır ve madde karakteristik eğrisinin alt asimtotudur. Sahte şans düzey parametresi olarak da adlandırılır. 3 PLM'de her ne kadar madde güçlük indeksi hala madde karakteristik

eğrisinin büküm noktasında yer alsa da, bu nokta maddeye doğru cevap verme olasılığının 0.50 olduğu özellik düzeyinde değildir. Büküm noktası, alt asimtot nedeniyle ötelenmiştir (Embretson & Reise, 2000). 2 PLM, c parametresinin sıfıra eşit olduğu durumda 3PLM'in özel halidir (Seungho-Yang, 2007). 3 PLM için madde karakteristik eğrisine ait eşitlik aşağıdaki gibidir.

$$P(\theta) = c + (1 - c) \frac{e^{Da(\theta-b)}}{1 + e^{Da(\theta-b)}}$$

(Eşitlik 1.30)

Rasch modelin, 2 ve 3 PLM göre bazı avantajları vardır. Birincisi, bireyin testteki doğru cevap sayısı madde güçlük indeksinin kestirimi için ve toplam test puanı da bireyin yetenek kestirimi için yeterli bir istatistiktir. İkincisi Rasch modelde test performansını etkileyen tek parametre madde güçlük indeksi olduğundan daha az madde parametresi kestirilir (Hambleton ve diğerleri, 1991; Seungho-Yang, 2007).

MTK'da bir test formundan elde edilen kestirimlerin diğer test formundan elde edilen kestirimlere dönüştürülmesinde ayrı kestirim ve eş zamanlı kestirim olmak üzere iki yöntem kullanılır. Ayrı kestirim yöntemleri karakteristik eğri ve moment yöntemleri olmak üzere ikiye ayrılır. Bu bölümde, araştırma kapsamında ayrı kestirim yöntemleri kullanıldığından bu yöntemler hakkında bilgi verilmiştir.

#### 1.7.2.2.4. Ayrı Kestirim (Seperate Estimation)

Ayrı kestirim yöntemleri moment yöntemler ve karakteristik eğri yöntemleri olmak üzere ikiye ayrılır. Moment yöntemler ortalama-ortalama ve ortalama-standart sapma, karakteristik eğri yöntemleri ise Haebara ile Stocking ve Lord yöntemleridir. Ortalama-Ortalama yöntemi,  $\alpha$  (eğim) ve  $\beta$  (kesişim) eşitleme katsayılarını elde etmek için ortak maddelerin ayırıcılık ve güçlük parametresinin ortalamasını, ortalama-standart sapma yöntemi,  $\beta$  (kesişim) parametresinin ortalama ve standart sapmasını kullanır (Hambleton ve diğerleri, 1991). Moment yöntemleri, aynı yetenek aralığındaki bireylerde benzer madde karakteristik eğrilerine sahip ancak farklı a,b ve c parametre kestirimleri olan maddeler için hatalı sonuçlar üretecektir. Moment yöntemler kestirimde benzer madde karakteristik eğrisine sahip ancak farklı güçlük parametresine sahip bir madde için, her ne kadar benzer madde karakteristik eğrisine sahip olsa da ortalama-standart sapma yöntemi b parametre kestirimleri arasındaki farktan dolayı fazlaca



etkilenecektir. Bunun nedeni ise bu yöntemlerin ölçek dönüştürmede tüm madde parametre kestirimlerini aynı anda dikkate almamasından kaynaklanır. Haebara (1980) bu problemlerin üstesinden gelmek için tüm parametre kestirimlerini aynı anda göz önüne alan bir yöntem önermiştir. 1983 yılında ise Stocking ve Lord, Haebara yöntemine benzer başka bir yöntem önermiştir. Haebara ile Stocking ve Lord yöntemleri karakteristik eğriler arasındaki farklılığı ele alma şekillerine göre farklılaşırlar.

**Haebara Yöntemi:** Belirli bir yetenek düzeyi için madde karakteristik eğrileri arasındaki fark, her bir madde için madde karakteristik eğrileri arasındaki farkın kareleri toplamını ifade eder. Bu yöntemin matematiksel gösterimi aşağıdaki gibidir.

$$Hdiff(\theta_i) = \sum_{j:v} \left[ p_{ij}(\theta_{ij}; \hat{a}_{Jj}, b_{Jj}, c_{Jj}) - p_{ij}(\theta_{ji}; \frac{\hat{a}_{Ij}^*}{\alpha}, \alpha b_{Ij}^* + \beta, c_{Ij}) \right]^2$$

(Eşitlik 1.31)

**Stocking ve Lord Yöntemi:** Haebara yönteminin aksine Stocking ve Lord yöntemi toplamların farkının karesini kullanır. Kareleri alınmadan önce her bir madde seti için madde karakteristik eğrisinin toplamını alır.

$$\tau(\theta_i) = \sum_j p_{ij}(\theta_i)$$

(Eşitlik 1.32)

SL yöntemi, her belirli bir yetenek düzeyi için madde karakteristik eğrileri farkının karesini kullanır. Bu yöntemin matematiksel gösterimi ise aşağıdaki gibidir.

$$SLdiff(\theta_i) = \left[ \sum_{j:v} p_{ij}(\theta_{ij}; \hat{a}_{Jj}, b_{Jj}, c_{Jj}) - \sum_{j:v} p_{ij}(\theta_{ji}; \frac{\hat{a}_{Ij}^*}{\alpha}, \alpha b_{Ij}^* + \beta, c_{Ij}) \right]^2$$

(Eşitlik 1.33)

Belirli bir  $\theta$  düzeyinde madde karakteristik eğrileri aynı olduğunda  $Hdiff(\theta) = 0$  olur, oysa  $SLdiff(\theta)$  madde karakteristik eğrileri farklı olduğunda bile 0 olabilir. Bu nedenle Haebara yöntemi Stocking ve Lord yöntemine göre katıdır.

Karakteristik eğri yöntemleri ile moment yöntemlerinin karşılaştırıldığı çalışmalarda karakteristik eğri yöntemlerinin daha kararlı olduğu sonucuna ulaşılmıştır (Hanson & Béguin, 2002; Karkee & Wright, 2004; Kim & Cohen, 1998; Kim & Lee, 2004;

Kim & Kolen, 2006; Lee & Ban, 2010). Kernel eşitleme ile MTK gerçek puan eşitleme yöntemlerinin birlikte incelendiği çalışmalarda (Godfrey, 2007; Meng, 2012; Norman Dbrovak, 2009) ayrı kestirim yöntemlerinden Stocking ve Lord ve ortalama-standart sapma yöntemlerinin kullanıldığı görülmüştür. Bu nedenle bu çalışmada karakteristik eğri yöntemlerinden Haebara yöntemi kullanılmıştır.

## 2. İLGİLİ ARAŞTIRMALAR

Bu bölümde, ilgili alanyazında yürütülmüş olan araştırmalar bulunmaktadır.

Livingston (1993) yaptığı çalışmada, küçük örneklerde zincirleme eşit yüzdelikli eşitlemede log lineer düzgünleştirmeyi incelemiştir. Testi 93283 kişinin cevapladığı 100 soruluk çoktan seçmeli test kullanılarak çalışmada kullanılacak X ve Y formu ve ortak test oluşturulmuştur. Her bir koşul için 50 tekrar yapılmıştır. Çalışmada örneklem büyüklüğü 25, 50, 100 ve 200 olarak toplam veriden random olarak seçilmiştir. Log lineer düzgünleştirme için 4 durum oluşturulmuştur. İlk durumda düzgünleştirme yapılmamıştır. Birinci düzgünleştirmede, ilk iki tek değişkenli moment (ortalama-standart sapma) ve ilk iki değişkenli moment (alt formlar ile ortak test arasındaki korelasyon), ikinci düzgünleştirme de birinci düzgünleştirmeye ek olarak üçüncü tek değişkenli moment (çarpıklık), üçüncü düzgünleştirmede ise dördüncü tek değişkenli moment (basıklık) eklenmiştir. Elde edilen sonuçlar için kriter eşitleme yöntemi olarak tek grup deseninde eşit yüzdelikli eşitleme yöntemi kullanılarak RMSD değeri hesaplanmıştır. Çalışmada örneklem büyüklüğü küçük olduğunda düzgünleştirmenin daha önemli olduğu, düşük uç değerlerde RMSD'nin daha büyük değerler aldığı sonucuna ulaşmıştır.

Liou, Cheng ve Johnson (1997) yaptıkları çalışmada, denk olmayan gruplarda ortak madde test deseninde Kernel eşitlemede tek biçimli ve gauss Kernel metotlarında standart hatayı incelemiştir. Ayrıca gauss ve tek biçimli Kernel için iki ayrı bant genişliği (1 ve 3) seçmişlerdir. Her bir koşul için 100 tekrar yapılmıştır. İlk çalışmada 55 sorudan oluşan iki adet İngilizce testi kullanılmış, daha sonra 25 soruluk ortak test verilmiştir. Her iki test de 719 birey tarafından cevaplanmıştır. Elde edilen puan dağılımları log lineer model ile düzgünleştirilmiştir. Birinci ve ikinci testi cevaplayan bireylerden random olarak 100 ve 1000 kişilik örneklem seçilmiştir. İkinci çalışmada 1990 yılına ait 9229 bireyin cevapladığı 112 maddelik çoktan seçmeli test kullanılmıştır. Maddeler 3 PLM göre analiz edilmiş ve bir soru testten çıkarılmıştır. 112 soruluk test 50 soruluk iki teste ayrılmış ve 11 soru ortak madde olarak belirlenmiştir. Bu çalışmada içinde random olarak 100 ve 1000 kişilik örneklem seçilmiştir. Örneklem büyüklüğü 1000 olduğunda, standart hata eşitliklerinin iyi performans verdiklerini, küçük örneklemde ise tek biçimli ve gauss Kernel metotlarının birbirine benzer sonuçlar verdiğini bulmuşlardır. Ancak bant

geniřliđi üç olduđunda, Gauss Kernel metotun standart hatasının azaldığı sonucuna ulařmıřlardır.

Mao, von Davier ve Rupp (2006) yaptıkları arařtırmada, Kernel eřitleme ile gözlenen puan eřitleme yöntemlerinden zincirleme dođrusal eřitleme, zincirleme eřit yüzdelikli eřitleme, Tucker dođrusal eřitleme yöntemlerini NEAT ve eřdeđer grup desenlerinde karřılařtırmıřlardır. alıřmada gerek veri kullanılmıřtır. Eřdeđer grup deseninde, Kerneleřitlemede ideal bant geniřliđi kullanıldıđından Kernel eřitleme ile geleneksel eřit yüzdelikli eřitleme sonuçlarının ok yakın olduđu, geniř bant geniřliđi kullanıldıđında ise neredeyse aynı sonuçlar verdiđini bulmuřlardır. NEAT deseninde i ve dıř ortak testte, Kernel son tabakalama eřitlemede ideal bant geniřliđi kullanıldıđında Kernel eřitleme ile frekans kestirimi sonuçlarının benzer olduđu, geniř bant geniřliđi kullanıldıđında Tucker eřitleme sonuçlarıyla benzer sonuçlar verdiđi sonucuna ulařmıřlardır. Ancak Kernel eřitlemeyle diđer eřitleme yöntemleri arasında sadece dıřuk puan aralıđında geniř farklılıklar bulmuřlardır.

von Davier ve diđerleri (2006), alıřmalarında gerek veri seti üzerinden KE ile eřit yüzdelikli eřitleme metotlarından zincirleme ve frekans kestirimi eřitleme yöntemleri ile dođrusal eřitleme metotlarından Tucker ve Levine gözlenen puan eřitleme yöntemlerini NEAT deseninde i ve dıř ankor test kullanarak karřılařtırmıřlardır. RMSE hesaplanabilmesi iin kriter olarak eřdeđer grup deseni seilmiřtir. Sonuç olarak, Kernel eřitleme sonuçlarının diđer eřitleme sonuçlarına benzer sonuçlar verdiđini bulmuřlardır. Ayrıca Kernel eřitlemenin kriter olarak belirlenen eřdeđer grup desenine göre elde edilen sonuçlara yakın sonuçlar verdiđi görülmüřtür.

Mao (2006) yaptıđı alıřmada, Kernel eřitlemede random grup deseninde örneklem büyüklüđü, yetenek dađılımı, bant geniřliđinin seimi ve düzgünleřtirme parametre derecesi faktörlerine göre standart hata deđerlerini karřılařtırmıřtır. Yaptıđı simülyasyon alıřmasında gerek test kořullarındaki durumlarını yansıtmaları iin 4 adet gerek veri setinden elde ettiđi deđerler kullanmıřtır. Elde edilen sonuçlara göre, Kernel eřitlemede deneysel standart hata ve kestirilen standart hatanın dođruluđu örneklem büyüklüđü, bant geniřliđi ve düzgünleřtirme parametresinden etkilenmektedir. Ancak farklı puan dađılımlarında kestirilen standart hatanın örneklem büyüklüđü, bant geniřliđi ve düzgünleřtirme

parametresi etkisinin bütün testlerde aynı olduğu bulgusuna ulaşmıştır. Çalışma sonucunda grupların dağılımı normal veya dağılımları aynı olduğunda standart hatanın kestirilmesini çok fazla etkilemediğini, bunun da Kernel eşitlemede standart hatanın normallik varsayımı gerektirmemesi olduğunu belirtmiştir. Düzgünleştirmenin random hatayı azalttığını ve özellikle örneklem sayısı küçük olduğunda eşitleme fonksiyonlarının kesinliğini geliştirdiğinden uygulamada kullanılması gerektiğini önermiştir.

Holland, von Davier, Sinharay ve Han (2006) yaptıkları çalışmada, Kernel eşitlemede zincirleme ve son tabakalama eşitleme yöntemlerini karşılaştırmışlardır. 119 maddelik gerçek veri seti kullanarak X ve Y formlarını ve farklı uzunluklarda üç ankor test oluşturmuşlardır. Grupların yetenek dağılımının farklılığının az olduğu durumda iki yöntemin benzer sonuçlar verdiği, yetenek dağılımı farklılığı büyük olduğunda zincirleme eşitleme yönteminin son tabakalama eşitleme yöntemine göre daha iyi sonuçlar verdiği bulgusuna ulaşmışlardır. Ayrıca ortak test uzunluğu ve güvenilirliği arttıkça son tabakalama eşitlemenin tutarlılığının arttığını ancak zincirleme eşitleme için tutarlı sonuçlar vermediğini belirtmişlerdir.

Sinharay ve Holland (2006a), Rasch ve 2 PL modele göre 1000 örneklem için simülasyon verisi üretmiş ve gerçek veri üzerinde analiz yapmışlardır. Mini, midi ve semi midi ortak testlerin toplam testin ham puanları ile korelasyonları incelenmiştir. Ortak testteki madde güçlüğü dağılımının toplam teste göre daha küçük olduğu durumda (midi ortak test), toplam test ile ortak test arasındaki korelasyonun daha yüksek olduğu sonucuna varılmıştır.

Sinharay ve Holland (2006b, 2007), tek boyutlu ve çok boyutlu Madde tepki Kuramına dayalı olarak ortak testte madde güçlük dağılımlarının farklı olduğu koşullara uygun simülasyon veri ve gerçek veri üzerinde çalışmışlardır. Bu çalışmada midi, mini ve semi-midi ortak testlerin eşitleme üzerinde etkisi incelenmiştir. Test uzunluğu, örneklem büyüklüğü, test formlarının güçlükleri arasındaki farklılık, bireylerin yetenek dağılımları arasındaki farklılık değişkenlerinin eşit yüzdelli eşitleme yöntemlerinden zincirleme eşitleme ve son tabakala eşitleme yöntemlerini toplam hata, eşitleme hatası ve yanlılık değerleri açısından karşılaştırmışlardır. İç ortak testte midi ve semi midi ortak testin, mini ortak test kadar iyi sonuçlar verdiği, dış ortak testte ise midi ortak testin daha avantajlı olduğu sonucuna varmışlardır. Ayrıca yanlılık değerleri açısından

zincirleme eşitleme yönteminin, RMSE değerleri açısından ise son tabakalı eşitleme yönteminin daha iyi sonuç verdiği görülmüştür.

Ricker ve von Davier (2007) çalışmalarında, 119 maddeden oluşan gerçek veri setinden NEAT desenine göre ortak maddelerinde oluşturulduğu yalancı (pseudotests) test oluşturmuşlardır. Elde edilen testlere eşitleme yapılmadan önce cinsiyete ve uygulandığı tarihe göre DMF analizi (Mantel-Haenzsel) yapılmış ve maddelerden hiçbiri DMF vermemiştir. Bu çalışmada NEAT deseninde dış ortak testin eşit yüzdelikli (frekans kestirimi), zincirleme eşit yüzdelikli, Kernel son tabakalama eşitleme (ideal ve geniş bant genişliği) yöntemlerine göre eşitleme sonuçları karşılaştırılmıştır. Eşitleme sonuçlarının değerlendirilmesinde yanlılık (bias) ve RMSE kullanılmıştır. Ayrıca kriter olarak eşdeğer grup eşit yüzdelikli (frekans kestirimi) eşitleme yöntemi seçilmiştir. Elde edilen sonuçlara göre, ortak madde oranı arttıkça yanlılığın azaldığı görülmüştür. Ancak bu durumun Kernel son tabakalama (ideal bant genişliği) ve frekans kestirimi eşitleme yöntemlerinin diğer yöntemlere göre daha az duyarlı olduğu bulgusuna ulaşılmıştır. Ayrıca bütün ortak madde oranlarında Kernel son tabakalama (geniş bant genişliği) eşitleme kriter olan eşdeğer grup eşit yüzdelikli eşitlemeyle karşılaştırıldığında en kötü sonuçları vermiştir.

Qu (2007) yaptığı çalışmada, dengelenmiş grup deseninde ağırlıklandırmanın Kerneleşitlemede etkisini incelemiştir. Dengelenmiş grup deseni oluşturmak için iki tane tek grup deseni oluşturularak, örneklem büyüklüğü ve sıra etkisi durumu faktörlerini ele almıştır. Kernel doğrusal ve Kernel eşit yüzdelikli eşitleme ile geleneksel eşitleme metotlarından eşit yüzdelikli eşitleme metodu hem gerçek veri hemde simülasyon veri kullanarak karşılaştırmıştır. Eşit yüzdelikli eşitleme metodu, değerlendirme kriterlerinin hesaplanması için kriter eşitleme metodu olarak belirlenmiştir. Bu çalışmada simülasyon veri için yanlılık, eşitlemenin standart hatası, toplam hata (RMSE) ve eşitlemenin standart hatası farkı, gerçek veri için ise RMSE ve eşitlemenin standart hatası kullanılmıştır. Çalışma sonucunda genel olarak her iki veri türü içinde ağırlıklandırmanın etkili olduğu elde edilmiştir. Dengelenmiş grup deseninde iki tek grup deseni yaklaşımının ağırlıklandırmayı ayarlayarak sistematik hatayı azalttığı, ancak yanlılıktaki azalmanın sadece örneklem büyüklüğü geniş ve sıra etkisi yeteri kadar büyük olduğunda önemli olduğu sonucu bulgusuna ulaşmışlardır.

Godfrey'in (2007) yaptığı çalışmanın iki amacı bulunmaktadır. Birincisi Kernel eşitlemede iki değişkenli puan dağılımında öndüzgünleştirmede log lineer modellerin etkisini belirlemek, ikincisi ise Kernel eşitleme ile zincirleme eşit yüzdelikli eşitleme, MTK gerçek puan eşitleme metotlarından Stocking-Lord ile aynı kalibrasyon yöntemlerinden elde edilen sonuçların karşılaştırılmasıdır. Bu çalışmada hem tek grup hem de NEAT deseni kullanılmış ve tek grup deseninde düzgünleştirilmemiş eşit yüzdelikli eşitlemeyi de kriter eşitleme metodu olarak seçmiştir. Çalışmada 2 PL modele göre veriler üretilmiş ve örneklem büyüklüğü 1000, 10.000 ve 100.000 olarak, test 100, 60 ve 20 olarak, ortak madde oranı 20 soruluk test için %50, 100 ve 60 soruluk test için ise %20, %35 ve %50 olarak belirlenmiştir. Çalışmanın birinci aşamasında ki-kare istatistiklerine göre veriye uyuyorsa log lineer modelleri, modellerin kararlılığını anlamak için ise Freeman-Tukey hata (artık-residual) kullanılmasını ve kabul edilebilir ki-kare uyum istatistikleri ve Freeman-Tukey hatalarla birlikte bir kaç momentin veriye uyumunun yeterli olduğunu önermişlerdir. Çalışmanın ikinci aşamasında, Kernel eşitleme ile diğer eşitleme yöntemlerinin benzer sonuçlar verdiği, bazı durumlarda ise kriter eşitlemeye çok yakın değerler verdiğini ve Stocking-Lord yönteminin en iyi sonuçları verdiğini, ancak bu durumun MTK'ya göre madde parametreleri üretildiğinden kaynaklanabileceği sonucuna ulaşmıştır. Özellikle gerçek puan eşitlemenin yapılamadığı durumlarda (örneklem sayısının veya test uzunluğunun küçük olduğu durumlar gibi) Kernel eşitlemenin geçerli bir seçenek olduğunu belirtmiştir.

Moses ve Holland (2007) yaptıkları çalışmada, eşitleme yöntemlerindeki log lineer düzgünleştirmenin yanlılık, standart hata, eşitlemenin standart hata farkı üzerinde etkisini incelemişlerdir. Çalışmada dış ortak NEAT deseninde zincirleme Kernel, zincirleme eşit yüzdelikli, Kernel son tabakalama, son tabakalama eşit yüzdelikli eşitleme metotlarını kullanmışlardır. Örneklem büyüklüğü 100, 200, 1000 olarak, üç ön düzgünleştirme derecesi (ilki tek değişkenli dağılımın ilk iki momenti olan ortalama ve varyansı ile test ve ortak test arasındaki korelasyonu, ikincisi tek değişkenli dağılımın ilk altı momenti ile test ve ortak test arasındaki korelasyonu, üçüncüsünde ise gerçek popülasyona ait momentler) ve süreklileştirme metodu olarak Kernel ve yüzdelik sıra kullanmışlardır. NEAT deseninde log lineer ön düzgünleştirmenin yanlılığı etkilediği ve hatta öndüzgünleştirme derecesi ile

yanlılık arasında direk ilişki olduğu; süreklileştirme metotları açısından ise Kernel süreklileştirmenin yüzdelik sıra metoduna göre önemli avantajları olduğu bulgusuna ulaşmışlardır.

Liu, Sinharay, Holland, Curley ve Feigenbaum (2009, 2011), gerçek veri ile (SAT Testi) yaptıkları çalışmada mini ve midi ortak testin gözlenen puan eşitlemesine dayalı yanlılık, random eşitleme hatası ve eşitleme hatası kriterlerine göz önünde bulundurularak incelenmiştir. Bu çalışmada eşitleme metotlarından Tucker, frekans kestirimi, doğrusal zincirleme ve eşit yüzdelikli zincirleme eşitleme kullanılmıştır. Elde edilen bulgularda her üç değerlendirme kriterine göre de midi testin, mini test kadar iyi eşitleme sonuçları verdiği görülmüştür. Ancak puan ölçeğinin üst bölümlerinde (highend of scorescale) eşitleme kesinliği önemli ise, mini ortak testin kullanılması gerektiğini önermişlerdir.

Norman Dvorak (2009) yaptığı çalışmada, NEAT desenine göre Kernel eşitleme ile MTK gerçek puan eşitleme yöntemini karşılaştırmıştır. MTK gerçek puan eşitleme de kalibrasyon yöntemlerinden Stocking-Lord yöntemini kullanmıştır. Monte Carlo simülasyon yöntemini kullanarak 2 PL faktör analitik modele göre veri üretmiştir. Örneklem büyüklüğü olarak 1000, 2000 ve 10.000, test uzunluğunu 25, 50 ve 75, ortak madde oranını 25 maddelik test için %20 ve %30, 50 ve 75 maddelik test uzunlukları için %10, %20 ve %30 olarak ve ortalama faktör yükleri ise 0.50 ve 0.62 olarak çalışmada belirlenmiştir. Yöntemleri değerlendirmek için RMSD kriteri kullanılmıştır. Test uzunluğu 75 ve ortalama faktör yükü 0.62 olduğunda, ortak madde oranı ve örneklem büyüklüğü koşulları gözetilmeksizin Kernel eşitleme metotları daha kesin sonuç verirken, ortalama yük 0.50, geniş örneklem ve ortak madde oranı %30 olduğunda MTK gerçek puan eşitlemenin daha iyi sonuçlar verdiği görülmüştür. Ayrıca KE metotlarının orta ile yüksek puan aralığında, MTK gerçek puan eşitlemenin ise bütün puan aralığında daha iyi sonuçlar verdiği bulgusuna ulaşılmıştır. Genel olarak, Kernel eşitlemenin MTK gerçek puan (Stocking-Lord) eşitleme ile karşılaştırıldığında iyi performans gösterdiğini vurgulamıştır.

Grant, Zhang ve Damiano (2009) yaptıkları çalışmada, gerçek veri kullanarak Kernel zincirleme eşitleme, Kernel son tabakalama eşitleme, Tucker eşitleme, zincirleme eşit yüzdelikli eşitleme ve Levine eşitleme sonuçları karşılaştırmıştır. Kernel zincirleme ve Kernel son tabakalama eşitlemede bant genişlikleri ideal ve



geniş olmak üzere iki tür seçmişlerdir. Geniş bant genişliğinin seçildiği Kernel zincirleme eşitleme sonuçlarının zincirleme doğrusal eşitleme sonuçlarıyla benzer olduğunu, Kernel son tabakalama eşitleme sonuçlarıyla Tucker ve Levine eşitleme sonuçları arasında küçük farklar olduğu bulgusuna ulaşmışlardır. Ayrıca Tucker doğrusal eşitlemenin Levine eşitlemeye göre Kernel eşitlemeye daha yakın sonuçlar verdiği görülmüştür. İdeal bant genişliğinin kullanıldığı durumda ise, Kernel son tabakalama ve Kernel zincirleme eşitleme sonuçlarıyla zincirleme eşit yüzdelikli eşitleme metotları karşılaştırılmış ve neredeyse bütün puan aralıklarında farkın küçük olduğu sonucuna ulaşmışlardır.

Choi (2009) yaptığı çalışmada, Kernel zincirleme ve Kernel son tabakalama eşitleme ile geleneksel eşitleme yöntemlerinden frekans kestirimi ve Tucker doğrusal eşitleme yöntemlerini NEAT ve eşdeğer grup deseninde incelemiştir. Ayrıca bu çalışmada eşdeğer grup için eşitlemenin standart hatasının hesaplanmasında parametrik bootstrap metodu ile parametrik olmayan bootstrap metodu ve delta modele dayanan analitik model de karşılaştırılmıştır. Kernel eşitleme için küçük, ideal ve geniş olmak üzere üç ayrı bant genişliği ve öndüzenleştirilmiş eşit yüzdelikli eşitleme için ise 2, 6, 8 ve 10 olmak üzere dört tane düzgenleştirme parametresi kullanılmıştır. Her bir koşul için 300 tekrar yapılmıştır. 2 PL modele göre veriler üretilmiş ve örneklem büyüklüğü 300, 500, 700, 1000, 3000 ve 5000 olarak ve eşdeğer grup için test uzunluğu 25, 50 ve 75, NEAT deseni için test uzunlukları ise 60, 80 ve 125 olarak belirlenmiştir. Değerlendirme kriterlerinden yanlılık, RMSD ve standart hata kullanılarak, eşdeğer grup için geleneksel eşit yüzdelikli eşitleme ile karşılaştırılmışlardır. Çalışma sonucunda eş değer grup için KE ile Kernel eşitlemenin karşılığı olan doğrusal yöntemler karşılaştırılabileceğini, ancak NEAT deseninde sadece ideal ve küçük bant genişliğinde son tabakalama eşitleme ile frekans kestirimi metodunun benzer sonuçlar verdiği bulgusuna ulaşmışlardır. Bunun yanında bütün test ve örneklem büyüklüğü koşullarında parametrikbootstrap metotunun diğer yöntemlere göre daha iyi sonuç verdiğini bulmuşlardır.

Hagge (2010) tarafından yapılan çalışmada, İleri Yerleştirme Sınavının (Advanced Placement Examination program) İspanyol, İngilizce ve Kimya testleri kullanılmıştır. Bu çalışmada karma format testlerde midi ve mini ortak testin MTK gerçek puan, MTK gözlenen puan, frekans kestirimi ve zincirleme eşit yüzdelikli

eşitlemenin sınıflama tutarlılığı, RMSE, ağırlıklandırılmış ortalama standart hata (the weighted average standard error of equating-WASE) ve ağırlıklandırılmış ortalama yanlılık (weighted average root means quared bias-WARMSB) kriterlerine göre performansı incelenmiştir. İlk durumda çoktan seçmeli ve yapılandırılmış maddeler eşit güçlükte, ikinci durumda bu farklı madde tiplerinin güçlükleri farklıdır. Elde edilen bulgulara göre bireylerin yetenek dağılımı ve ortak maddeler eşitleme sonuçlarını etkilemektedir. Grupların yetenek dağılımları arasında farklılık büyük ise MTK gözlenen ve gerçek puan eşitlemesinin tercih edilmesini ve bu farklılığın daha büyük yanlılık değerlerine sebep olduğu görülmüştür. Ancak seçmeli maddeler ile yarı yapılandırılmış maddeler arasında korelasyon yüksek ise grupların yetenek dağılımları farkı büyük olsa da yanlılık görece olarak daha düşük çıktığı sonucuna varılmıştır. Araştırma bulgularına göre maddeler eşit güçlükte iken, mini testin çok az daha düşük WARMSB ve BIAS değerleri elde edilmiştir. Madde türleri farklı güçlükte olduğunda ise mini testin daha düşük yanlılık değerlerine sahip olduğu bulunmuştur.

Meng (2012) yaptığı çalışmada, Kernel eşitleme ile MTK gerçek puan eşitleme metotlarını (Stocking-Lord ve Ortalama-Standart sapma) karşılaştırmıştır. Gerçek veri seti üzerinden 3 PL modele göre örneklem büyüklüğü 200, 500 ve 2000, ortak madde oranı %12,5 ve %37,5 ve yetenek dağılımının benzer (0,1) ve farklı (0.5,1) olarak farklı veri setleri üretilmiştir. Değerlendirme kriterlerinden yanlılık, eşitlemenin standart hatası (SEE) ve eşitleme hatası (RMSE) seçilmiştir. Ayrıca gerçek veri setiyle de eşitleme yapılmış ve değerlendirme kriteri olarak kullanılmıştır. Üç eşitleme yönteminde de ortak madde oranı ve örneklem büyüklüğü arttıkça eşitleme hatasının azaldığı görülmüştür. Grupların yetenek dağılımı farklı olduğunda, Kernel eşitlemenin yüksek yanlılık ürettiği bulgusuna ulaşılmıştır. Genel olarak Kernel eşitlemenin MTK gerçek puan eşitleme ye göre daha az kararlı sonuçlar ürettiği, ancak test puan dağılımının orta değerlerinde daha iyi performans gösterdiği sonucuna ulaşılmıştır.

Kim (2014), NEAT deseninde düzgünleştirme metotlarının performansını incelemiştir. Bunun için Kernel eşitleme, frekans kestirimi, modifiye edilmiş frekans kestirimi, eşit yüzdelikli eşitleme ve zincirleme eşitleme metotlarını kullanmıştır. Düzgünleştirme parametresi ise polynomial log-linear öndüzgünleştirme ve kübik spline son düzgünleştirme olmak üzere iki türlü seçilmiştir. Madde parametrelerinin

gerçek test koşullarını yansıtması için madde parametre havuzundan 3 PL modele göre seçilmiştir. Çalışmada kullanılan değişkenlerden örneklem büyüklüğü 300, 2000 ve 6000, gruplar arası yetenek dağılımı farkı, ortak madde oranı %20 ve %40, ortak madde testi iç ve dış ankor olarak ve test formlarının güçlükleri ise benzer  $(0,1;0.05,1)$  ve farklı  $(0,1; 0.2,1)$ , ortak madde testinin güçlük dağılımı ise mini ve midi test olmak üzere toplamda 192 koşul oluşturulmuştur. Koşullara göre elde edilen sonuçlar için yanlılık, standart hata ve RMSE değerlendirme kriteri olarak seçilmiştir. Frekans kestirimi ve modifiye edilmiş frekans kestirimi kullanıldığında, toplam hata ve sistematik hata açısından polynomial log-linear öndüzgünleştirme, kübik spline son düzgünleştirmeye göre daha iyi sonuçlar verdiği, kübik spline son düzgünleştirmenin bütün eşitleme yöntemlerinde daha düşük random hata verdiği bulgusuna ulaşılmıştır. Ayrıca Kernel eşitleme koşulların büyük çoğunluğunda zincirleme eşitlemeden daha doğru eşitleme ilişkisi sonucu verdiği bulunmuştur.

Liang ve von Davier (2014) yaptıkları çalışmada, Kernel eşitlemede bant genişliği seçiminde ceza fonksiyonu ile CV (Cross validation) metodu (çapraz geçerleme-verisetini alt bölümlere ayırarak, bir alt bölüm üzerinden analizleri gerçekleştirir ve analizleri diğer alt bölümlere uygular)karşılaştırmışlardır. CV metotta ilk standartlaştırılmış ve standartlaştırılmamış Kernel olmak üzere iki ayrı yöntem kullanmıştır. Random grup desenine göre iki test formu simüle edilmiştir. Örneklem büyüklüğü 200 ve 2000 ve grupların dağılımı beta dağılımından simetrik, pozitif çarpık ve negatif çarpık dağılım olarak belirlenmiştir. Toplamda 188 koşulun her biri için 500 tekrar yapılmıştır. Elde edilen sonuçlar yanlılık, eşitlemenin standart hatası, eşitlenmiş puanlar arasındaki fark ve görelî hata yüzdesi kriterlerine göre değerlendirmişlerdir. Dağılım simetrik olduğunda, yoğunluk fonksiyonunun ilk iki momentinin daha hatasız sonuçlar verdiğini bulmuşlardır. Ayrıca CV metodun diğer metotlara göre daha düşük standart hata, görelî hata yüzdesi ve yanlılık verdiğini ve örneklem büyüklüğü arttıkça yanlılığın azaldığı sonucuna ulaşmışlardır.

Andersson ve von Davier (2014) yaptıkları çalışmada, Kernel eşitlemede bant genişliği seçiminde cezafonksiyonunun sadece ilk bölümünün kullanıldığı, ceza fonksiyonunun tamamının kullanıldığı ve thumb kuralının kullanıldığı üç metodu karşılaştırmışlardır. Çalışmada gerçek ve simülasyonverifarklı dağılımlarda ve

farklı veri toplama desenleri (NEAT ve eşdeğer grup desenler) kullanmışlardır. Her bir koşul için 5000 bootstrap tekrarı kullanılmış ve eşitlemenin standart hatasını hesaplamışlardır. Bant genişliği seçiminde önerilen thump kuralı ile ceza fonksiyonu sonuçlarının birbirine yakın sonuçlar verdiğini bulmuşlardır. Eşdeğer grup deseninde grupların yetenek dağılımları farklı olduğunda standart hatadaki en büyük farklılık thumb kuralında elde etmişlerdir. Thump kuralının, ceza fonksiyonuna göre hem daha basit hemde daha hızlı olması sebebiyle kullanılmasını önermişlerdir.

Antal, Proctor ve Melican (2014), 2 PL modele göre yapılan simülasyon çalışmasında NEAT desenine göre 6 farklı test formunu eşitlemiştir. Dış ortak teste sahip bu testlerde midi ve mini ortak testin MTK gerçek puan gerçek puan eşitlemesi RMSD, ortalama mutlak fark (MAD) ve ortalama işaretli fark (MSD) kriterlerine göre değerlendirilmiştir. Elde edilen bulgular Sinharay ve Holland'ın (2006a, 2006b,2007) yaptıkları çalışmayı desteklemiş ve mini ve midi ortak testin benzer sonuçlar verdiği görülmüştür.

Sinharay, Haberman, Holland ve Lewis (2012) tarafından yapılan çalışmada, birkaç Madde tepki kuramına dayalı simülasyon verisi üretilmiştir. Elde edilen sonuçlara göre midi testin mini teste göre toplam testle daha yüksek korelasyon verdiğini ve midi testin eşitlemede mini test kadar iyi sonuçlar verdiğini ifade etmişlerdir.

Kim (2014) tarafından yapılan çalışmada, çeşitli koşullar altında MTK gerçek puan olarak örneklem büyüklüğü, yetenek dağılımı, ortak madde özelliği, test formları arasındaki güçlük farkı, test uzunluğu, ortak madde tipi ve ortak maddelerin güçlük dağılımı koşulları göz önünde tutularak veri üretilmiştir. Yanlılık, RMSE ve standart hata kriterleriyle düzgülendirme metotlarının (frekans kestirimi, genelleştirilmiş frekans kestirimi, zincirleme yüzdellikli eşitleme ve Kernel eşitleme) performansı karşılaştırılmıştır. Ortak maddelerin ortalama güçlükleri farklı olduğunda, sistematik, random ve toplam hatanın arttığı, özellikle sistematik hatanın random hatadan daha fazla artış gösterdiği görülmüştür. Ayrıca iç ortak testin dış ortak teste göre daha büyük hatalara sahip olduğu sonucu elde etmiştir. Midi ortak test kullanıldığında sistematik ve toplam hatanın daha az olduğu, sistematik hatanın ise mini ortak testte biraz daha az olduğu sonucuna ulaşılmıştır.

Fitzpatrick ve Skorupski (2016) tarafından yapılan çalışmada, aynı kalibrasyon, Stocking Lord, Ortalama ortalama ve ortalama standart sapma MTK gerçek puan eşitleme yöntemleri iç ortak testte mini ortak test ve midi ortak test koşuluna göre incelenmiştir. Değerlendirme kriterlerinden ortalama RMSE, RMSE, bias ve ortalama bias hesaplanmıştır. Örneklem büyüklüğü 1000 ve test uzunluğu 80 olarak belirlenmiştir. 4 yetenek dağılımı koşulu ile ortak testin güçlük parametresi değiştirilmiştir. Birçok koşulda midi ortak testin mini ortak test kadar iyi sonuçlar verdiği bulgusuna ulaşmışlardır.

## 2.1. İlgili Araştırmalar Özeti

Kernel eşitleme 1981'de Holland ve Thayer tarafından ortaya atılmış ve 2004 yılında von Davier, Holland ve Thayer tarafından geliştirildikten sonra alanyazında kullanılmaya başlanmıştır. İlk yıllarda bilgisayar programının olmaması Kernel eşitleme yöntemlerinin kullanılmasında araştırmacıları kısıtlamıştır. Daha sonra Eğitimsel Test Hizmetlerindeki (ETS) araştırmacılar tarafından KE (2004) programı geliştirilmiş ve bu tarihten sonra eşitleme alanında Kernel eşitleme çalışmaları daha sık kullanılmaya başlanmıştır. İlk olarak Livingson (1993) tarafından log lineer düzgünleştirme kullanılarak küçük örneklerde eşitleme çalışması yapılmıştır. 2004 yılından sonra Kernel eşitleme yöntemlerinin karşılaştırıldığı veya Kernel eşitleme yöntemleri ile geleneksel eşitleme yöntemlerinin karşılaştırıldığı çalışmalar daha fazla iken, son yıllarda MTK gerçek ve gözlenen puan eşitleme yöntemleriyle karşılaştırıldığı çalışmalarında arttığı dikkat çekmektedir. Çalışmalar incelendiğinde, eşitlemenin doğruluğu ve kesinliği için eşitleme deseni, örneklem büyüklüğü, yetenek dağılımı, test uzunluğu, ortak madde oranı, ortak madde tipi, ortak madde güçlük dağılımı, bant genişliği gibi değişkenler manipüle edilerek vefarklı değerlendirme kriterleri kullanılarak çalışmaların yapıldığı görülmüştür. Bu çalışmanın hem MTK gerçek puan eşitleme hemde Kernel eşitlemenin doğrusal ve eşit yüzdelikli yöntemlerini kapsamı, ele alınan koşulların diğer çalışmalarda birlikte ele alınmamış olması ve simülasyon koşullarıyla bu koşulların yöntemlere göre nasıl performans gösterdiği belirlenmeye çalışıldığından alanyazına katkı sağlayacağı düşünülmektedir. Ayrıca Türkiye'de yapılan çalışmalar incelendiğinde, Kernel eşitleme ile herhangi bir çalışmaya rastlanılmamıştır. Bu çalışmanın bir amacı da bu yöntemi ülkemizde tanıtmaktır.

### 3. YÖNTEM

Bu bölümde temel alınan araştırma türü, simülasyon koşulları ve gerçekleştirilen analizlerle ilgili bilgiler yer almaktadır.

#### 3.1. Araştırmanın Türü

Bu çalışmada, örneklem büyüklüğü, yetenek dağılımı, ortak madde oranı, ortak madde tipi ve ortak madde güçlük dağılımı koşullarına göre eşitleme yöntemlerinin karşılaştırılması amaçlanmıştır. Bu amaçla simülasyon veri kullanılarak en az hata veren yöntemin belirlenmesi amaçlanmıştır. Belirtilen koşullar altında kuramsal çalışmalara katkıda bulunulması amaçlandığından, bu araştırmanın temel araştırma niteliği taşıdığı söylenebilir.

#### 3.2. Araştırmada Kullanılan Eşitleme Deseni

Bu çalışmada test formlarının eşitlenmesinde denk olmayan gruplarda ortak madde/test deseni (NEAT) kullanılmıştır. Bunun için test formları (X ve Y) ve ortak test formları oluşturulmuştur. Bu çalışmada X formu eski test formunu, Y formu ise eşitlenecek (yeni) test formunu ifade etmektedir. X formunu alan grup Grup 1, Y formunu alan grup ise Grup 2 ile temsil edilmektedir. Gruplardaki birey sayısı ve ortak testte yer alan madde sayısı araştırmanın amacı doğrultusunda farklılaşmaktadır. Araştırmada kullanılan eşitleme deseni Tablo 3.1'de yer almaktadır.

**Tablo 3.1: Araştırmanın Eşitleme Deseni**

Örneklem	X Formu	Ortak Test	Y Formu
Grup 1	✓	✓	
Grup 2		✓	✓

*Not: Test maddelerinin ve bireylerin üretilmesine ilişkin ayrıntılı bilgiler "3.4. Verilerin Üretilmesi" başlığı altında yer almaktadır.*

#### 3.3. Simülasyon Koşulları

Holland, Dorans ve Petersen (2006), başarılı bir eşitleme için eşitleme deseni, ana testlerin kalitesi, ortak testin özellikleri, örneklem büyüklüğü, grupların yetenek dağılımı ve eşitleme yöntemi seçiminin temel özellikler olduğunu belirtmiştir. Bu çalışmada, örneklem büyüklüğü, yetenek dağılımı, ortak madde oranı, ortak madde tipi ve ortak madde güçlük dağılımı değişkenleri incelenmiştir. Araştırmada

simülasyon verisinin kullanılmasının nedeni ise gerçek veride ele alınan koşulların tümünün bulunmamasıdır. Koşullara ilişkin bilgiler Tablo 3.2’de verilmiştir.

**Tablo 3.2: Araştırmanın Simülasyon Koşulları**

Faktörler	Koşullar	Koşul Sayısı
Örneklem Büyüklüğü	500-1500-5000	3
Gruplar arası yetenek dağılımı (Eski&Yeni)	(0;1)&(0.05;1)-(0;1)&(0.5;1)	2
Ortak Madde Tipi	İç-Dış	2
Ortak Madde Oranı	%20-%30-%40	3
Ortak Madde Güçlük Dağılımı	Mini-Midi	2

Tablo 3.2’de görüldüğü gibi örneklem büyüklüğü 3, gruplar arası yetenek dağılımı 2, ortak madde tipi 2, ortak madde oranı 3 ve ortak madde dağılımı 2 farklı koşul içermek üzere toplamda 72 ( $3 \times 2 \times 2 \times 3 \times 2$ ) koşul ele alınmıştır. Araştırmada kullanılan 5 eşitleme yönteminin performansını karşılaştırmak için 360 ( $72 \times 5$ ) veri seti oluşturulmuştur. Araştırma sonuçlarının tutarlı ve genellenebilir olabilmesi için her veri seti üzerinde 100 tekrar yapılmıştır. Bu amaçla R programında yazılan kod EK-3’de verilmiştir

### 3.3.1. Örneklem Büyüklüğü

Eşitlemede örneklem büyüklüğünün eşitleme sonuçlarının doğruluğu ve kesinliğini arttırdığı çeşitli araştırmalar ile ortaya konmuştur (Cui & Kolen, 2008; Godfrey, 2007; Hanson & Beguin, 2002; Meng, 2012; Norman Dvorak, 2009; Nozawa, 2009; Qu, 2007; Zhao, 2008). Kolen ve Brennan (2004), MTK gerçek puan eşitleme yöntemlerinde NEAT deseninde 3 PLM’de her bir form için örneklem büyüklüğünün 1500 olması gerektiğini ve Skaggs ve Lissitz (1986) ise, yatay eşitleme deseninde 3PLM için örneklem büyüklüğünün en az 1000 olması gerektiğini belirtmiştir. Norman Dvorak (2009) ve Meng (2012) Kernel eşitleme ile MTK gerçek puan eşitleme yöntemlerini karşılaştırdıkları çalışmalarda örneklem büyüklüğü arttıkça, eşitleme yöntemlerinden daha iyi sonuçlar elde ettiklerini belirtmişlerdir. Bu çalışmada örneklem büyüklüğü 500 (küçük), 1500 (orta) ve 5000 (büyük) olarak belirlenmiştir. Ayrıca her iki test formu için örneklem büyüklüğü eşit olarak ele alınmıştır.

### 3.3.2. Gruplararası Ortalama Yetenek Farkı

Alanyazında, eşitlemede grupların yetenek dağılımlarının benzer olması gerektiği belirtilmektedir. Wang, Lee, Brennan ve Kolen (2008), grupların yetenek dağılımlarının ortalaması arasındaki farkı 0.05 ve 0.1 aralığında ise geniş ve 0.25 ve daha büyük değer aralığında ise çok geniş olarak adlandırılmışlardır. Bu çalışmada yukarıda belirtilen değerler göz önünde bulundurularak grupların ortalama yetenek farkının az olduğu ve çok geniş olduğu koşul olmak üzere iki koşul belirlenmiştir. Yeni formu alan grubun yetenek dağılımı  $N(0.05,1)$  ve  $N(0.5,1)$ , eski formu alan grubun yetenek dağılımı ise  $N(0,1)$  olarak standart normal dağılım gösterecek şekilde veriler üretilmiştir.

### 3.3.3. Ortak Madde Tipi ve Oranı

NEAT deseninde ortak test formu iç ve dış ortak test olmak üzere ikiye ayrılır. Kernel eşitleme ile MTK'ya dayalı eşitleme yöntemlerinin karşılaştırıldığı çalışmalarda dış ortak test kullanılarak karşılaştırma yapıldığından, bu çalışmada hem iç hem de dış ortak test kullanılmıştır. Ayrıca ortak madde setinin tüm testteki madde sayısının %20'si kadar olması gerektiği (Angoff, 1971; Budescu, 1985; Kolen & Brennan, 2004) önerilmektedir. Hambleton ve diğerleri (1991) ise ortak maddeler için gereksinim duyulan madde sayısının, testteki madde sayısının yaklaşık olarak %20 - %25'i kadar olması gerektiğini ifade etmişlerdir. Ortak madde sayısının artmasının eşitleme hatasını azalttığı yapılan çalışmalarla gösterilmiştir (Godfrey, 2007; Keller III, 2007; Kim & Cohen, 1998; Kim & Cohen, 2002; Meng, 2012; Norman Dvorak, 2009; Wingersky, Cook & Eignor, 1987). Bu nedenle bu çalışmada eşitleme yöntemlerinde ortak madde oranının etkisini görebilmek için %20, %30 ve %40 olmak üzere üç farklı ortak madde oranı seçilmiştir.

### 3.3.4. Ortak Madde Güçlük Dağılımı

Birçok araştırmacı (Budescu, 1985; Kolen, 1988; Kolen, 2007; Petersen, Kolen & Hoover, 1989; von Davier, Holland & Thayer, 2004 ) ortak madde setinin toplam testin küçük bir versiyonu olması gerektiğini ifade etmiştir. Ancak ortak madde setinin testin bütününden farklı güçlük ve standart sapmaya sahip olduğunda eşitleme hataları ve yanlılığın değişip değişmediği inceleyen araştırmalarda (Antal, Proctor & Melican, 2014; Fitzpatrick & Skorupski, 2016; Hagge, 2010; Kim, 2014;

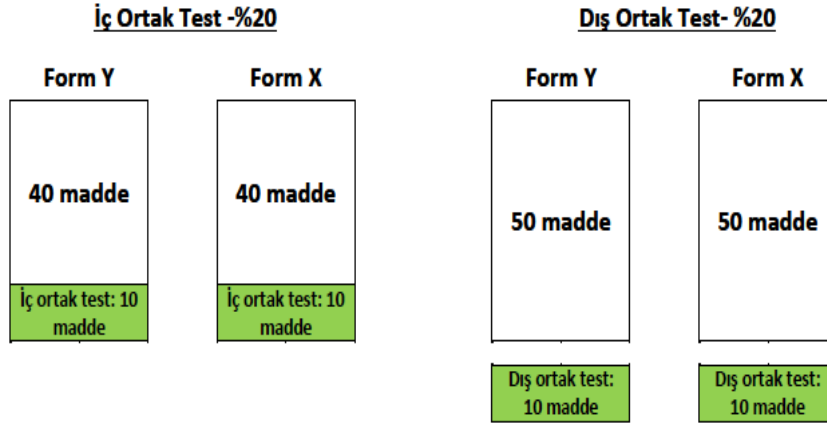


Liu, Sinharay, Holland, Feigenbaum & Curley, 2011; Sinharay & Holland, 2006a; Sinharay & Holland, 2006b; Sinharay & Holland, 2007; Sinharay, Haberman, Holland & Lewis, 2011) farklı güçlük dağılımına sahip ortak testlerin daha iyi sonuçlar verdiği sonucuna ulaşılmıştır. Crocker ve Algina (1986), ortak madde test deseninde MTK gerçek puan eşitleme yöntemleri için ortak test ile toplam testin aynı özelliği/yeteneği ölçmesi gerektiğini, geleneksel eşitleme yöntemlerinden doğrusal ve eşit yüzdelikli eşitleme için ise böyle bir gerekliliğin olmadığını belirtmiştir. Sinharay ve Holland (2006a, 2006b, 2007) ortak testin güçlük dağılımının toplam test ile aynı olması gerektiği konusunda bir kanıt olmadığını ve bu durumun çok kısıtlayıcı olduğunu, ortak testin kapsam ve ortalama güçlüğüne eşitlenen test ile aynı, ancak ortak maddelerin madde güçlük dağılımının toplam testin güçlüğünden daha küçük olmasının daha iyi eşitleme sonuçlarını verebileceğini ifade etmişlerdir. Sinharay ve Holland'a (2006) göre ortak testler üç farklı şekilde oluşturulabilir. Eğer ortak test, toplam test ile aynı kapsam ve istatistiksel açıdan benzer ise mini-test; ortak test kapsam açısından toplam test ile aynı, ancak madde güçlükleri açısından farklılaşıyorsa ve madde güçlüklerinin hepsi orta güçlükte ise midi-test; maddelerin güçlükleri dağılımı mini-test ile midi-test arasında değer alıyorsa bu durumda da semi-midi test olarak adlandırmışlardır. Bu çalışmada madde güçlük dağılımları açısından mini ve midi ortak test kullanılmıştır.

### **3.4. Verilerin Üretilmesi**

Geniş ölçekli testlerde kullanılan pek çok test çoktan seçmeli formattadır ve iki kategorili (dichotomously) olarak puanlanır. Bazı çalışmalarda çok kategorili maddelerin analizleri yapılırken maddeler sanki iki kategoriliymiş gibi ele alınarak puanlama yapıldığı sıklıkla görülmektedir (Potenza ve Dorans, 1995). Veri setleri üretilmeden önce ana testler ve ortak testler için madde parametrelerine bu maddeleri cevaplayacak bireylere ait yetenek parametreleri için, ilgili araştırmalar incelenerek, madde ve yetenek parametre değerleri türetilmiştir. Çalışmada, test uzunluğu 50 olarak belirlenmiş ve 3 PL modele uygun veriler R 3.2.5 yazılımı yardımıyla üretilmiştir. NEAT deseni kullanıldığından, test ve ortak test formları oluşturulmuştur. Test ve ortak test formlarına ait madde ayırt edicilik parametresi ( $a$ ) tek biçimli (uniform) dağılımdan (0.5 - 0.2) aralığında,  $c$  parametresi tek biçimli (uniform) dağılımdan (0.05 - 0.20) aralığında ve  $X$  ve  $Y$  formlarına ait güçlük

parametresi olan  $b$  parametresi ise normal dağılımdan  $(0,1)$  üretilmiştir. Ortak test formunun mini olduğu koşulda toplam testle aynı güçlük parametresine sahip olacak şekilde yani ortalaması ve standart sapması aynı, midi olduğu koşulda ise ortalaması toplam test ile aynı, ancak standart sapması 0.2 alınarak oluşturulmuştur. İç ve dış ortak testte test formlarının nasıl oluşturulduğuna ilişkin bir gösterim %20 ortak madde örneğiyle Şekil 3.1’de verilmiştir.



**Şekil 3.1. İç ve Dış Ortak Testin Test Formlarında Gösterimi**

### 3.5. Eşitleme Süreci ve Verilerin Analizi

Araştırmada, beş farklı eşitleme yönteminin performansı; örneklem büyüklüğü (3 koşul), yetenek dağılımı (2 koşul), ortak madde tipi (2 koşul), ortak madde oranı (3 koşul) ve ortak madde güçlük dağılımı (2 koşul) olmak üzere toplam 72 koşulda incelenmiştir. Bu koşullar altında Kernel son tabakalama eşit yüzdelikli (ideal h), Kernel son tabakalama doğrusal (geniş h), Kernel zincirleme eşit yüzdelikli (ideal h), Kernel zincirleme doğrusal (geniş h) ve MTK gerçek puan (Haebara) eşitleme yöntemleri karşılaştırılmıştır.

Hem Kernel hem de MTK gerçek puan eşitleme yöntemlerinde kullanılan test formlarındaki maddeler için verilerin üretilmesi aşamasında belirtilen değerlerle madde ve birey parametre değerleri simüle edildikten sonra R programında “*irtoys*” paketi (Partchev, 2016) kullanılarak MTK 3 PL modele uyumlu iki kategorili (1-0) cevaplar türetilmiştir. Kernel eşitleme yöntemleri ile test formlarının eşitlenmesi için “*kequate*” paketi (Andersson, Branberg & Wiberg, 2013) kullanılmıştır. Kernel eşitleme yöntemleri için elde edilen veri setlerine ilk aşamada log lineer modelle

düztünleřtirme yapılmıřtır. Denk olmayan gruplarda ortak madde test deseninde iç ortak test kullanıldıđında yapısal sıfırlar durumu ortaya çıkmaktadır. Gözlenen puan dađılımlarına iki deđiřkenli log lineer öndüztünleřtirme yapıldıđında, yapısal sıfırlar genellikle çok küçük olsa da bazı pozitif olasılıklar verir (Kim, 2014). Brennan, Wang, Kim ve Soel (2009) yapısal sıfırlar durumu için bir yaklařım önermiřlerdir. Bu yaklařımda, toplam test formu X, ortak test V ve test formu X\* řeklinde gösterilirse; toplam test puanı  $X=X^*+V$  olur. Yapısal sıfırlar durumunun önüne geçmek için  $X \times V$  matrisine yerine  $X^* \times V$  matrisine iki deđiřkenli düztünleřtirme uygulanmaktadır. Bu arařtırmada da bu yaklařım kullanılmıřtır. Dıř ortak test formunda bu durum ortaya çıkmamaktadır.

İkinci ařamada, ilk ařamada elde edilen düztünleřtirilmiř puan dađılımları, puan olasılık dađılımlarını kestirmek için kullanılmıřtır. Ancak bu durum zincirleme ve son tabakalama eřitleme yöntemleri için farklıdır. Zincirleme eřitleme yöntemlerinde (dođrusal ve eřit yüzdelikli) öncelikle iki ayrı tek grup deseni oluřturulur. Daha sonra X formundan ortak test formuna (A) bađlanılır ve sonra ortak test formundan (A) Y formuna bađlantı kurulur. Son tabakalama eřitleme yöntemlerinde (dođrusal ve eřit yüzdelikli) ise öndüztünleřtirme yapıldıktan sonra sentetik evren (T) üzerinde X ve Y formu için marjinal dađılımlar kestirilir. Son tabakalama eřitleme yöntemlerinde NEAT desenindeki her iki grup birleřtirilerek tek bir grup oluřturulur ve bu oluřan gruba sentetik grup adı verilir. Sentetik grup içindeki her bir grup  $w_1$  ve  $w_2$  ile ađırlıklandırılır. von Davier ve diđerlerinin (2004) Kernel son tabakalama eřitleme yöntemlerinde ađırlıklandırmanın etkisini inceledikleri çalıřmada  $w_1=0$ ,  $w_1=1/2$ , ve  $w_1=1$  olmak üzere 3 kořul için X'e ait yığılmalı dađılım fonksiyonlarını hesaplamıř, ancak w'nin seğıiminin etkisinin oldukça küçük olduđunu görmüřlerdir. Grupların ađırlıklandırma ile ilgili genel olarak ađırlıklandırmanın eřit olduđu  $w_1=w_2=0.5$  durum kullanılmaktadır (Kolen & Brennan, 2014). Bu çalıřmada da son tabakalama eřit yüzdelikli ve son tabakalama dođrusal eřitleme yöntemleri için ađırlıklandırmanın eřit olduđu durum kullanılmıřtır. Üçüncü ařamada ise kesikli dađılımları süreklileřtirmek için Gauss Kernel düztünleřtirme kullanılmıř ve dördüncü ařamada ise test formları eřitlenmiřtir. Kernel eřitleme hem eřit yüzdelikli hem de dođrusal eřitleme fonksiyonlarını içerir. Kernel eřitlemede h parametresi ideal olduđunda eřit yüzdelikli, geniř olduđunda ise dođrusal eřitleme fonksiyonuna göre eřitleme

fonksiyonları elde edilir. Bu çalışmada doğrusal ve eşit yüzdeliği eşitleme yöntemleri için  $h$  parametresinin seçimi “*kequate*” paketi tarafından belirlenmiştir.

MTK gerçek puan eşitleme için test formlarının cevaplarına ait 3 PL modele uygun olarak madde ve yetenek parametre kestirimleri ise R programında “*ltm*” paketi (Rizopoulos, 2015) ile yapılmıştır. Madde parametre kestirimleri için Marjinal En Çok Olabilirlik, yetenek parametrelerinin kestirimi için ise Beklenen Sonsal Dağılım yöntemleri kullanılarak kestirimler yapılmıştır (Rizopoulos, 2015). Daha sonra kestirilen parametreler, MTK gerçek puan eşitleme için ayrı kalibrasyon yöntemlerinden Haebara yöntemiyle aynı ölçeğe yerleştirilerek gerçek puan eşitleme yapılmıştır. Madde parametrelerine ait kalibrasyonlar ve gerçek puan eşitleme için “*plink*” paketi (Weeks, 2010) kullanılmıştır.

Ayrıca araştırma kapsamında ele alınan değişkenlerin eşitleme yöntemleri üzerindeki ortak etkisini belirleyebilmek için “*lattice*” paketi (Sarkar, 2017) kullanılmıştır. Böylece değişkenlerin değerlendirme kriterlerine göre eşitleme yöntemleri üzerindeki etkilerini ayrı ayrı belirlemenin yanında, bu değişkenlerin etkileşimlerinin ortak etkisinde incelenmiştir.

### **3.6. Değerlendirme Kriterleri**

Eşitleme desenleri ve yöntemleri eşitleme hatasını en küçük yapacak şekilde seçilmektedir. Eşitlemede, sistematik ve random olmak üzere iki hata türü bulunmaktadır (Kolen, 1988; Kolen & Brennan, 2014).

Random hata, özellikle küçük örneklemeler kullanıldığı zaman ortaya çıkar. Geniş örneklemeler kullanılması ve uygun eşitleme deseninin seçimiyle random hata azaltılabilir. Sistematik hata ise eşitleme yöntemine ait varsayımların ve eşitleme koşullarının ihlal edilmesi sonucunda oluşur. Tek grup deseninde, pratik etkisi ve yorgunluk, random grup deseninde ise sarmal sürecin uygulanmaması sistematik hatayı ortaya çıkarır. Sistematik hata özellikle denk olmayan grup deseninde, ortak maddelerin toplam testi kapsam ve istatistiksel olarak temsil etmemesi, her iki formda da aynı yerde olmaması veya grupların birbirinden çok farklı olması durumlarında oluşur. Random hata eşitlemenin standart hatasıyla hesaplanabilir (Kolen, 1988; Kolen & Brennan, 2004).

Harris ve Course (1993) farklı eşitleme kriterlerini değerlendirdiği çalışmalarında, tek bir eşitleme kriterin bütün eşitleme yöntemleri için uygun olmayacağını ve birden fazla kriterin kullanılması gerektiğini belirtmişlerdir. Bu kriterler RMSE, RMSD, WRMSE (weighted RMSE), BIAS, WAB (weighted absolute bias), SEE, WSE (weighted Standard error), WASE (the weighted average standard error of equating), I. sıra ve II. sıra eşitlik olarak belirtilebilir.

Farklı eşitleme yöntemlerinin karşılaştırıldığı çalışmalarda en büyük problem, eşitleme yöntemlerini değerlendirmek için kullanılacak olan gerçek eşitleme fonksiyonuna karar vermektir. Genellikle, gerçek eşitleme fonksiyonu olarak eşit yüzdellikli eşitleme yöntemi kullanılmaktadır (Wang ve diğerleri, 2008).

Kernel eşitleme ile ilgili yapılan çalışmalar incelendiğinde, eşitlenmiş puanların gerçek değerinin elde edilmesi için farklı eşitleme yöntemlerinin kullanıldığı görülmüştür. Godfrey (2007), Livingston (1993), Ricker ve Von Davier (2007), Qu (2007) ile von Davier ve diğerleri'nin (2006) eşit yüzdellikli eşitleme yöntemini; Kim'in (2014) ise MTK gözlenen puan eşitleme yöntemini kullandığı görülmektedir. Bu çalışmada kriter eşitleme olarak eşit yüzdellikli eşitleme yöntemi (düzgünleştirme yapılmadan) kullanılmıştır.

**Eşit Yüzdellikli eşitleme:** Eşit yüzdellikli eşitlemede aynı puan dağılımına sahip olmayan formların dağılımları eşitlenir. Bunun için ilk olarak her bir formun yığılmalı frekansı hesaplanarak tablolaştırılır ve bu yığılmalı frekanslara göre aynı yüzdellik puanlara karşılık gelen puanlar eşitlenir. Eşit yüzdellikli eşitlemeyle aynı yüzdellik sırasındaki eşitlenmiş puanlara sahip bireylerin aynı yetenek düzeyinde olduğu kabul edilir. Eşit yüzdellikli eşitlemede doğrusal eşitlemeye göre, eşitlenmiş puan dağılımları daha fazla benzerlik gösterir. Eşit yüzdellikli eşitlemeyle formlar eşitlenirse formlardan elde edilen puanların ortalama, standart sapma ve dağılım şekli (basıklık, çarpıklık, vb.) benzer olur (Kolen, 1988).

Eğer test formlarından elde edilen puanlar aynı standart sapmaya sahipse, ortalama ve doğrusal eşitleme; puan dağılımları aynıysa doğrusal ve eşit yüzdellikli eşitleme aynı sonuçları verecektir. Eşit yüzdellikli eşitleme doğrusal ve ortalama eşitlemeye göre daha fazla örnekleme ihtiyaç duyar. Ayrıca denk olmayan grupta ortak madde test deseninde ortalama ve doğrusal eşitlemeye göre eşit yüzdellikli eşitleme daha karmaşık işlemler gerektirir (Crocker & Algina, 1986; Kolen, 1988;

Kolen & Brennan, 2004). Eşit yüzdelikli eşitleme yöntemi için “*equate*” paketi (Albano, 2010) kullanılmıştır.

Çalışmada eşitlenmenin tutarlılığını incelemek için eşitleme hatası (RMSD) ve eşitlemenin standart hatası (SEE) hesaplanmıştır. Değerlendirme kriterlerinin elde edilmesi için aşağıda yer alan denklemler kullanılmıştır.

**RMSD (Ortalama Hatanın Kare Kökü Farkı) veya Hata:** Her bir puan için, eşitleme yöntemi ile kriter olarak kullanılan eşitleme yönteminden elde edilen eşitlenmiş puanlar arasındaki farkın ortalamasına ve bu farkın standart sapmasıyla toplamının kareköküne eşittir. Mao, von Davier ve Rupp (2005) ve Tsai (1995), RMSE indeksini adapte ederek RMSD indeksini oluşturmuşlardır. RMSD indeksi, eşitleme kriteri ile karşılaştırıldığında eşitleme sonuçlarının ne kadar yanlı ve doğru olduğunu yansıtır (Qu, 2007).

$$\text{RMSD} = \sqrt{\bar{d}^2 + sd_d^2}$$

(Eşitlik 3.1)

$\bar{d}$  : Her bir eşitlenmiş puan için kriter eşitleme ile eşitleme yapılan yöntem arasındaki farkın ortalaması,

$sd_d$ : Elde edilen farkın standart sapması.

**Standart Hata (SE):** Random hatayı veren standart hata, her bir kestirilen değer ile elde edilen ortalama değer arasındaki farkın kareleri toplamının tekrar (replikasyon) sayısına oranının kareköküne eşittir.

$$\text{SE}[\hat{e}_Y(x_i)] = \sqrt{\frac{1}{R} \sum_{1}^R [\hat{e}_Y(x_i) - \bar{\hat{e}}_Y(x_i)]^2}$$

(Eşitlik 3.2)

$\hat{e}_Y(x_i)$ : Her bir tekrar için elde edilen eşitlenmiş puan

$\bar{\hat{e}}_Y(x_i)$ : Tekrar sonucu elde edilen eşitlenmiş puanların ortalaması

R : Tekrar sayısı.

## 4. BULGULAR VE TARTIŞMA

Bu bölümde, alt problemlere göre verilmiş araştırma bulguları eşitlemenin hatası (RMSD) ve eşitlemenin standart hatası (SEE) kriterlerine göre incelenmiştir. Bütün alt problemlerin bulguları verildikten sonra tartışma alt başlığı altında çalışmanın bulguları ilgili alanyazın göz önünde bulundurularak tartışılmıştır.

### 4.1. Alt Problem 1'e İlişkin Bulgular ve Yorumlar

“Testler, Kernel son tabakalama doğrusal, Kernel zincirleme doğrusal, Kernel son tabakalama eşit yüzdelikli, Kernel zincirleme eşit yüzdelikli ve MTK gerçek puan eşitleme yöntemleriyle küçük örneklerde (500) eşitlendiğinde eşitlenmiş puanlara ilişkin;

#### 4.1.a. Eşitleme hatası çalışmada ele alınan koşullara göre nasıl değişmektedir?

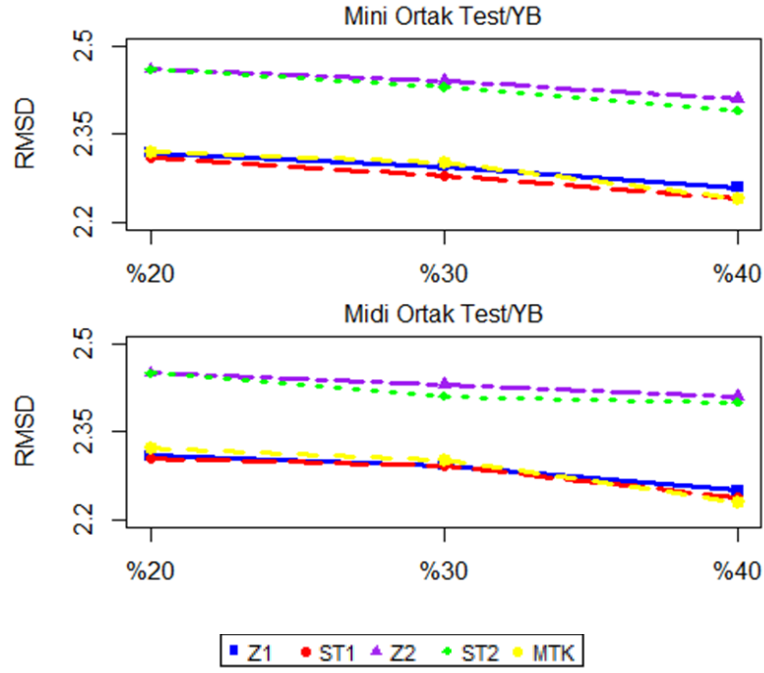
Bu alt problemin çözümünde eşitleme yöntemlerinin karşılaştırmalarında eşitleme hatalarından RMSD kullanılmıştır. RMSD değerleri; her bir puan için, eşitleme yöntemi ile kriter olarak kullanılan eşitleme yönteminden elde edilen eşitlenmiş puanlar arasındaki farkın ortalamasına ve bu farkın standart sapmasıyla toplamının kareköküne eşittir.

Alt problem 1.a'ya ilişkin bulguların anlaşılmasını kolaylaştırmak için araştırma kapsamında ele alınan koşullara ait eşitleme hatası sonuçları iç ve dış ortak test için ayrı ayrı yapılandırılmıştır. 500 örneklem için ele alınan koşullara göre eşitleme hatalarına ait grafikler Şekil 4.1-Şekil 4.4 arasında verilmiştir. Bu grafikler EK-5'de yer alan tabloya göre oluşturulmuştur.

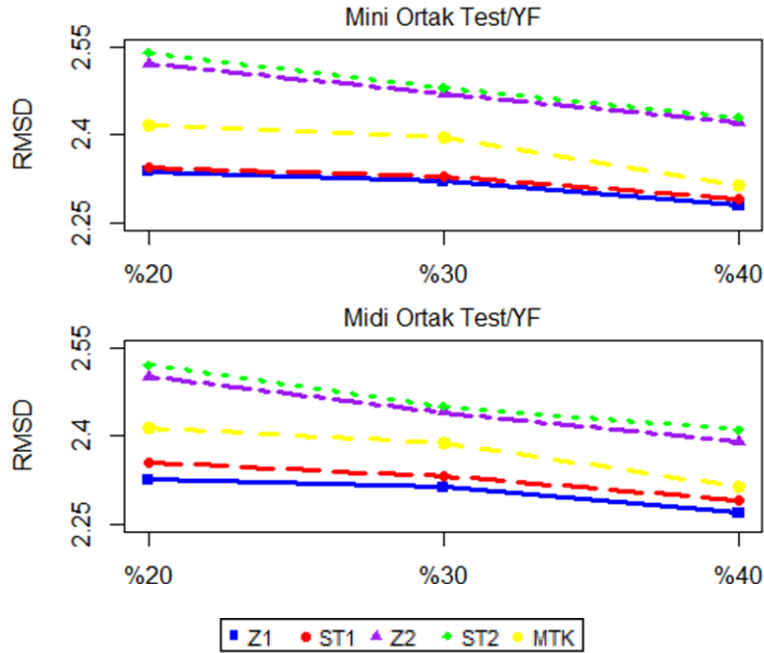
Z1 Zincirleme eşit yüzdelikli eşitleme, Z2 zincirleme doğrusal eşitleme, ST1 son tabakalama eşit yüzdelikli eşitleme, ST2 son tabakalama doğrusal eşitleme ve MTK MTK gerçek puan eşitleme yöntemini belirtmektedir.

#### *İç ortak teste göre eşitleme hatası nasıl değişmektedir?*

İç ortak testte, ortak madde oranının %20, %30 ve %40 olduğu koşullarda gruplar arası yetenek dağılımı benzer olduğunda elde edilen eşitleme hatasına ait grafikler Şekil 4.1'de, yetenek dağılımı farklı olduğunda elde edilen eşitleme hatasına ait grafikler Şekil 4.2.'de verilmiştir.



**Şekil 4.1. 500 Örneklem için İç Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen RMSD Değerleri**



**Şekil 4.2. 500 Örneklem için İç Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen RMSD Değerleri**

Ortak madde oranının %20 olduğu durumda;

Gruplar arası yetenek dağılımı benzer olduğunda (YB) en düşük RMSD değeri mini ve midi ortak test koşulunda son tabakalama eşit yüzdelikli (ideal h) eşitleme; en yüksek RMSD değeri ise mini ortak test koşulunda zincirleme doğrusal (geniş



h) eşitleme ve midi ortak test koşulunda MTK gerçek puan eşitleme yönteminden elde edilmiştir. Gruplar arası yetenek dağılımı farklı olduğunda (YF) en düşük RMSD değerinin her iki ortak test koşulunda zincirleme eşit yüzdellikli (ideal h) eşitleme ve en yüksek RMSD değerinin MTK gerçek puan eşitleme yönteminden elde edildiği görülmektedir.

Ortak madde oranının %30 olduğu durumda;

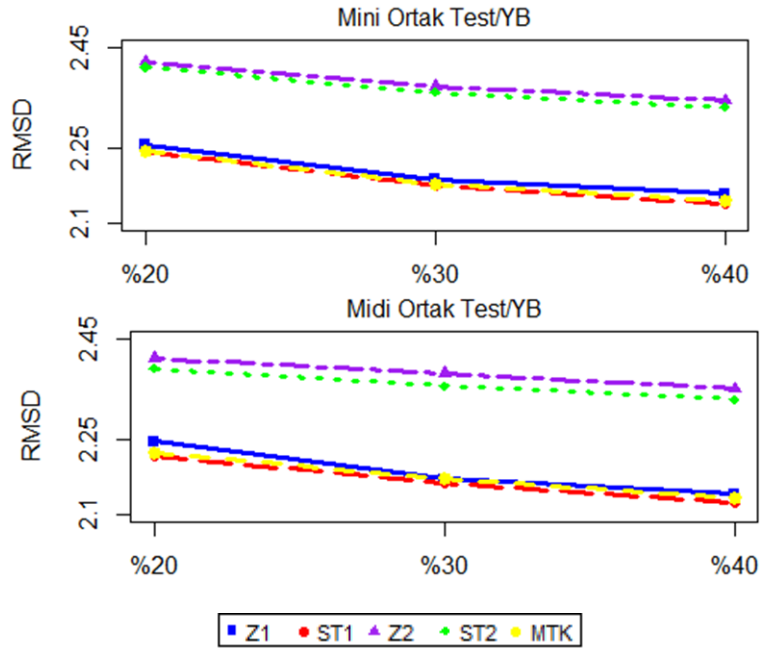
Gruplar arası yetenek dağılımı benzer olduğunda en düşük RMSD değeri her iki ortak test koşulunda son tabakalama (ideal h) eşitleme, en yüksek RMSD ise zincirleme (geniş h) eşitleme yönteminden elde edilmiştir. Gruplar arası yetenek dağılımı farklı olduğunda en düşük RMSD değerinin her iki ortak test koşulunda zincirleme (ideal h) eşitleme; en yüksek RMSD değerinin ise MTK gerçek puan eşitleme yönteminden elde edildiği görülmektedir.

Ortak madde oranının %40 olduğu durumda;

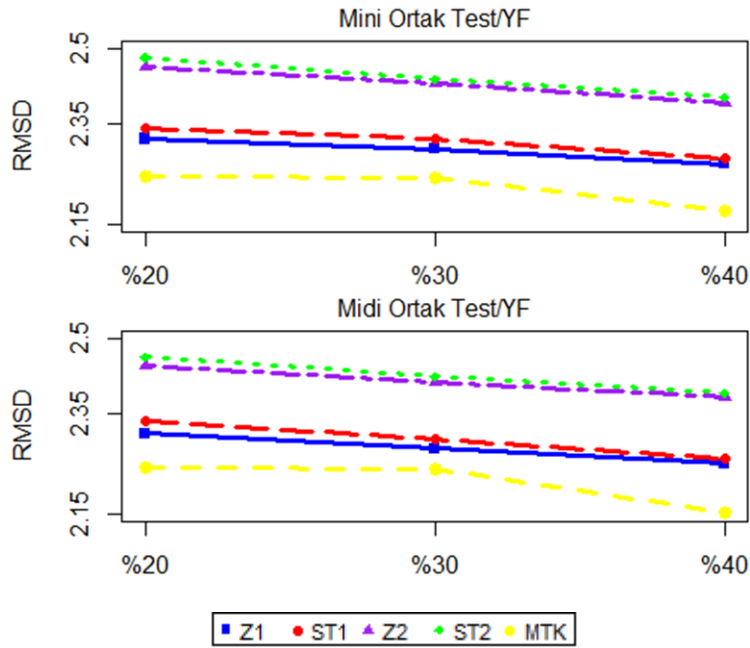
Gruplar arası yetenek dağılımı benzer olduğunda en düşük RMSD değeri her iki ortak test koşulunda son tabakalama (ideal h) eşitleme ve en yüksek RMSD değeri ise MTK gerçek puan eşitleme yönteminden elde edilmiştir. Gruplar arası yetenek dağılımı farklı olduğunda en düşük RMSD değerinin her iki ortak test koşulunda zincirleme (ideal h) eşitleme yönteminden ve en yüksek RMSD değerinin ise MTK gerçek puan eşitleme yönteminden elde edildiği görülmektedir.

### ***Dış ortak teste göre eşitleme hatası nasıl değişmektedir?***

Dış ortak testte, ortak madde oranının %20, %30 ve %40 olduğu koşullarda gruplar arası yetenek dağılımı benzer olduğunda elde edilen eşitleme hatasına ait grafikler Şekil 4.3'de, yetenek dağılımı farklı olduğunda elde edilen eşitleme hatasına ait grafikler Şekil 4.4 'de verilmiştir.



**Şekil 4.3. 500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen RMSD Değerleri**



**Şekil 4.4. 500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen RMSD Değerleri**

Ortak madde oranının %20 olduğu durumda;

Gruplar arası yetenek dağılımı benzer olduğunda en düşük RMSD değeri her iki ortak test koşulunda son tabakalama (ideal h) eşitleme ve en yüksek RMSD değeri MTK gerçek puan eşitleme yönteminden elde edilmiştir. Gruplar arası yetenek

dağılımı farklı olduğunda en düşük RMSD değerinin her iki ortak test koşulunda zincirleme (ideal h) eşitleme yönteminden, en yüksek RMSD değerinin ise MTK gerçek puan eşitleme yönteminden elde edildiği görülmektedir.

Ortak madde oranının %30 olduğu durumda;

Gruplar arası yetenek dağılımı benzer olduğunda en düşük RMSD değeri her iki ortak test koşulunda son tabakalama (ideal h) eşitleme yönteminden ve en yüksek RMSD değeri ise MTK gerçek puan eşitleme yönteminden elde edilmiştir. Gruplar arası yetenek dağılımı farklı olduğunda en düşük RMSD değerinin her iki ortak test koşulunda zincirleme (ideal h) eşitleme ve en yüksek RMSD değerinin her iki ortak testte MTK gerçek puan eşitleme yönteminden elde edildiği görülmektedir.

Ortak madde oranının %40 olduğu durumda;

Gruplar arası yetenek dağılımı benzer olduğunda en düşük RMSD değerinin her iki ortak test koşulunda son tabakalama (ideal h) eşitleme ve en yüksek RMSD değerini ise her iki ortak test için MTK gerçek puan eşitleme yöntemi vermiştir. Gruplar arası yetenek dağılımı farklı olduğunda en düşük RMSD değeri her iki ortak test koşulunda zincirleme (ideal h) eşitleme yönteminden, en yüksek RMSD değerinin ise her iki ortak testte MTK gerçek puan eşitleme yönteminden elde edildiği görülmektedir.

Bütün eşitleme yöntemlerinde iç ve dış ortak testte ortak madde oranı arttıkça RMSD değerlerinin azaldığı görülmektedir. Ortak madde oranı arttıkça iç ortak testte son tabakalama eşitleme yöntemlerinde, dış ortak testte ise eşit yüzdelikli eşitleme yöntemlerinde ve MTK gerçek puan eşitlemede hata miktarları daha çok azalmıştır. Ayrıca dış ortak testten elde edilen eşitleme hatalarının iç ortak teste göre daha düşük olduğu görülmektedir. Gruplar arası yetenek dağılımının benzer ve farklı olduğu koşullarda, dış ortak testin midi test olduğu durumda elde edilen hata değerlerinin, mini test koşuluna göre daha düşük olduğu görülmektedir. Ancak iç ortak test midi test olduğunda hata değerleri bazı durumlarda mini testten elde edilen değerlerle benzer, bazı durumlarda ise artmış veya azalmıştır. Gruplar arası yetenek dağılımının benzer olduğu koşuldaki RMSD değerleri yetenek dağılımlarının farklı olduğu koşula göre daha düşük elde edilmiştir. Gruplar arası yetenek dağılımı farklı olduğunda hem iç hem de dış ortak test için midi ortak testten daha düşük eşitleme hatası elde edilmiştir. En düşük eşitleme hatasını;

gruplar arası yetenek dağılımı benzer olduğu durumda son tabakalama eşit yüzdelikli eşitleme, yetenek dağılımının farklı olduğu durumda ise zincirleme eşit yüzdelikli eşitleme vermiştir. Doğrusal eşitleme yöntemlerinin, eşit yüzdelikli eşitleme yöntemlerine göre daha yüksek hata verdiği görülmektedir.

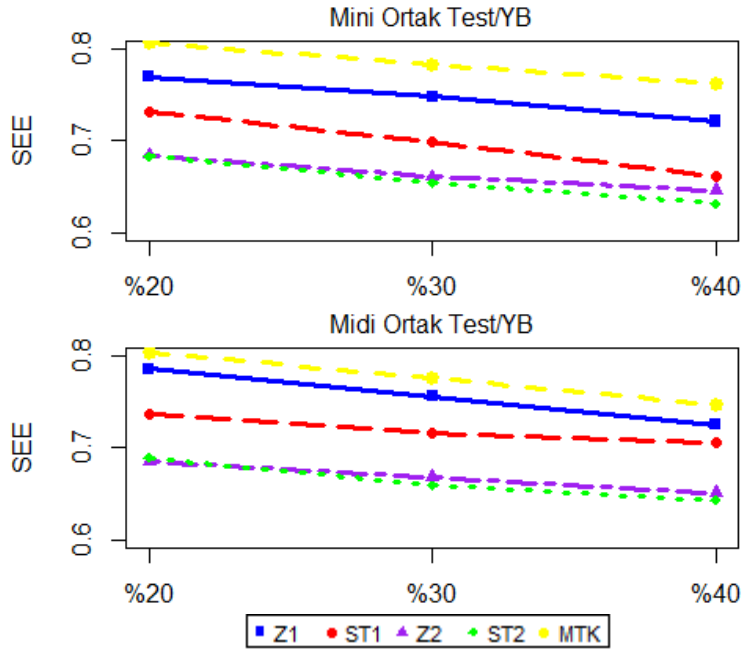
#### **4.1.b. Eşitlemenin standart hatası ve her bir puan için elde edilen standart hata çalışmada ele alınan koşullara göre nasıl değişmektedir?**

Bu alt problemin çözümünde eşitleme yöntemlerinin karşılaştırmalarında eşitlemenin standart hatası (SEE) kullanılmıştır.

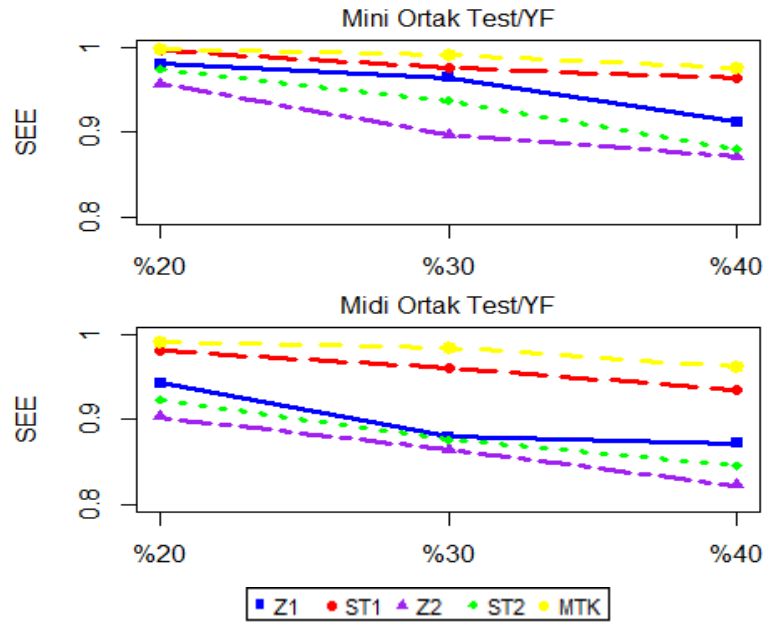
Alt problem 1.b'ye ilişkin bulguların anlaşılmasını kolaylaştırmak için araştırma kapsamında ele alınan koşullara ait eşitlemenin standart hatası sonuçları iç ve dış ortak test için ayrı ayrı yapılandırılmıştır. 500 örneklem için ele alınan koşullara göre eşitlemenin standart hatasına ait grafikler Şekil 4.5 - Şekil 4.8 arasında verilmiştir. Bu grafikler EK-8'de yer alan tabloya göre oluşturulmuştur.

#### ***İç ortak teste göre standart hata nasıl değişmektedir?***

İç ortak testte, ortak madde oranının %20, %30 ve %40 olduğu koşullarda gruplar arası yetenek dağılımı benzer olduğunda elde edilen eşitlemenin standart hatasına ait grafikler Şekil 4.5'de, yetenek dağılımı farklı olduğunda elde edilen eşitlemenin standart hatasına ait grafikler Şekil 4.6'da verilmiştir.



**Şekil 4.5. 500 Örneklem için İç Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen SEE Değerleri**



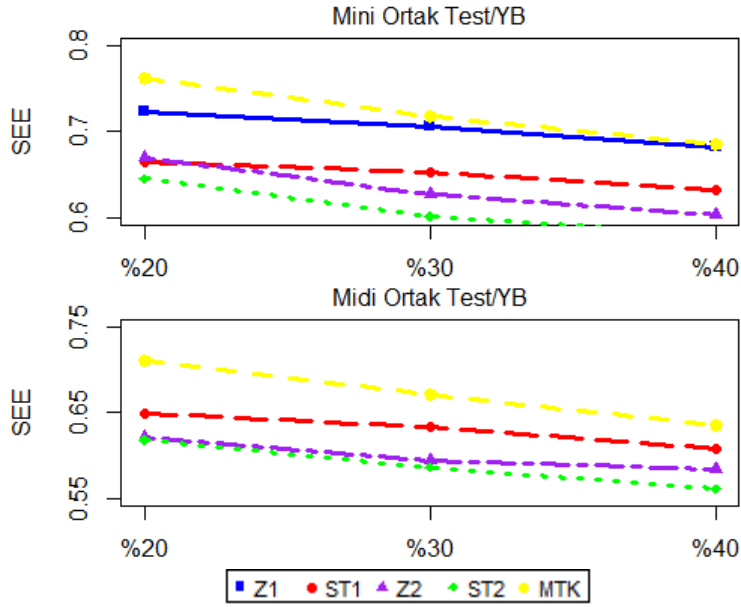
**Şekil 4.6. 500 Örneklem için İç Ortak Testte Gruplar Arası Yetenek Farklı Olduğunda Elde Edilen SEE Değerleri**

Şekil 4.5 ve 4.6'da yer alan grafikler incelendiğinde, her iki yetenek dağılımı koşulunda hem KE eşitleme yöntemlerinde hem de MTK gerçek puan eşitleme yönteminde ortak madde oranı arttıkça standart hataların azaldığı görülmüştür. Bütün eşitleme yöntemlerinde gruplar arası yetenek dağılımı farklı olduğunda

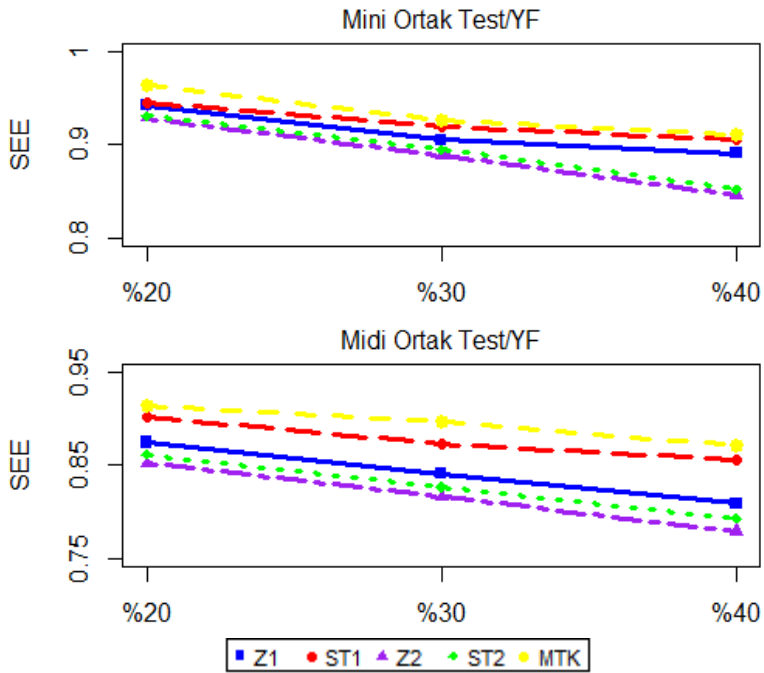
standart hatalarının artış göstermiştir. KE doğrusal eşitleme yöntemleri, gruplar arası yetenek dağılımının benzer olduğu koşulda birbirine yakın standart hata değerlerine sahipken, yetenek dağılımı farklı olduğunda birbirinden farklılaşmaktadır. Bütün koşullarda KE doğrusal eşitleme yöntemleri, KE eşit yüzdeliği eşitleme yöntemlerine göre daha düşük standart hata vermiştir. Gruplar arası yetenek dağılımının benzer olduğu durumda KE son tabakalama doğrusal eşitleme yöntemi, yetenek dağılımı farklı olduğunda ise KE zincirleme doğrusal eşitleme yöntemi daha düşük standart hata vermiştir. Gruplar arası yetenek dağılımının farklı olduğu mini ortak testte, eşitleme yöntemlerinden elde edilen standart hataların diğer koşullara göre birbirine daha yakın olduğu görülmektedir. Standart hata değerlerinin mini ve midi ortak test koşulunda birbirine yakın olduğu, hatta mini ortak testin midi ortak teste göre biraz daha düşük olduğu görülmektedir. Bütün koşullarda MTK gerçek puan eşitleme yönteminin en yüksek standart hataya sahip olduğu, ancak yetenek dağılımının benzer ve ortak madde güçlük dağılımının mini olduğu koşulda farkın daha fazla olduğu görülmektedir. Gruplar arası yetenek dağılımının farklı ve midi ortak test olduğu durumda son tabakalama eşit yüzdeliği eşitleme ile MTK gerçek puan eşitleme yönteminden elde edilen standart hata değerleri birbirine oldukça yakındır.

#### ***Dış ortak teste göre standart hata nasıl değişmektedir?***

Dış ortak testte, ortak madde oranının %20, %30 ve %40 olduğu koşullarda gruplar arası yetenek dağılımı benzer olduğunda elde edilen eşitlemenin standart hatasına ait grafikler Şekil 4.7'de, yetenek dağılımı farklı olduğunda elde edilen eşitlemenin standart hatasına ait grafikler Şekil 4.8'de verilmiştir.



**Şekil 4.7. 500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen SEE Değerleri**



**Şekil 4.8. 500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen SEE Değerleri**

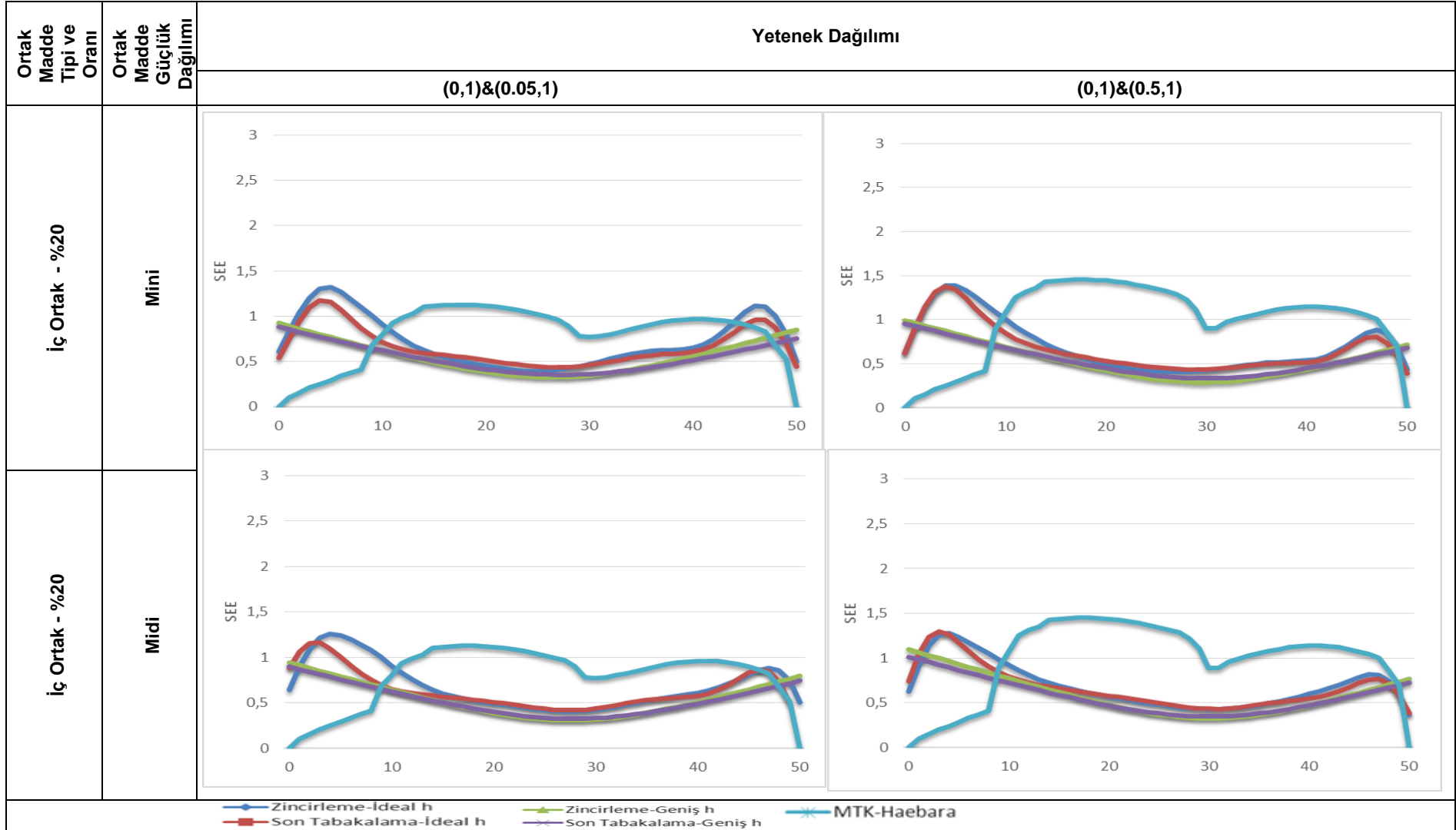
Şekil 4.7 ve Şekil 4.8 incelendiğinde, iç ortak testte olduğu gibi dış ortak testte de bütün koşullarda MTK gerçek puan eşitleme yönteminin en yüksek standart hataya sahip olduğu görülmektedir. Eşitleme yöntemlerinden elde edilen standart

hataların mini ortak test koşulunda gruplar arası yetenek dağılımının farklı olduğu durumda birbirine daha yakın; yetenek dağılımının benzer olduğu durumda ise farkın fazla olduğu görülmektedir. Bütün eşitleme yöntemlerinde, ortak madde güçlük dağılımının midi olduğu durumda elde edilen standart hata değerlerinin mini olduğu koşula göre daha düşük olduğu görülmektedir. KE son tabakalama eşitleme yöntemleri gruplar arası yetenek dağılımının farklı olduğu koşullarda, KE zincirleme eşit yüzdeli eşitleme yöntemlerine göre daha yüksek; gruplar arası yetenek dağılımının benzer olduğu koşulda ise daha düşük standart hata vermiştir. Midi ortak test koşulunda her iki yetenek dağılımında da mini ortak test koşuluna göre daha düşük standart hata elde edilmiştir. Şekil 4.5-4.6 ile Şekil 4.7-4.8'deki grafikler karşılaştırıldığında, dış ortak testten elde edilen standart hata değerlerinin iç ortak testten elde edilen değerlere göre daha düşük olduğu görülmektedir.

***Her bir eşitlenmiş puana ilişkin standart hatalar çalışmada ele alınan koşullara göre nasıl değişmektedir?***

500 örneklem için, iç ortak testte, %20 ortak madde oranında 4 (gruplar arası yetenek dağılımı ve ortak madde güçlük dağılımı) koşulda her bir eşitlenmiş puan için elde edilen standart hata grafikleri Şekil 4.9'da verilmiştir. Böylece eşitleme yöntemlerine göre hangi puan aralıklarında standart hatanın daha düşük olduğu belirlenmiştir. Diğer koşullara ilişkin tüm puan ölçeği boyunca elde edilen standart hatalara ait grafikler EK -11 de Şekil 1 – Şekil 5 arasında verilmiştir.





Şekil 4.9. 500 Örneklem için ortak madde oranının %20 olduğu koşulda elde edilen SEE değerleri

Şekil 4.9'da yer alan grafikler incelendiğinde, KE son tabakalama ve eşit yüzdellikli eşitleme yöntemlerinden elde edilen SEE değerleri uç puanlarda yüksek; orta puanlarda ise daha düşük değerlere sahip ve düşük uç puanlarda yüksek uç puanlara göre SEE değerleri daha yüksek çıkmıştır. KE zincirleme (geniş h) ve son tabakalama (geniş h) yöntemlerinde de uç puanlardaki SEE değerlerinin orta puanlara göre daha yüksek olduğu, ancak doğrusal yöntemlerde, uç ve orta puan ölçeğinde elde edilen SEE değerlerinin birbirine daha yakın olduğu görülmektedir. Ayrıca tüm ölçek puanlarında, doğrusal eşitleme yöntemlerinin (geniş h), eşit yüzdellikli eşitleme (ideal h) yöntemlerinden daha düşük SEE değeri verdiği görülmektedir. Gruplar arası yetenek dağılımı benzer olduğunda düşük puanlarda zincirleme eşit yüzdellikli eşitleme yönteminde, son tabakalama eşit yüzdellikli eşitleme yönteminden daha yüksek SEE değeri elde edilmişken, yetenek dağılımı farklı olduğunda SEE değerleri birbirine yaklaşmıştır. MTK gerçek puan eşitleme yönteminde uç ve orta puanlarda düşük, diğer puanlarda ise daha yüksek SEE elde edilmiştir. Orta puanlarda elde edilen SEE değerlerinin diğer yöntemlerden elde edilen standart hatalara göre yüksek olduğu görülmektedir. MTK gerçek puan eşitleme yönteminde gruplar arası yetenek dağılımı farklı olduğunda, hem mini hem de midi ortak test koşulunda SEE değerlerinin arttığı görülmektedir. Ancak SEE değerlerindeki artış mini ortak test koşulunda daha fazladır.

#### **4.2. Alt Problem 2'ye İlişkin Bulgular ve Yorumlar**

“Testler, Kernel son tabakalama doğrusal, Kernel zincirleme doğrusal, Kernel son tabakalama eşit yüzdellikli, Kernel zincirleme eşit yüzdellikli ve MTK gerçek puan eşitleme yöntemleriyle orta büyüklükteki örneklerde (1500) eşitlendiğinde eşitlenmiş puanlara ilişkin;

##### **4.2.a. Eşitleme hatası çalışmada ele alınan koşullara göre nasıl değişmektedir?**

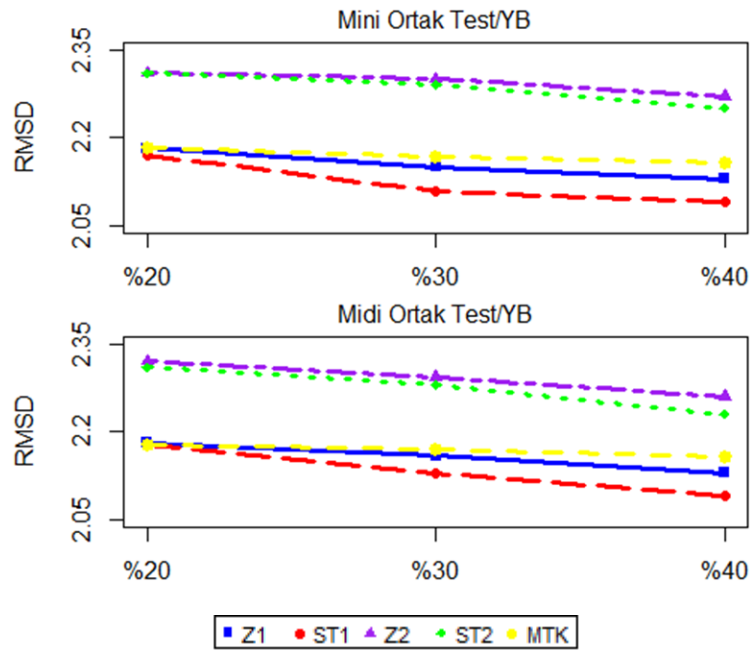
Bu alt problemin çözümünde eşitleme yöntemlerinin karşılaştırmalarında eşitleme hatalarından RMSD kullanılmıştır.

Alt problem 2.a'ya ilişkin bulguların anlaşılmasını kolaylaştırmak için araştırma kapsamında ele alınan koşullara ait eşitleme hatası sonuçları iç ve dış ortak test için ayrı ayrı yapılandırılmıştır. 1500 örneklem için ele alınan koşullara göre

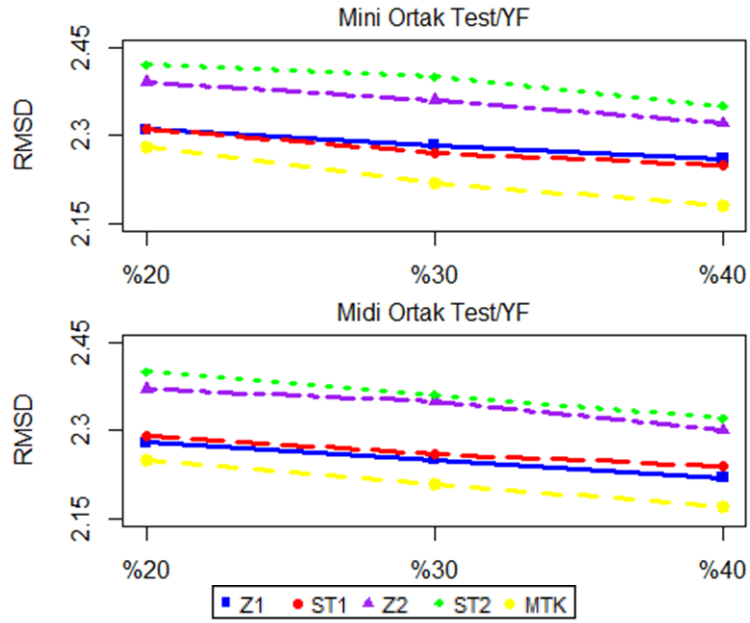
eşitleme hatalarına ait grafikler Şekil 4.10 - Şekil 4.13 arasında verilmiştir. Bu grafikler EK-6'da yer alan tabloya göre oluşturulmuştur.

### ***İç ortak teste göre eşitleme hatası nasıl değişmektedir?***

İç ortak testte, ortak madde oranının %20, %30 ve %40 olduğu koşullarda gruplar arası yetenek dağılımı benzer olduğunda elde edilen eşitleme hatasına ait grafikler Şekil 4.10'de, yetenek dağılımı farklı olduğunda elde edilen eşitleme hatasına ait grafikler Şekil 4.11'de verilmiştir.



**Şekil 4.10. 1500 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen RMSD Değerleri**



**Şekil 4.11. 1500 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen RMSD Değerleri**

Ortak madde oranının %20 olduğu durumda;

Gruplar arası yetenek dağılımı benzer olduğunda en düşük RMSD değerinin mini ortak test koşulunda son tabakalama (ideal h) eşitleme ve en yüksek RMSD değerini zincirleme (geniş h) eşitleme yöntemi vermiştir. Midi test koşulunda ise en düşük RMSD değeri son tabakalama (ideal h) ile MTK gerçek puan eşitleme yöntemlerinden ve en yüksek RMSD değeri ise zincirleme (geniş h) eşitleme yönteminden elde edilmiştir. Gruplar arası yetenek dağılımı farklı olduğunda her iki ortak test koşulunda en düşük RMSD değerinin zincirleme (ideal h) ve en yüksek RMSD değerinin son tabakalama (geniş h) eşitleme yönteminden elde edildiği görülmektedir.

Ortak madde oranının %30 olduğu durumda;

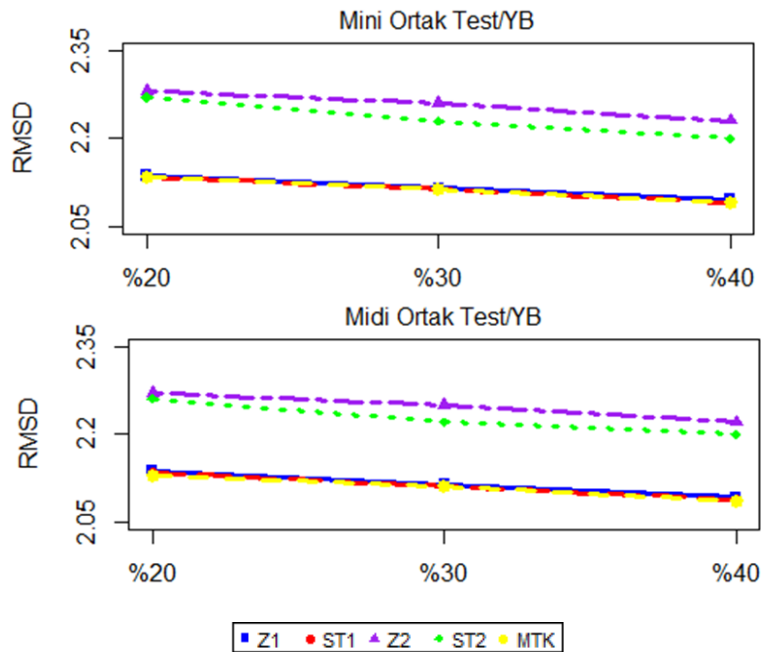
Gruplar arası yetenek dağılımı benzer olduğunda en düşük hatayı her iki ortak test koşulunda son tabakalama (ideal h) eşitleme ve en yüksek hatayı zincirleme (geniş h) eşitleme yöntemi vermiştir. Gruplar arası yetenek dağılımı farklı olduğunda en düşük RMSD değeri mini ortak test koşulunda zincirleme (ideal h) eşitleme, midi ortak test koşulunda MTK gerçek puan eşitleme, en yüksek RMSD değeri ise her iki ortak testte son tabakalama (geniş h) eşitleme yönteminden elde edilmiştir.

Ortak madde oranının %40 olduđu durumda;

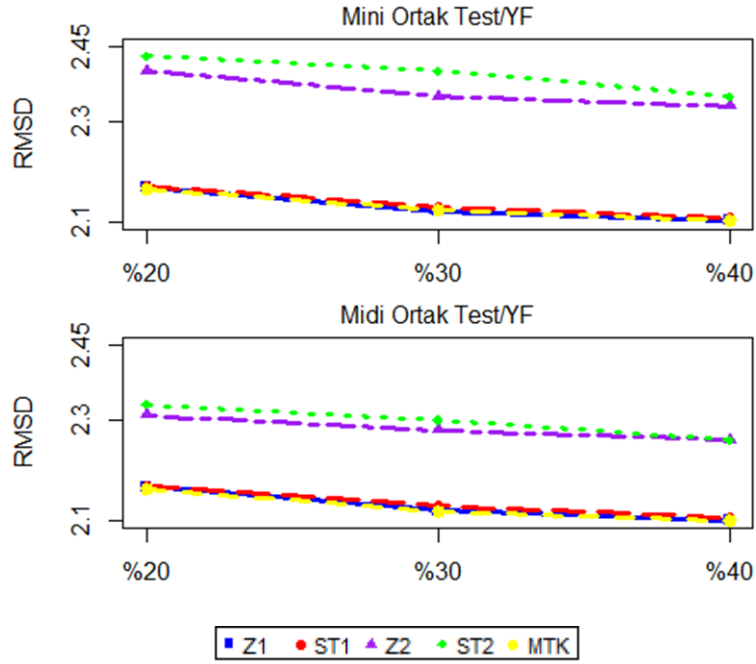
Gruplar arası yetenek dağılımı benzer olduđuunda en düşük RMSD değeri her iki ortak test koşulunda MTK gerçek puan eşitleme ve en yüksek RMSD değeri zincirleme (geniş h) eşitleme yönteminden elde edilmiştir. Gruplar arası yetenek dağılımı farklı olduđuunda en düşük RMSD değerinin her iki ortak test koşulunda MTK gerçek puan eşitleme ve en yüksek RMSD değerinin ise son tabakalama (geniş h) eşitleme yönteminden elde edildiđi görülmektedir.

### ***Dış ortak teste göre eşitleme hatası nasıl değişmektedir?***

Dış ortak testte, ortak madde oranının %20, %30 ve %40 olduđu koşullarda gruplar arası yetenek dağılımı benzer olduđuunda elde edilen eşitleme hatasına ait grafikler Şekil 4.12'de, yetenek dağılımı farklı olduđuunda elde edilen eşitleme hatasına ait grafikler Şekil 4.13'de verilmiştir.



**Şekil 4.12. 1500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduđuunda Elde Edilen RMSD Değerleri**



**Şekil 4.13. 1500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen RMSD Değerleri**

Ortak madde oranının %20 olduğu durumda;

Gruplar arası yetenek dağılımı benzer olduğunda en düşük hatayı mini ortak test koşulunda son tabakalama (ideal h) eşitleme ve en yüksek hatayı zincirleme (geniş h) eşitleme yöntemi vermiştir. Midi ortak test koşulunda ise en düşük hata MTK gerçek puan eşitleme ve en yüksek hata zincirleme (geniş h) eşitleme yönteminden elde edilmiştir. Gruplar arası yetenek dağılımı farklı olduğunda en düşük hatanın her iki ortak test koşulunda zincirleme ile MTK gerçek puan eşitleme ve en yüksek hatanın ise son tabakalama (geniş h) eşitleme yönteminden elde edildiği görülmektedir.

Ortak madde oranının %30 olduğu durumda;

Gruplar arası yetenek dağılımı benzer olduğunda en düşük hata her iki ortak test koşulunda son tabakalama (ideal h) eşitleme yönteminden ve en yüksek hata zincirleme (geniş h) eşitleme yönteminden elde edilmiştir. Gruplar arası yetenek dağılımı farklı olduğunda en düşük hata mini ortak test koşulunda zincirleme (ideal h) eşitleme ve midi test koşulunda MTK gerçek puan eşitleme; en yüksek hata ise her iki ortak test için son tabakalama (geniş h) eşitleme yönteminden elde edilmiştir.

Ortak madde oranının %40 olduđu durumda;

Gruplar arası yetenek dağılımı benzer olduđunda en düşük toplam hata mini ortak test koşulunda son tabakalama (ideal h) ve MTK gerçek puan eşitleme yöntemlerinden, midi ortak test koşulunda ise MTK gerçek puan eşitleme yönteminden elde edilmiştir. En yüksek hatayı ise her iki ortak test için zincirleme (geniş h) eşitleme yöntemi vermiştir. Gruplar arası yetenek dağılımı farklı olduđunda en düşük hatanın mini ortak test koşulunda MTK gerçek puan eşitlemede ve midi test koşulunda ise MTK gerçek puan eşitleme ile zincirleme (ideal h) eşitleme yöntemlerinde; en yüksek hata ise her iki ortak testte son tabakalama (geniş h) eşitleme yönteminden elde edildiđi görölmektedir.

Bütün eşitleme yöntemlerinde iç ve dış ortak testte ortak madde oranı arttıkça toplam hatanın azaldıđı görölmektedir. Toplam hata, iç ortak testte ortak madde oranı arttıkça gruplar arası yetenek dağılımı benzer olduđunda doğrusal eşitleme yöntemlerinde; gruplar arası yetenek dağılımı farklı olduđunda MTK gerçek puan eşitleme yönteminde daha çok azalmıştır. Dış ortak testte ise doğrusal eşitleme yöntemlerinde hata miktarlarının daha çok azaldıđı görölmektedir. Dış ortak test midi test olduđunda elde edilen hata, mini test koşuluna göre daha düşüktür. Ancak iç ortak test midi test olduđunda hata değerlerine ilişkin düzenli bir örüntü elde edilememiştir. Hem iç hem de dış ortak test için gruplar arası yetenek dağılımının benzer olduđu durumdaki toplam hata, yetenek dağılımlarının farklı olduđu duruma göre daha düşük elde edilmiştir. Doğrusal eşitleme yöntemlerinin (geniş h), eşit yüzdelikli eşitleme yöntemlerine göre daha yüksek hata verdiđi görölmektedir. MTK gerçek puan eşitlemenin, dış ortak testte bütün koşullarda ya en düşük hataya ya da en düşük hataya çok yakın hata verdiđi, iç ortak testte ise %30 ve %40 ortak madde oranı için gruplar arası yetenek dağılımının farklı olduđu koşulda daha düşük hata verdiđi görölmektedir. Gruplar arası yetenek dağılımı benzer olduđunda son tabakalama eşit yüzdelikli eşitleme; yetenek dağılımı farklı olduđunda ise zincirleme eşit yüzdelikli eşitleme yöntemi daha düşük hata vermiştir.

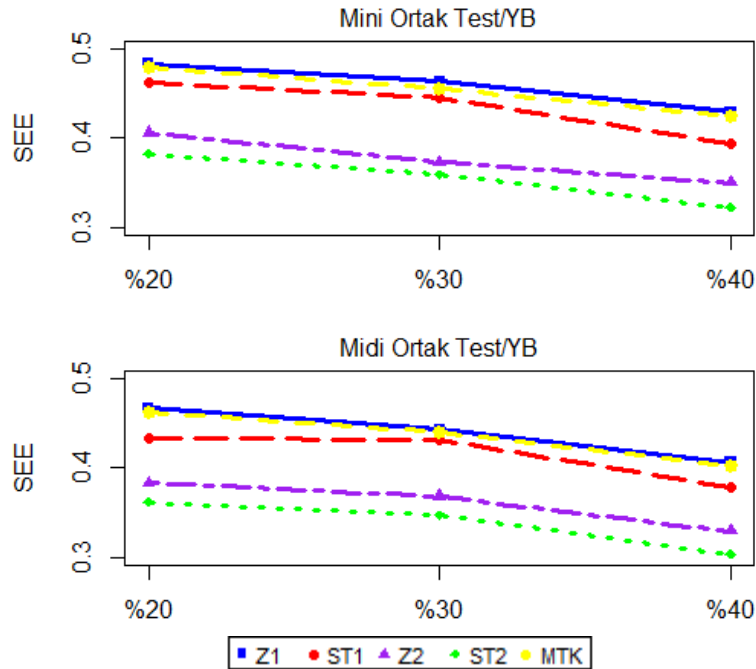
#### 4.2.b. Eşitlemenin standart hatası ve her bir puan için elde edilen standart hata çalışmada ele alınan koşullara göre nasıl değişmektedir?

Bu alt problemin çözümünde eşitleme yöntemlerinin karşılaştırmalarında eşitlemenin standart hatası (SEE) kullanılmıştır.

Alt problem 2.b'ye ilişkin bulguların anlaşılmasını kolaylaştırmak için araştırma kapsamında ele alınan koşullara ait eşitlemenin standart hatası sonuçları iç ve dış ortak test için ayrı ayrı yapılandırılmıştır. 1500 örneklem için ele alınan koşullara göre eşitlemenin standart hatasına ait grafikler Şekil 4.14 - Şekil 4.17 arasında verilmiştir. Bu grafikler EK-9'da yer alan tabloya göre oluşturulmuştur.

#### *İç ortak teste göre standart hata nasıl değişmektedir?*

İç ortak testte, ortak madde oranının %20, %30 ve %40 olduğu koşullarda gruplar arası yetenek dağılımı benzer olduğunda elde edilen eşitlemenin standart hatasına ait grafikler Şekil 4.14'de, yetenek dağılımı farklı olduğunda elde edilen eşitlemenin standart hatasına ait grafikler Şekil 4.15'de verilmiştir.

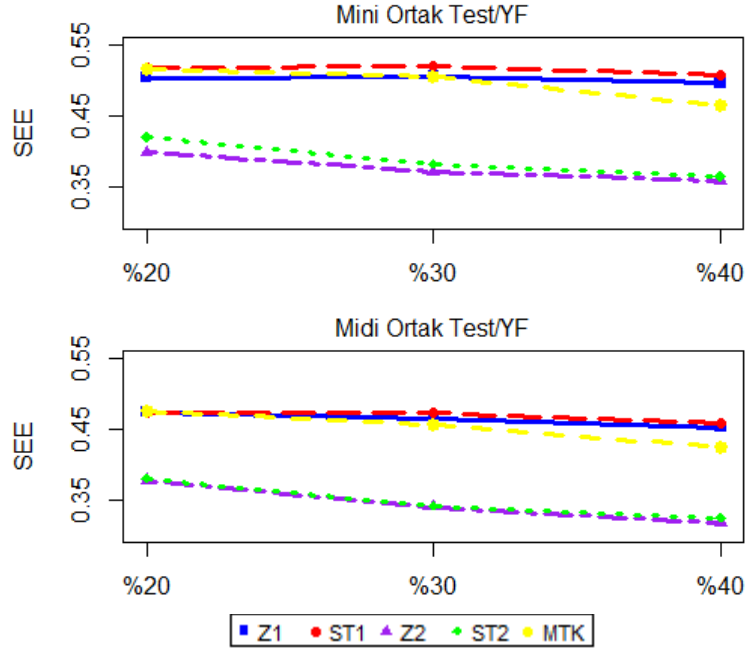


**Şekil 4.14. 1500 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen SEE Değerleri**

Şekil 4.14 ve 4.15'de yer alan grafikler incelendiğinde, bütün eşitleme yöntemleri için ortak madde oranı arttıkça standart hata değerinin azaldığı görülmektedir.



Özellikle MTK gerçek puan eşitlemede ortak madde oranının %30'dan %40'a arttığı durumda standart hatadaki düşüş daha fazla olmuştur. Bütün eşitleme yöntemlerinde gruplar arası yetenek dağılımının benzer olduğu durumda elde edilen standart hata değerleri, farklı olduğu koşula göre daha düşük çıkmıştır, yani gruplar yetenek dağılımı açısından farklı olduğunda standart hata artış göstermiştir. Bütün ortak madde oranı ve yetenek dağılımları için KE doğrusal eşitleme yöntemleri, KE eşit yüzdelikli eşitleme yöntemlerine göre daha düşük standart hata vermiştir. Doğrusal eşitleme yöntemleri ile eşit yüzdelikli eşitleme yöntemlerinden elde edilen standart hata değerleri arasındaki farkın, gruplar arası yetenek dağılımı farklı ve mini ortak test koşulunda diğer koşullara göre daha benzer olduğu görülmektedir.



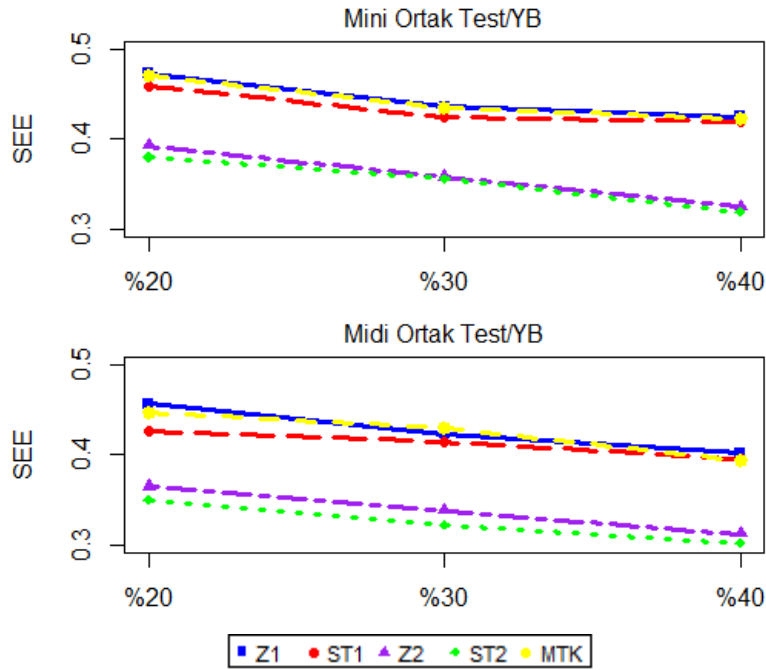
**Şekil 4.15. 1500 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen SEE Değerleri**

Gruplar arası yetenek dağılımının farklı ve testin midi olduğu koşulda mini test koşuluna göre eşitleme yöntemlerinden elde edilen standart hata değerleri birbirine çok yakın çıkmıştır. Gruplar arası yetenek dağılımının benzer ve mini ortak test koşulunda standart hata değerlerinin midi ortak test koşuluna göre biraz daha düşük, gruplar arası yetenek dağılımı farklı olduğunda ise midi ortak testin mini ortak teste göre daha düşük standart hata verdiği görülmektedir. Gruplar arası yetenek dağılımının benzer olduğu durumda son tabakalama doğrusal eşitleme

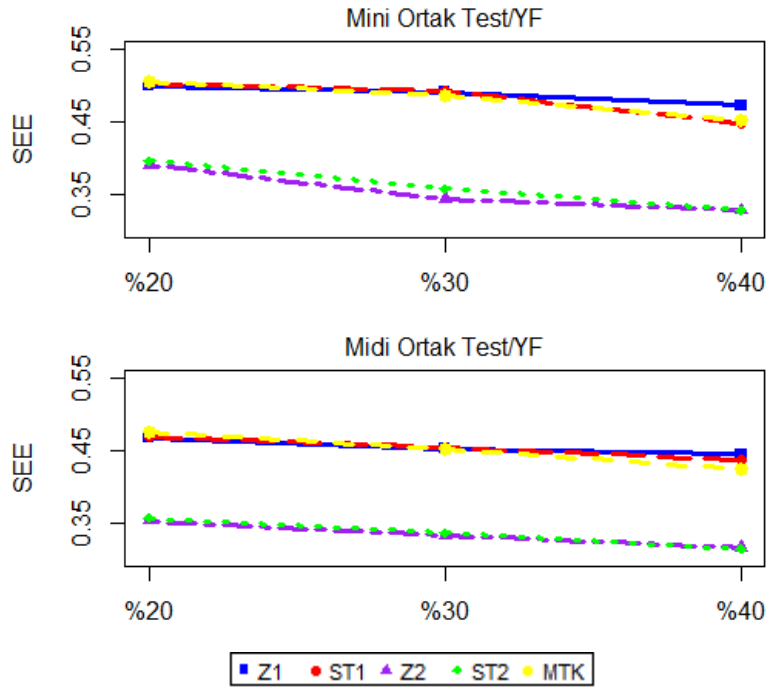
yöntemi düşük standart hata değeri verirken, yetenek dağılımı farklı olduğunda zincirleme doğrusal eşitleme yöntemi daha düşük standart hata değeri üretmiştir. Bütün durumlarda MTK gerçek puan eşitleme yönteminin, doğrusal eşitleme yöntemlerinden daha yüksek, eşit yüzdeliği eşitleme yöntemlerine ise yakın veya daha düşük standart hataya sahip olduğu görülmektedir. Gruplar arası yetenek dağılımının benzer olduğu durumda zincirleme eşit yüzdeliği eşitlemeden düşük ancak son tabakalı eşit yüzdeliği eşitlemeden yüksek standart hata elde edilmiştir. Yetenek dağılımlarının farklı olduğu durumda midi ortak test koşulunda her iki eşit yüzdeliği eşitleme yönteminden düşük, mini ortak test koşulunda ise %20 ile %30 ortak madde oranında son tabakalama eşit yüzdeliği eşitlemeye yakın, %40 ortak madde koşulunda ise daha düşük standart hata değerleri elde edilmiştir.

### ***Dış ortak teste göre standart hata nasıl değişmektedir?***

Dış ortak testte, ortak madde oranının %20, %30 ve %40 olduğu koşullarda gruplar arası yetenek dağılımı benzer olduğunda elde edilen eşitlemenin standart hatasına ait grafikler Şekil 4.16'da, yetenek dağılımı farklı olduğunda elde edilen eşitlemenin standart hatasına ait grafikler Şekil 4.17'de verilmiştir.



**Şekil 4.16. 1500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen SEE Değerleri**



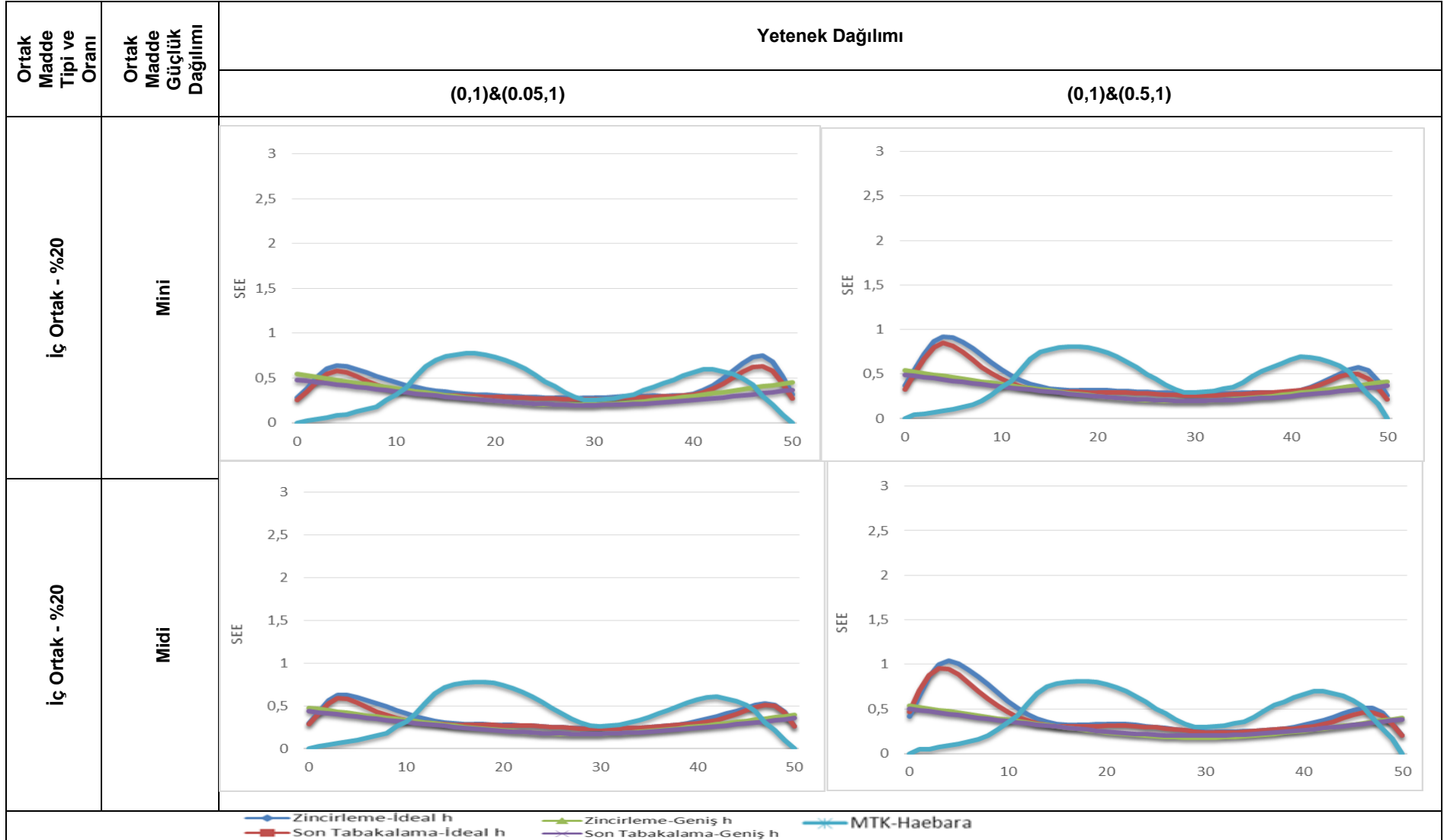
**Şekil 4.17. 1500 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen SEE değerleri**

Şekil 4.16 ve 4.17’de yer alan grafikler incelendiğinde, bütün eşitleme yöntemlerinde iç ortak testte olduğu gibi dış ortak testte de ortak madde oranı arttıkça eşitleme hatası azalmıştır. Özellikle doğrusal eşitleme yöntemlerinde bu fark daha net olarak görülmektedir. Her iki ortak test koşulunda gruplar arası yetenek dağılımı farklı olduğunda standart hata değerleri artış göstermiştir. Bütün koşullarda Kernel doğrusal eşitleme yöntemleri, Kernel eşit yüzdelikli eşitleme yöntemlerine göre daha düşük standart hata vermiştir. Gruplar arası yetenek dağılımı benzer olduğunda zincirleme eşitleme yöntemlerinin son tabakalama eşitleme yöntemlerine göre yüksek, farklı olduğunda ise daha düşük standart hata verdiği görülmektedir. Gruplar arası yetenek dağılımının farklı olduğu mini test koşulunda %30 ortak madde oranında doğrusal ile eşit yüzdelikli eşitleme yöntemlerinin standart hata değerleri arasındaki farkın diğer durumlara göre daha fazla olduğu görülmektedir. Her iki yetenek dağılımı için, midi test koşulunda mini test koşuluna göre daha düşük standart hata elde edilmiştir. Şekil 4.16 ile Şekil 4.17 karşılaştırıldığında dış ortak testten elde edilen standart hata değerlerinin iç ortak testten elde edilen değerlere göre daha düşük olduğu görülmektedir.

***Her bir eşitlenmiş puana ilişkin standart hatalar çalışmada ele alınan koşullara göre nasıl değişmektedir?***

1500 örneklem için, iç ortak testte, %20 ortak madde oranında 4 (gruplar arası yetenek dağılımı ve ortak madde güçlük dağılımı) koşulda her bir eşitlenmiş puan için elde edilen standart hata grafikleri Şekil 4.18'de verilmiştir. Böylece eşitleme yöntemlerine göre hangi puan aralıklarında standart hatanın düşük olduğu belirlenmiştir. Diğer koşullara ilişkin tüm puan ölçeği boyunca elde edilen standart hatalara ait grafikler EK-12'de Şekil 1-Şekil 5 arasında verilmiştir.

Şekil 4.18'de yer alan grafikler incelendiğinde, KE son tabakalama ve eşit yüzdelikli eşitleme yöntemlerinden elde edilen SEE değerlerinin uç puanlarda yüksek, orta puanlarda ise daha düşük olduğu görülmektedir. KE zincirleme (geniş h) ve son tabakalama (geniş h) yöntemlerinde uç ve orta puanlarda elde edilen SEE değerlerinin birbirine daha yakın olduğu görülmektedir. Ayrıca tüm ölçek puanlarında, doğrusal eşitleme yöntemlerinin (geniş h), eşit yüzdelikli eşitleme (ideal h) yöntemlerinden daha düşük standart hata verdiği, ancak orta puan ölçeğinde hataların birbirlerine yakın değerler aldığı görülmektedir. Gruplar arası yetenek dağılımı farklı olduğunda eşit yüzdelikli eşitleme yöntemlerinde düşük uç puanlarda elde edilen standart hata değerlerinin arttığı görülmektedir. MTK gerçek puan eşitleme yönteminde uç ve orta puanlarda düşük, diğer puanlarda ise daha yüksek SEE değerleri elde edilmiştir. Orta puanlardan elde edilen SEE değerlerinin diğer yöntemlerden elde edilen SEE değerleri yakındır. MTK gerçek puan eşitleme yönteminde gruplar arası yetenek dağılımı farklı olduğunda, hem mini hem de midi ortak test koşulunda SEE değerlerinin özellikle 35-45 puan aralığında arttığı görülmektedir. Ancak SEE değerlerindeki artış mini ortak test koşulunda daha fazladır.



Şekil 4.18. 1500 örneklem için ortak madde oranının %20 olduğu koşulda elde edilen SEE değerleri

### **4.3. Alt Problem 3'e İlişkin Bulgular ve Yorumlar**

“Testler, Kernel son tabakalama doğrusal, Kernel zincirleme doğrusal, Kernel son tabakalama eşit yüzdelikli, Kernel zincirleme eşit yüzdelikli ve MTK gerçek puan eşitleme yöntemleriyle büyük örneklerde (5000) eşitlendiğinde eşitlenmiş puanlara ilişkin;

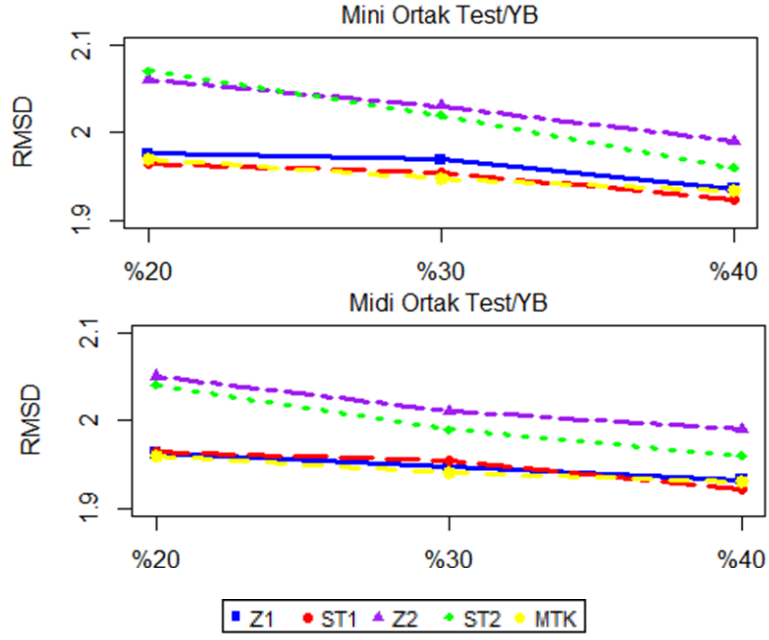
#### **4.3.a. Eşitleme hatası çalışmada ele alınan koşullara göre nasıl değişmektedir?**

Bu alt problemin çözümünde eşitleme yöntemlerinin karşılaştırmalarında eşitleme hatalarından RMSD kullanılmıştır.

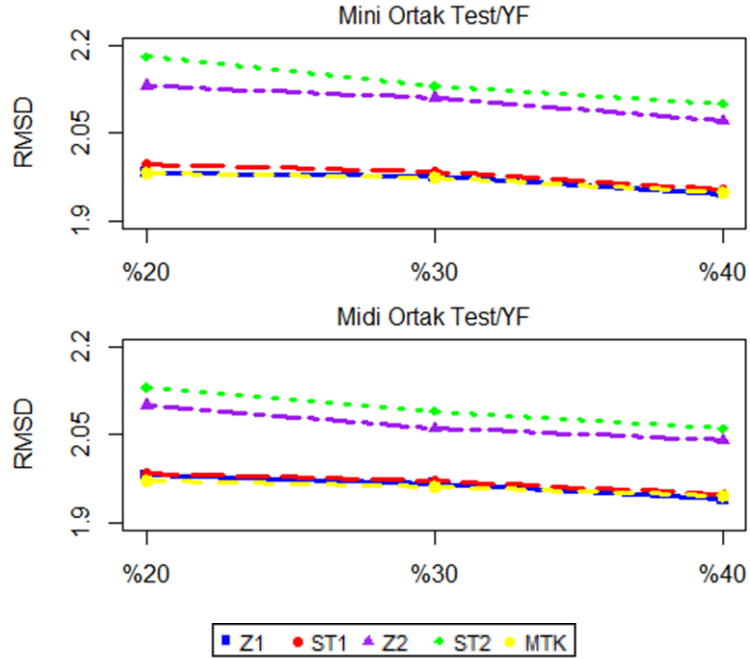
Alt problem 3.a'ya ilişkin bulguların anlaşılmasını kolaylaştırmak için araştırma kapsamında ele alınan koşullara ait eşitleme hatası sonuçları iç ve dış ortak test için ayrı ayrı yapılandırılmıştır. 5000 örneklem için ele alınan koşullara göre eşitleme hatalarına ait grafikler Şekil 4.19 - Şekil 4.22 arasında verilmiştir. Bu grafikler EK-7'de yer alan tabloya göre oluşturulmuştur.

#### ***İç ortak teste göre eşitleme hatası nasıl değişmektedir?***

İç ortak testte, ortak madde oranının %20, %30 ve %40 olduğu koşullarda gruplar arası yetenek dağılımı benzer olduğunda elde edilen eşitleme hatasına ait grafikler Şekil 4.19'da, yetenek dağılımı farklı olduğunda elde edilen eşitleme hatasına ait grafikler Şekil 4.20'de verilmiştir.



**Şekil 4.19. 5000 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen RMSD Değerleri**



**Şekil 4.20. 5000 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen RMSD Değerleri**

Ortak madde oranının %20 olduğu durumda;

Gruplar arası yetenek dağılımı benzer olduğunda, en düşük toplam hata mini ortak test koşulunda son tabakalama (ideal h) eşitleme; midi ortak test koşulunda ise MTK gerçek puan eşitleme yönteminden elde edilmiştir. En yüksek hatayı mini ortak test koşulunda zincirleme (geniş h) eşitleme ve midi ortak test koşulunda son

tabakalama (geniş h) eşitleme yöntemi vermiştir. Gruplar arası yetenek dağılımı farklı olduğunda en düşük hatanın her iki ortak test koşulunda MTK gerçek puan eşitleme ve en yüksek hatanın her iki ortak test koşulunda son tabakalama (geniş h) eşitleme yönteminden elde edildiği görülmektedir.

Ortak madde oranının %30 olduğu durumda;

Gruplar arası yetenek dağılımı benzer olduğunda en düşük toplam hatayı her iki ortak test koşulunda MTK gerçek puan eşitleme; en yüksek hatayı ise zincirleme (geniş h) eşitleme yöntemi vermiştir. Gruplar arası yetenek dağılımı farklı olduğunda en düşük hata her iki ortak test koşulunda MTK gerçek puan eşitleme ve en yüksek hata ise son tabakalama (geniş h) eşitleme yönteminden elde edilmiştir.

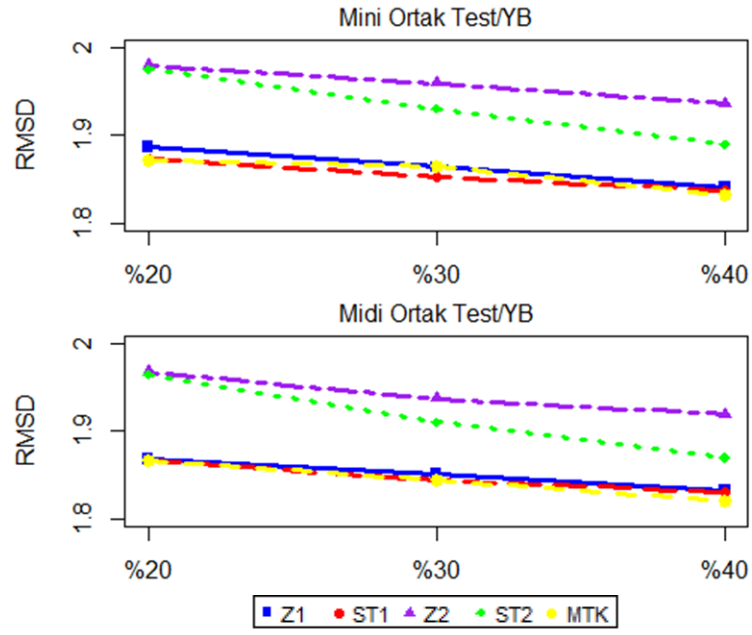
Ortak madde oranının %40 olduğu durumda;

Gruplar arası yetenek dağılımı benzer olduğunda en düşük toplam hata her iki ortak test koşulunda son tabakalama (ideal h) eşitleme, en yüksek hata zincirleme (geniş h) eşitleme yönteminden elde edilmiştir. Gruplar arası yetenek dağılımı farklı olduğunda en düşük hata her iki ortak test koşulunda zincirleme (ideal h) eşitleme ve en yüksek hata ise son tabakalama (geniş h) eşitleme yönteminden elde edilmiştir.

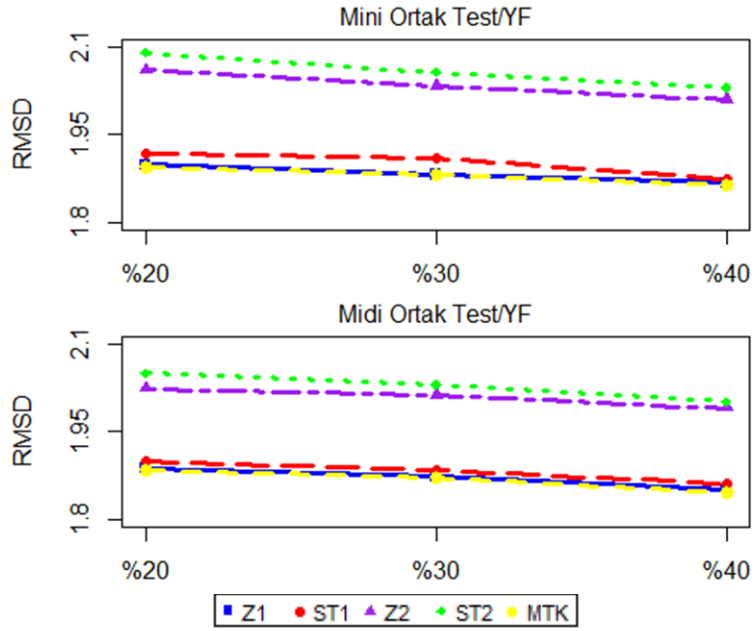
### ***Dış ortak teste göre eşitleme hatası nasıl değişmektedir?***

Dış ortak testte, ortak madde oranının %20, %30 ve %40 olduğu koşullarda gruplar arası yetenek dağılımı benzer olduğunda elde edilen eşitleme hatasına ait grafikler Şekil 4.21'de, yetenek dağılımı farklı olduğunda elde edilen eşitleme hatasına ait grafikler Şekil 4.22'de verilmiştir.





**Şekil 4.21. 5000 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen RMSD Değerleri**



**Şekil 4.22. 5000 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen RMSD Değerleri**

Ortak madde oranının %20 olduğu durumda;

Gruplar arası yetenek dağılımı benzer olduğunda en düşük toplam hata mini ortak test koşulunda MTK gerçek puan eşitleme; midi ortak test koşulunda yine MTK gerçek puan eşitleme ile son tabakalama (ideal h) eşitleme yöntemlerinden elde edilmiştir. En yüksek hata ise her iki ortak test için zincirleme (geniş h) eşitleme

yönteminden elde edilmiştir. Gruplar arası yetenek dağılımı farklı olduğunda en düşük toplam hatayı her iki ortak test koşulunda MTK gerçek puan eşitleme; en yüksek hatayı ise son tabakalama (geniş h) eşitleme yöntemi vermiştir.

Ortak madde oranının %30 olduğu durumda;

Gruplar arası yetenek dağılımı benzer olduğunda en düşük toplam hata her iki ortak test koşulunda son tabakalama (ideal h) eşitleme, en yüksek hata zincirleme (geniş h) eşitleme yönteminden elde edilmiştir. Gruplar arası yetenek dağılımı farklı olduğunda en düşük hata mini ortak test koşulunda zincirleme (ideal h) ve MTK gerçek puan eşitleme ve midi ortak test koşulunda ise MTK gerçek puan eşitleme yönteminden elde edilmiştir. En yüksek hatayı ise her iki ortak testte son tabakalama (geniş h) eşitleme yöntemi vermiştir.

Ortak madde oranının %40 olduğu durumda;

Bütün koşullarda en düşük hata MTK gerçek puan eşitleme yönteminden elde edilmiştir. En yüksek hatanın gruplar arası yetenek dağılımı benzer olduğunda zincirleme (geniş h) eşitleme; gruplar arası yetenek dağılımı farklı olduğunda ise son tabakalama (geniş h) eşitleme yönteminden elde edildiği görülmektedir.

Çalışmadaki tüm koşullar dikkate alındığında, bütün eşitleme yöntemlerinde hata değerlerinin birbirine yakın olduğu görülmektedir. Doğrusal eşitleme yöntemlerinin (geniş h), eşit yüzdelli eşitleme yöntemlerine göre daha yüksek hata verdiği belirlenmiştir. Ortak madde oranı arttıkça hem iç hem de dış ortak testte toplam hatanın azalmıştır. Toplam hata iç ortak testte ortak madde oranı arttıkça her iki yetenek dağılımı koşulu için zincirleme eşit yüzdelli eşitleme yönteminde daha çok azalmıştır. Dış ortak testte ise gruplar arası yetenek dağılımı benzer olduğunda mini ortak test koşulunda son tabakalama eşit yüzdelli eşitleme, diğer koşullarda ise son tabakalama doğrusal eşitleme yönteminde hata miktarlarının daha çok azaldığı görülmektedir. Hem iç hem de dış ortak test için midi ortak test koşulunda elde edilen hatanın, mini test koşuluna göre daha düşük ya da aynı olduğu bulunmuştur. Toplam hata, gruplar arası yetenek dağılımı benzer olduğunda, farklı olduğu duruma göre daha düşüktür. MTK gerçek puan eşitlemenin, dış ortak testte bütün koşullarda, iç ortak testte ise %20 ve %30 ortak madde oranı koşullarında en düşük hatayı verdiği görülmektedir. Kernel eşitleme yöntemleri arasından gruplar arası yetenek dağılımı benzer olduğunda son

tabakalama eşit yüzdelikli eşitleme, yetenek dağılımı farklı olduğunda ise zincirleme eşit yüzdelikli eşitleme daha düşük hata vermiştir.

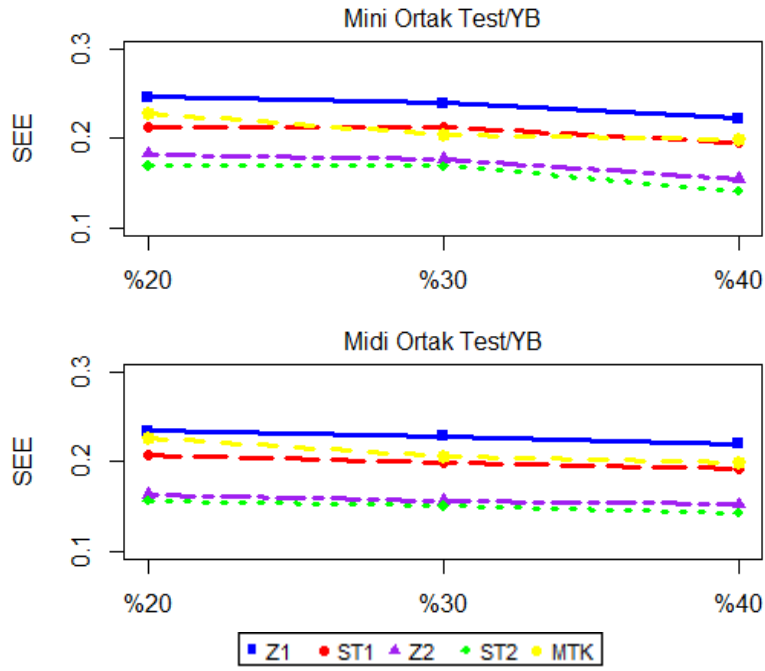
#### **4.3.b. Eşitlemenin standart hatası ve her bir puan için elde edilen standart hata çalışmada ele alınan koşullara göre nasıl değişmektedir?**

Bu alt problemin çözümünde eşitleme yöntemlerinin karşılaştırmalarında eşitlemenin standart hatası (SEE) kullanılmıştır.

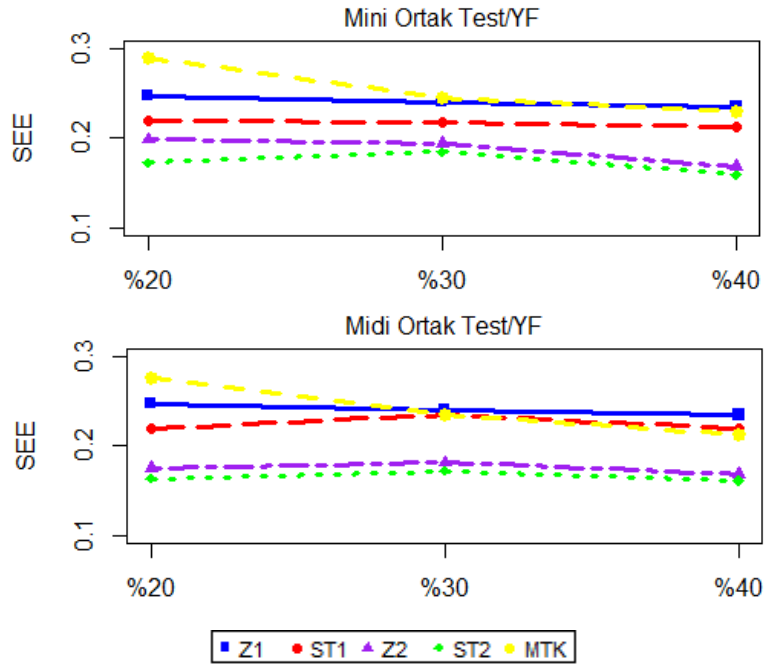
Alt problem 3.b'ye ilişkin bulguların anlaşılmasını kolaylaştırmak için araştırma kapsamında ele alınan koşullara ait eşitlemenin standart hatası sonuçları iç ve dış ortak test için ayrı ayrı yapılandırılmıştır. 5000 örneklem için ele alınan koşullara göre eşitlemenin standart hatasına ait grafikler Şekil 4.23-Şekil 4.26 arasında verilmiştir. Bu grafikler EK-10'da yer alan tabloya göre oluşturulmuştur.

#### ***İç ortak teste göre standart hata nasıl değişmektedir?***

İç ortak testte, ortak madde oranının %20, %30 ve %40 olduğu koşullarda gruplar arası yetenek dağılımı benzer olduğunda elde edilen eşitlemenin standart hatasına ait grafikler Şekil 4.23'de, yetenek dağılımı farklı olduğunda elde edilen eşitlemenin standart hatasına ait grafikler Şekil 4.24'de verilmiştir.



**Şekil 4.23. 5000 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen SEE Değerleri**



**Şekil 4.24. 5000 Örneklem İçin İç Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen SEE Değerleri**

Şekil 4.23 ve 4.24 incelendiğinde, örneklem büyüklüğünün 5000 olduğu koşullarda bütün eşitleme yöntemlerinde ortak madde oranı arttıkça standart hata değeri azalmaktadır. Gruplar arası yetenek dağılımının benzer olduğu durumda elde edilen standart hata değerleri, farklı olduğu koşula göre daha düşüktür. Bütün koşullarda KE doğrusal eşitleme yöntemleri, KE eşit yüzdelikli eşitleme

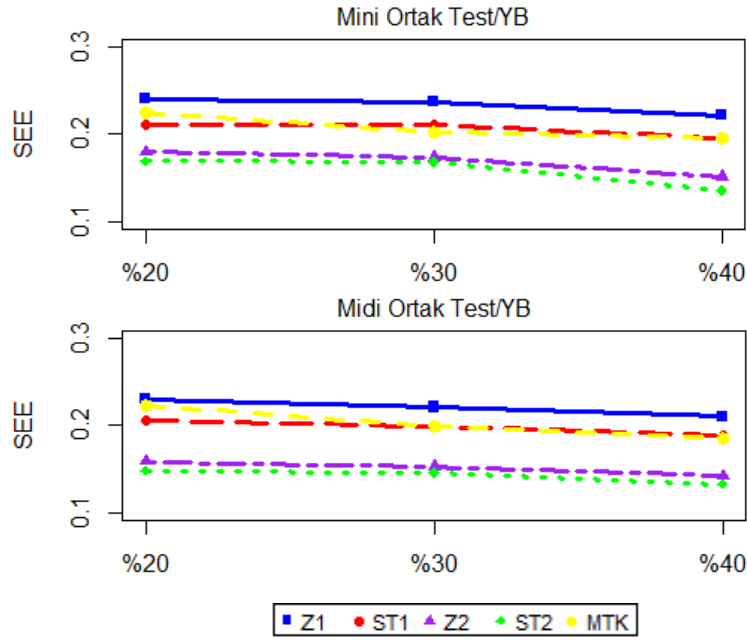
yöntemlerine göre daha düşük standart hata vermiştir. Gruplar arası yetenek dağılımı benzer olduğunda doğrusal ve eşit yüzdelliği eşitleme yöntemlerinden elde edilen standart hatalar birbirine daha yakın, yetenek dağılımı farklı olduğunda fark daha da artmıştır. Gruplar arası yetenek dağılımının benzer olduğu durumda zincirleme eşit yüzdelliği eşitlemenin son tabakalama eşit yüzdelliği eşitlemeye göre daha yüksek standart hata verdiği ve bu farkın yetenek dağılımının farklı olduğu duruma göre daha fazla olduğu görülmektedir.

Gruplar arası yetenek dağılımı benzer, mini ortak test koşulundaki standart hata değerlerinin midi ortak test koşuluna göre biraz daha düşük, gruplar arası yetenek dağılımı farklı olduğunda ise midi ortak testin mini ortak teste göre daha düşük standart hata verdiği görülmektedir. Gruplar arası yetenek dağılımının benzer olduğu durumda son tabakalama doğrusal eşitleme yöntemi düşük standart hata verirken, yetenek dağılımı farklı olduğunda zincirleme doğrusal eşitleme yöntemi ile son tabakalama doğrusal eşitleme benzer/yakın standart hata vermiştir.

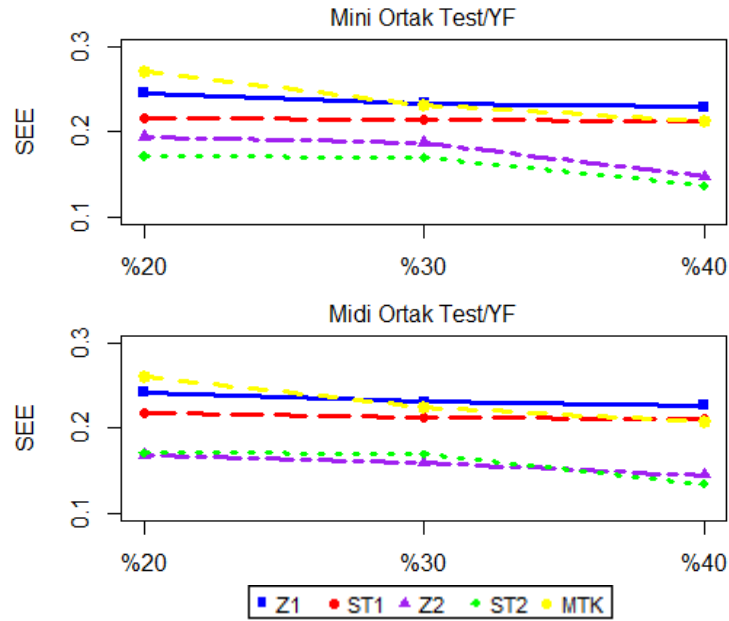
Bütün durumlarda MTK gerçek puan eşitleme yönteminin, doğrusal eşitleme yöntemlerinden daha yüksek standart hataya sahip olduğu görülmektedir. Ancak bu durum eşit yüzdelliği eşitleme yöntemleri için farklılık göstermektedir. Gruplar arası yetenek dağılımının benzer olduğu ve mini test koşulunda zincirleme eşit yüzdelliği eşitlemeden düşük son tabakalı eşit yüzdelliği eşitlemeye yakın, midi test koşulunda ise %30 ortak madde oranında son tabakalama doğrusal eşitlemeden daha düşük standart hata elde edilmiştir. MTK gerçek puan eşitleme yöntemi gruplar arası yetenek dağılımının farklı olduğu ve ortak madde oranının %20 olduğu koşulda en yüksek, %30 ve %40 midi ortak test koşulunda her iki eşit yüzdelliği eşitleme yönteminden düşük standart hataya sahiptir. Mini ortak test koşulunda ise %20 ile %30 ortak madde oranında son tabakalama eşit yüzdelliği eşitlemeye yakın, %40 ortak madde koşulunda ise daha düşük standart hata değerleri elde edilmiştir.

#### ***Dış ortak teste göre standart hata nasıl değişmektedir?***

Dış ortak testte, ortak madde oranının %20, %30 ve %40 olduğu koşullarda gruplar arası yetenek dağılımı benzer olduğunda elde edilen eşitlemenin standart hatasına ait grafikler Şekil 4.25'de, yetenek dağılımı farklı olduğunda elde edilen eşitlemenin standart hatasına ait grafikler Şekil 4.26'da verilmiştir.



**Şekil 4.25. 5000 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Benzer Olduğunda Elde Edilen SEE Değerleri**



**Şekil 4.26. 5000 Örneklem İçin Dış Ortak Testte Gruplar Arası Yetenek Dağılımı Farklı Olduğunda Elde Edilen SEE Değerleri**

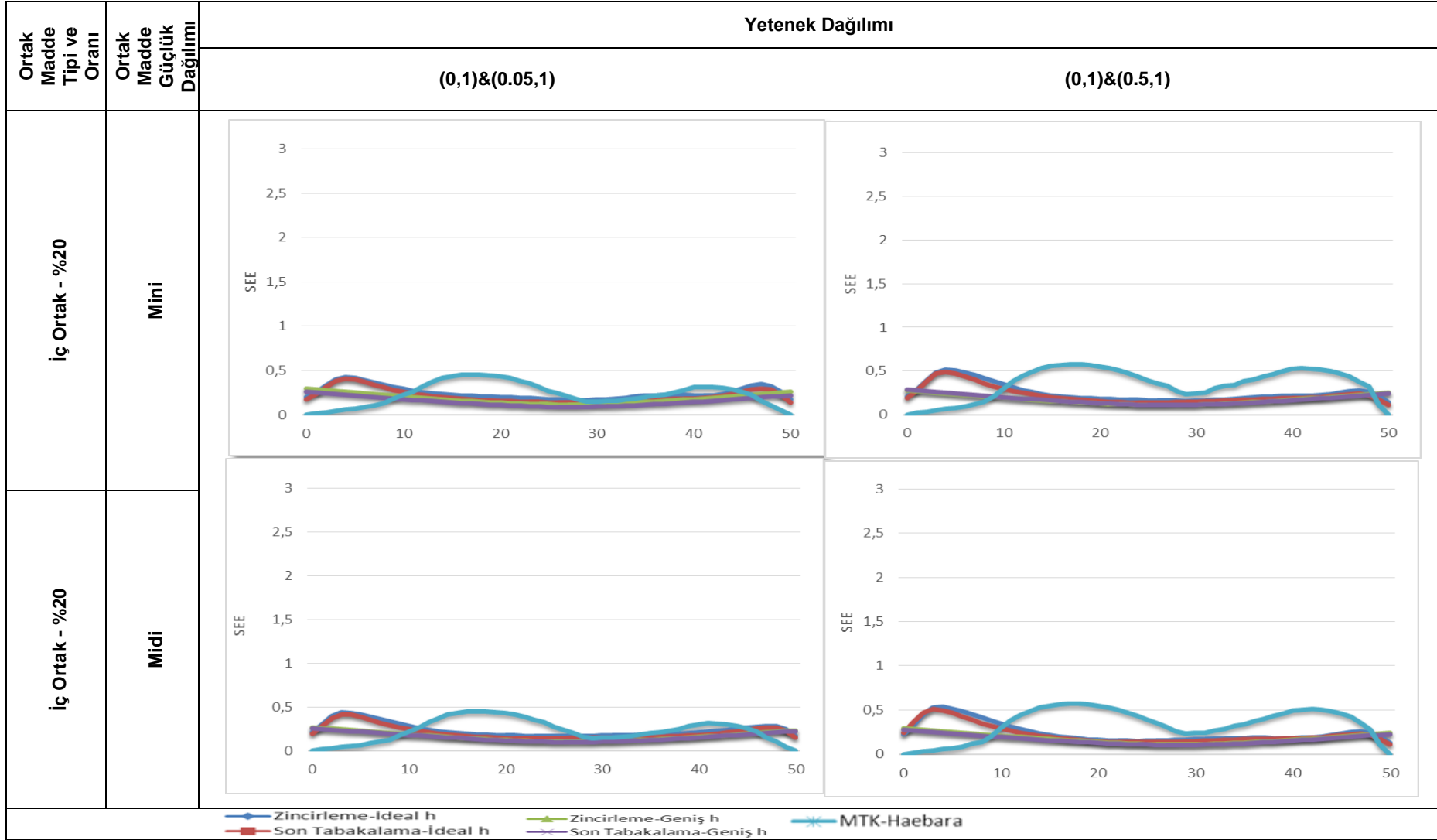
Şekil 4.25 ve 4.26'da yer alan grafikler incelendiğinde, bütün eşitleme yöntemlerinde iç ortak testte olduğu gibi dış ortak testte de ortak madde oranı arttıkça eşitleme hatası azalmıştır. Özellikle doğrusal eşitleme yöntemlerinde bu fark daha net olarak görülmektedir. Her iki ortak test koşulunda gruplar arası yetenek dağılımı farklı olduğunda standart hata değerleri artış göstermiştir. Bütün

koşullarda Kernel doğrusal eşitleme yöntemleri, Kernel eşit yüzdelikli eşitleme yöntemlerine göre daha düşük standart hata vermiştir. Gruplar arası yetenek dağılımı benzer olduğunda zincirleme eşitleme yöntemlerinin son tabakalama eşitleme yöntemlerine göre yüksek, farklı olduğunda ise daha düşük standart hata verdiği görülmektedir. Gruplar arası yetenek dağılımının farklı olduğu mini test koşulunda %30 ortak madde oranında doğrusal ile eşit yüzdelikli eşitleme yöntemlerinin standart hata değerleri arasındaki farkın diğer durumlara göre daha fazla olduğu görülmektedir. Her iki yetenek dağılımı için, midi test koşulunda mini test koşuluna göre daha düşük standart hata elde edilmiştir. Şekil 4.23-24 ile Şekil 4.25-26 karşılaştırıldığında dış ortak testten elde edilen standart hata değerlerinin iç ortak testten elde edilen değerlere göre daha düşük olduğu görülmektedir.

***Her bir eşitlenmiş puana ilişkin standart hatalar çalışmada ele alınan koşullara göre nasıl değişmektedir?***

5000 örneklem için, iç ortak testte, %20 ortak madde oranında 4 (gruplar arası yetenek dağılımı ve ortak madde güçlük dağılımı) koşulda her bir eşitlenmiş puandan elde edilen standart hata grafikleri Şekil 4.27'de verilmiştir. Böylece eşitleme yöntemlerine göre hangi puan aralıklarında standart hatanın düşük olduğu belirlenmiştir. Diğer koşullara ilişkin tüm puan ölçeği boyunca elde edilen standart hatalara ait grafikler EK-13'de Şekil 1 - Şekil 5 arasında verilmiştir.

Şekil 4.27'de yer alan grafikler incelendiğinde, Kernel son tabakalama ve eşit yüzdelikli eşitleme yöntemlerinden elde edilen SEE değerlerinin düşük uç puanlarda diğer puanlara göre yüksek olduğu görülmektedir. KE zincirleme (geniş h) ve son tabakalama (geniş h) yöntemlerinde tüm puan ölçeğinde elde edilen SEE değerleri birbirine yakın çıkmıştır. Doğrusal eşitleme yöntemlerinin (geniş h) ile eşit yüzdelikli eşitleme (ideal h) yöntemlerinin düşük uç puanlar haricinde birbirlerine yakın SEE değerlerine sahip olduğu görülmektedir. Gruplar arası yetenek dağılımı farklı olduğunda düşük uç puanlarda SEE değerleri artmıştır. MTK gerçek puan eşitleme yönteminde uç ve orta puanlarda düşük, diğer puanlarda ise daha yüksek SEE değerleri elde edilmiştir. Orta puanlardan elde edilen SEE değerleri diğer yöntemlerden elde edilen SEE değerlerine yakın çıkmıştır. MTK gerçek puan eşitleme yönteminde gruplar arası yetenek dağılımı farklı olduğunda 30-50 puan aralığındaki SEE değerlerinin, yetenek dağılımının benzer olduğu duruma göre arttığı görülmektedir.



Şekil 4.27. 5000 Örneklem İçin Ortak Madde Oranının %20 olduğu Koşulda Elde Edilen SEE Değerleri



#### 4.4. Alt Problem 4'e İlişkin Bulgular ve Yorumlar

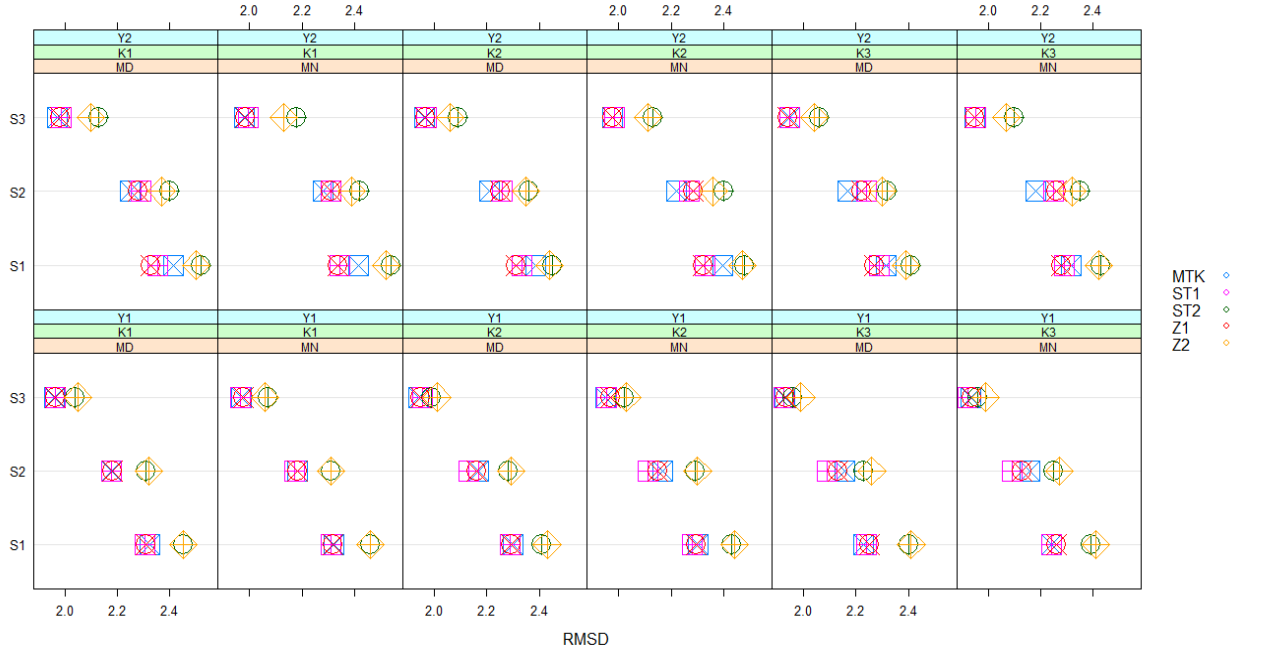
*“Çalışmada yer alan koşulların Kernel son tabakalama doğrusal, Kernel zincirleme doğrusal, Kernel son tabakalama eşit yüzdelikli, Kernel zincirleme eşit yüzdelikli ve MTK gerçek puan eşitleme yöntemlerinden elde edilen eşitleme hatasına (RMSD) ortak etkisi nasıldır?”*

Bu alt problemde, çalışmada yer alan örneklem büyüklüğü, gruplararası yetenek dağılımı, ortak madde oranı ve ortak madde güçlük dağılımı etkileşimlerinin iç ve dış ortak testte eşitleme yöntemlerinden elde edilen eşitleme hatası üzerindeki ortak etkisi incelenmiştir. İç ortak test için çalışmada ele alınan koşulların eşitleme hatası üzerindeki ortak etkisi Şekil 4.28'de, dış ortak için ise Şekil 4.29'da yer almaktadır. Şekil 4.28 ve 4.29'da; S1,S2 ve S3 sırayla 500, 1500, 5000 örneklem büyüklüğünü, Y1 yetenek dağılımının benzer ve ve Y2 farklı olduğu koşulları; K1, K2, K3 ise sırayla %20, %30 ve %40 ortak madde oranını, MN mini ortak test koşulunu ve MD midi ortak test koşulunu belirtmektedir.

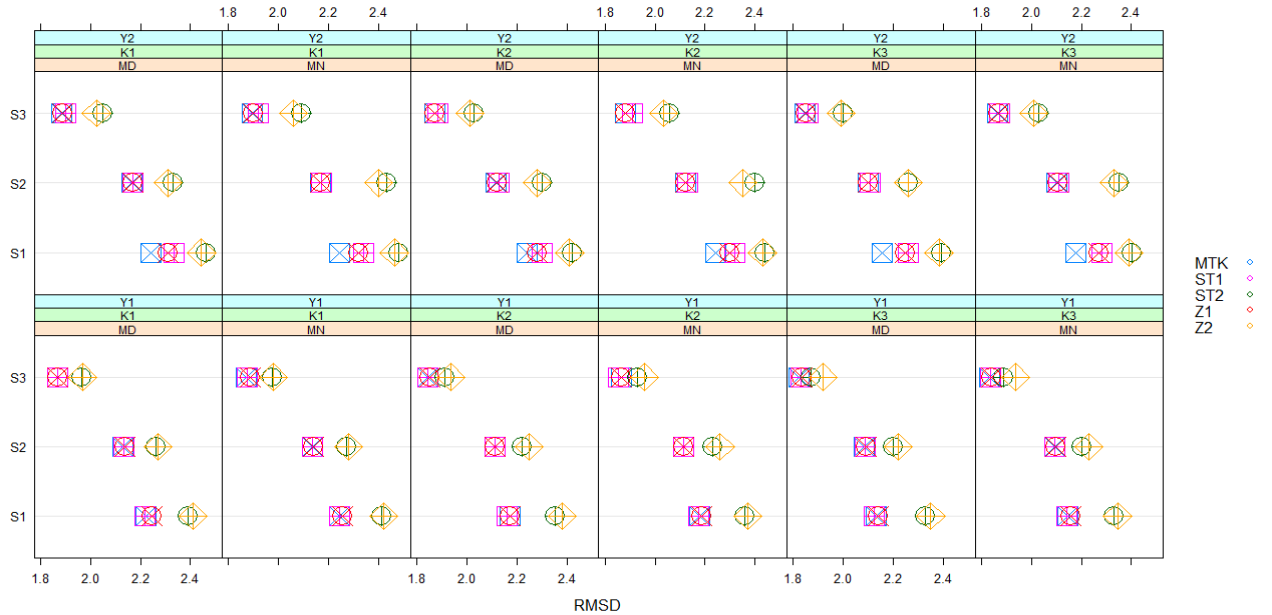
Şekil 4.28 ve 4.29'da yer alan grafikler incelendiğinde, bütün eşitleme yöntemlerinde örneklem büyüklüğü arttıkça yetenek dağılımı, ortak madde tipi, ortak madde oranı ve ortak madde güçlük dağılımının tüm düzeylerinde eşitleme hatasının azaldığı görülmektedir. Gruplar arası yetenek dağılımının farklılaştığı koşullarda, RMSD değerleri artış göstermiştir. Doğrusal eşitleme yöntemleri bütün koşullarda en yüksek eşitleme hatasına sahiptir; gruplar arası yetenek dağılımının benzer olduğu koşulda, ortak madde oranı arttıkça diğer eşitleme yöntemlerine daha yakın eşitleme hatası vermiştir. Örneklem büyüklüğünün 5000 olduğu koşulda gruplar arası yetenek dağılımı farklı olduğunda, MTK ve eşit yüzdelikli eşitleme yöntemleri benzer RMSD vermiştir. Gruplar arası yetenek dağılımı farklı olduğu koşulda, iç ortak testte dış ortak teste göre doğrusal eşitleme yöntemlerinin MTK ve eşit yüzdelikli eşitleme yöntemlerinden daha yüksek eşitleme hatasına sahip olduğu görülmektedir. Gruplar arası yetenek dağılımı benzer olduğunda ortak madde oranı arttıkça yöntemlerden elde edilen hataların birbirine yaklaşmıştır.

Ortak madde oranı arttıkça, bütün eşitleme yöntemlerinin eşitleme hatasının azaldığı görülmektedir. Ortak madde güçlük dağılımının mini olduğu koşulda eşitleme

yöntemleri arasındaki farklılaşmanın midi ortak test koşuluna göre daha fazla olduğu ve ortak madde arttıkça bu farklılaşmanın azaldığı görülmektedir.



**Şekil 4.28. İç ortak testte çalışmada ele alınan koşulların eşitleme hatasına (RMSD) ortak etkisi**



**Şekil 4.29. Dış ortak testte çalışmada ele alınan koşulların eşitleme hatasına (RMSD) ortak etkisi**

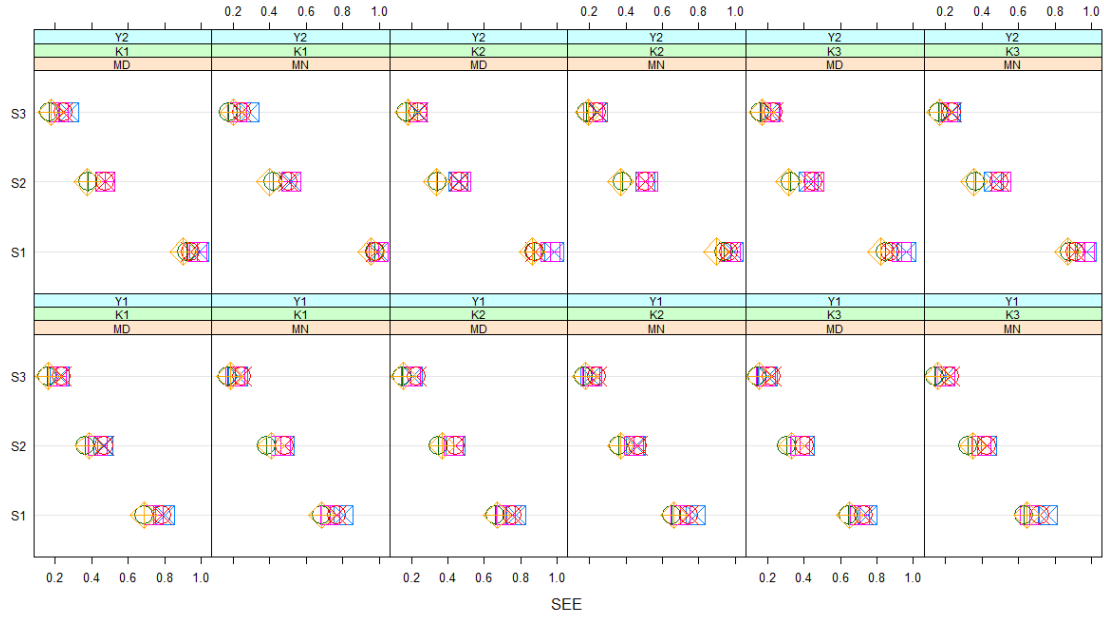
#### 4.5. Alt Problem 5'e İlişkin Bulgular ve Yorumlar

*“Çalışmada yer alan koşulların Kernel son tabakalama doğrusal, Kernel zincirleme doğrusal, Kernel son tabakalama eşit yüzdelikli, Kernel zincirleme eşit yüzdelikli ve MTK gerçek puan eşitleme yöntemleriyle elde edilen eşitlemenin standart hatasına (SEE) ortak etkisi nasıldır?”*

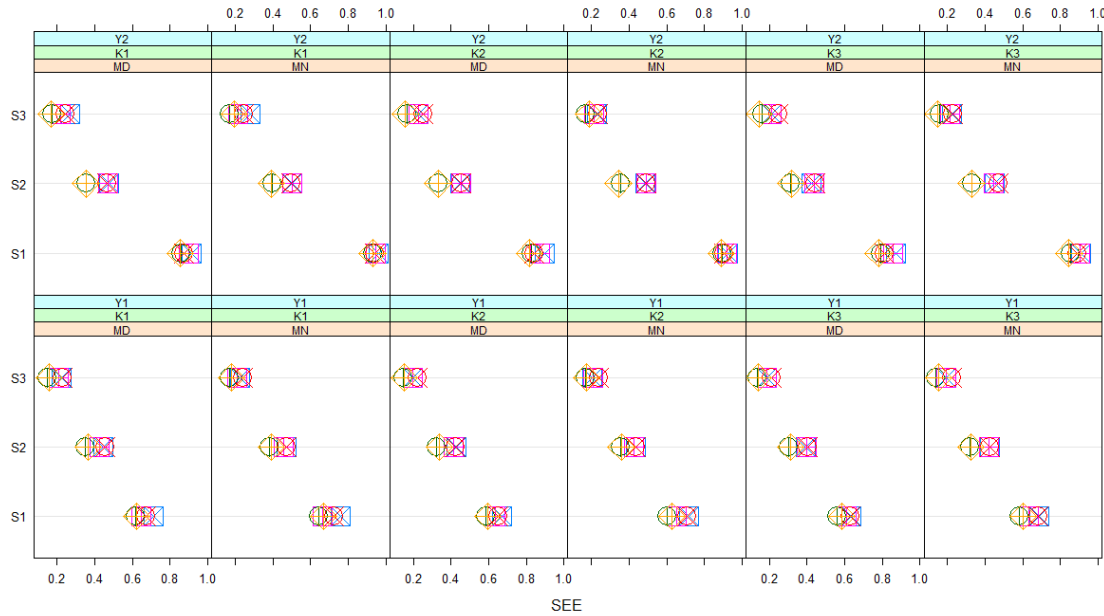
Bu alt problemde, çalışmada yer alan örneklem büyüklüğü, gruplararası yetenek dağılımı, ortak madde oranı ve ortak madde güçlük dağılımı etkileşimlerinin iç ve dış ortak testte eşitleme yöntemlerinden elde edilen eşitlemenin standart hatası üzerindeki ortak etkisi incelenmiştir. İç ortak test için çalışmada ele alınan koşulların eşitlemenin standart hatası üzerindeki ortak etkisi Şekil 4.30'da, dış ortak için ise Şekil 4.31'de yer almaktadır. Şekil 4.30 ve 4.31'da; S1,S2 ve S3 sırayla 500, 1500, 5000 örneklem büyüklüğünü, Y1 yetenek dağılımın benzer ve Y2 farklı olduğu koşulları; K1, K2, K3 ise sırayla %20, %30 ve %40 ortak madde oranını, MN mini ortak test koşulunu ve MD midi ortak test koşulunu belirtmektedir.

Şekil 4.30 ve 4.31'de yer alan grafikler incelendiğinde, bütün eşitleme yöntemlerinde örneklem büyüklüğü arttıkça yetenek dağılımı, ortak madde tipi, ortak madde oranı ve ortak madde güçlük dağılımının tüm düzeylerinde ortalama standart hatalarının azaldığı görülmektedir. Gruplar arası yetenek dağılımının farklı olduğu koşulda, ortalama standart hatadaki bu azalış örneklem büyüklüklerine göre farklılaşmaktadır. Örneklem büyüklüğünün 500'den 1500'e arttığında ortalama standart hatadaki azalma miktarı, örneklem büyüklüğünün 1500'den 5000'e arttığı duruma göre daha fazla olduğu görülmektedir. Ele alınan koşulların etkileşimleri örneklem büyüklüğüne göre incelendiğinde, 1500 ile 5000 örneklemelerden elde edilen standart hataların birbirine yakın olduğu söylenebilir. Gruplar arası yetenek dağılımının benzer olduğu koşulda ise örneklem arası ortalama standart hatadaki azalma miktarı daha azdır. Gruplar arası yetenek dağılımı farklı olduğunda, örneklem büyüklüğünün 500 iken ortalama standart hatanın arttığı görülmektedir. Ancak diğer örneklem büyüklüklerinin yetenek dağılımının farklılaşmasından çok fazla etkilenmemiştir. Gruplar arası yetenek dağılımı benzer olduğunda, ele alınan koşulların etkileşimlerine göre eşitleme yöntemlerinin çok fazla farklılaşmadığı ve benzer ortalama eşitleme hataları verdiği görülmektedir. Dış ortak testte, iç ortak teste göre daha düşük ortalama standart hata elde edilmiştir. Ancak ortalama standart hatalarındaki bu azalma

miktarının önemli ölçüde olmadığı görülmektedir. Ortak madde oranı arttıkça, bütün koşullarda standart hatanın azaldığı görülmüştür, ancak bu azalma miktarı da önemli ölçüde değildir. Ayrıca ortak madde güçlük dağılımının farklılaşmasının eşitleme yöntemlerinin ortalama standart hatalarını önemli ölçüde değiştirmedeği de söylenebilir.



**Şekil 4.30. İç ortak teste çalışmada ele alınan koşulların eşitlemenin standart hatasına (SEE) ortak etkisi**



**Şekil 4.31. Dış ortak teste çalışmada ele alınan koşulların eşitlemenin standart hatasına (SEE) ortak etkisi**

#### 4.6. Tartışma

Bu çalışmada, KE son tabakalama eşit yüzdelikli, KE son tabakalama doğrusal eşitleme, KE zincirleme eşit yüzdelikli, KE zincirleme doğrusal eşitleme ve MTK gerçek puan eşitleme yöntemlerinin örneklem büyüklüğü, yetenek dağılımı, ortak madde oranı, ortak madde tipi ve ortak madde güçlük dağılımı değişkenlerinden nasıl etkilendiği RMSD ve SEE değerlendirme kriterlerine göre incelenmiştir. Bu çalışmadan elde edilen bulgular ele alınan koşullara göre tartışılmıştır.

##### ***Örneklem büyüklüğü;***

Eşitleme yöntemleriyle ilgili yapılan çalışmalarda örneklem büyüklüğü sıklıkla kullanılan değişkenlerden biridir. Örneklem büyüklüğünün artması araştırmacıya puan dağılım kestirimleri ve MTK model uyumu hakkında daha fazla bilgi sağlamakta, bu da daha doğru ve daha kesin eşitleme sonuçlarının elde edilmesine olanak vermektedir (Kolen & Brennan, 2004). Örneklem büyüklüğünün özellikle standart hata üzerinde doğrudan etkisi bulunmaktadır (Budescu, 1985). Kolen ve Brennan (2004) NEAT deseninde her bir test formunun doğrusal eşitleme ve MTK Rasch eşitleme yöntemleri için 400, eşit yüzdelikli eşitleme ve MTK 3 PLM'ye göre eşitleme yöntemleri için ise en az 1500 kişilik örneklem büyüklüğüne sahip olması gerektiğini belirtmişlerdir. Bu çalışmada örneklem büyüklüğünün 500 olduğu koşulda en yüksek standart hatalar MTK gerçek puan eşitleme yönteminden, en düşük hatalar ise doğrusal eşitleme yöntemlerinin kullanıldığı durumlarda elde edilmiştir. Örneklem büyüklüğünün 1500 olduğu koşulda, bütün yöntemlerden elde edilen standart hataların azaldığı ve örneklem büyüklüğünün 5000 olduğu koşullarda ise standart hataları birbirine yaklaştığı görülmektedir. Örneklem büyüklüğü 5000 iken, bütün yöntemlerden elde edilen standart hatanın tüm puan ölçeği boyunca 0,5'in altında olduğu görülmüştür. Bu çalışmadan elde edilen bulgular Kolen ve Brennan'nın (2004) MTK gerçek puan ve eşit yüzdelikli eşitleme yöntemleri için örneklem büyüklüğünün en az 1500 olması ile ilgili önerisini desteklemektedir.

Bu çalışmada kullanılan Kernel ve MTK gerçek puan eşitleme yöntemlerinde örneklem büyüklüğü arttıkça, standart hata ve toplam hata azalmıştır. Örneklem büyüklüğü 5000 olduğunda standart hata değerlerinin birbirine daha çok benzediği ve MTK gerçek puan eşitleme yönteminde daha düşük toplam hatalar elde edildiği görülmüştür. Bu bulgu Cui (2006), Hou (2007), ile Kim (2014), Meng (2012), Norman

Dvorak (2009) ile Wang ve diğerleri'nin (2008) elde ettiği sonuçlar ile paralellik göstermektedir. Kim (2014), Kernel son tabakalama ve zincirleme eşit yüzdelikli eşitleme ile geleneksel eşitleme yöntemlerini karşılaştırdığı çalışmasında örneklem büyüklüğü arttıkça standart hatanın azaldığını belirtmektedir. Meng (2012), MTK gerçek puan eşitleme yöntemleriyle, Kernel son tabakalama eşit yüzdelikli eşitleme yöntemini karşılaştırdığı çalışmasında örneklem büyüklüğü arttıkça standart hatanın azaldığını rapor etmiştir. Norman Dvorak'un (2009) KE ile MTK gerçek puan eşitleme yöntemlerini örneklem büyüklüğü, ortalama faktör yükleri, test uzunluğu ve ortak madde oranı değişkenlerine göre karşılaştırdığı çalışmasında örneklem büyüklüğünün etkisinin ortalama faktör yükü, ortak madde oranı ve test uzunluğuna bağlı olduğunu ve örneklem büyüklüğünün 1000 ve 2000 olduğu koşulda bazı durumlarda KE eşitleme yöntemi bazı durumlarda ise MTK gerçek puan eşitleme yönteminin, ancak 10000 örnekleme MTK gerçek puan eşitlemenin daha doğru sonuçlar verdiğini ifade etmiştir. Cui (2006), örneklem büyüklüğü artışının standart hatayı önemli ölçüde azalttığını ve bu durumun sürpriz olmadığını belirtmiştir. Örneklem büyüklüğünün özellikle standart hata üzerinde daha etkili bir koşul olduğu görülmüştür. Elde edilen sonuçlar genel olarak değerlendirildiğinde, büyük örneklem kullanıldığında eşitleme yöntemlerinin daha doğru sonuçlar verdiği görülmüştür. Bu durum puan dağılımı kestirimlerinin ve model veri uyumunun örneklem büyüklüğünden etkilenmesinden kaynaklanmaktadır. Ayrıca büyük örneklemde, evren ile örneklemden elde edilen eşitleme ilişkisi birbirine yaklaşmakta ve bu durum da hatanın azalmasına neden olmaktadır.

### ***Gruplar arası yetenek dağılımı;***

Bu çalışmada ele alınan diğer bir koşul da gruplar arası yetenek dağılımıdır. Bütün eşitleme yöntemlerinde ele alınan koşullara göre gruplar arası yetenek dağılımı farklı olduğunda standart hata ve toplam hatanın arttığı görülmektedir. Bir başka deyişle, eşitleme yöntemleri daha düşük performans göstermiştir. Bu bulgu, Godfrey (2007), Hagge ve Kolen (2011), Holland ve diğerleri (2006), Hou (2007), Kim (2014), Meng (2012), Powers ve Kolen (2011), Power ve diğerleri (2011), Sinharay ve Holland (2006a), Sinharay ve Holland (2007) ve Wang ve diğerleri'nin (2008) bulguları ile tutarlılık göstermektedir. Ayrıca gruplar arası yetenek dağılımının farklılaşmasından KE zincirleme eşitleme yöntemlerinin, KE son tabakalama eşitleme yöntemlerine göre

daha az etkilendiği görülmüştür. Hou (2007), KE son tabakalama eşitleme yönteminde gruplar arası yetenek dağılımı farkı arttıkça toplam hatanın arttığını ve gruplar arası yetenek dağılımı farklılaştıkça X testinin A ortak testi ve Y testinin A ortak test üzerindeki koşullu puan dağılımı varsayımının aynı olma olasılığının daha düşük olduğunu ve bu durumun da eşitleme hatasının artmasına neden olduğunu ifade etmiştir. Kim (2014), gruplar arası yetenek dağılımı farklılaşmasından zincirleme ve KE zincirleme eşit yüzdellikli eşitleme yönteminin daha az etkilendiğini, bu yöntemlerin daha düşük hata verdiğini rapor etmiştir. Wang ve diğerleri (2008), gruplar arası yetenek dağılımı farklı olduğunda zincirleme eşit yüzdellikli; gruplar arası yetenek dağılımı farkı az ise frekans kestirimi eşit yüzdellikli eşitleme yönteminin daha düşük hata verdiğini ifade etmiştir. Standart hata açısından ise gruplar arası yetenek dağılımı benzer olduğunda frekans kestirimi eşitleme zincirleme eşitlemeden daha düşük standart hata vermiştir. Holland ve diğerleri (2006), KE zincirleme ve KE son tabakalama eşitleme yöntemlerinin gruplar arası yetenek dağılımı benzer olduğunda birbirine yakın, farklı olduğunda ise zincirleme eşitleme yönteminin daha iyi sonuçlar verdiğini belirtmişlerdir. Power ve diğerleri'nin (2011), karma testlerde gruplar arası yetenek dağılımının MTK gerçek, MTK gözlenen puan, zincirleme ve frekans kestirimi eşit yüzdellikli eşitleme yöntemlerinde karşılaştırdıkları çalışmada, frekans kestirimi eşitleme metodunun zincirleme eşitleme yöntemine göre gruplar arası yetenek dağılımı farklı olduğunda daha fazla sistematik hatanın elde edildiğini rapor etmişlerdir.

Gruplar arası yetenek dağılımı farkı arttığında, bütün eşitleme yöntemlerinde uç puanlardaki hata miktarlarının arttığı görülmüştür. Bu bulgu Godfrey (2007) eşitleme yöntemlerinin uç puanlarda kriter eşitleme yönteminden (eşit yüzdellikli eşitleme) sapmalar gösterdiğini belirttiği çalışmasının bulgularıyla benzerlik göstermektedir. Meng (2012) yaptığı çalışmada gruplar arası yetenek dağılımı farkı az olduğunda Kernel eşitlemenin, farkın arttığı koşulda ise MTK gerçek puan eşitlemenin daha iyi sonuçlar verdiğini bulmuştur.

Bu doğrultuda, gruplar arası yetenek dağılımının farklılaşmasının hata üzerinde oldukça önemli bir etkiye sahip olduğu ve hatanın artmasına neden olduğu bulunmuştur. Bu durumun nedeni, gruplar arası yetenek dağılımı farklı olduğunda,

ortalaması yüksek olan grubun daha çok maddeyi doğru cevaplarırken, ortalaması düşük olan grubun daha çok maddeyi yanlış cevaplamasıdır.

### **Ortak Madde Özellikleri;**

Bu çalışmada ele alınan diğer bir koşul ise ortak testin özelliklerinin ele alındığı ortak testin tipi (dış ve iç), ortak madde oranı ve ortak madde güçlük dağılımıdır.

**Ortak Testin Tipi:** Denk olmayan gruplarda ortak madde deseninde ortak testin özellikleri önemli bir role sahiptir. Test uygulamalarında, dış ortak test için ayrı bir süre verilirken, iç ortak test ana test ile birlikte verilmektedir. Bu nedenle iç ortak testte yer alan maddelerin içerik etkisi dış ortak teste göre daha önemli bir problem olarak ortaya çıkmaktadır (Kolen & Brennan, 2014). Ayrıca KE yöntemlerinde iç ortak testte log-lineer düzgünleştirmeden dolayı yapısal sıfırlar durumu problemi ortaya çıkarken, dış ortak testte bu durum ortaya çıkmamaktadır. Bu nedenle KE yöntemlerinde iç ortak test kullanırken eşitlemeye başlamadan önce yapısal sıfırlar durumunun ortaya çıkıp çıkmadığı kontrol edilmelidir. Çalışmadan elde edilen bulgulara göre dış ortak testte standart ve toplam hatanın iç ortak teste göre daha az olduğu bulunmuştur. Dış ortak testte, toplam test daha fazla maddeye sahip olduğundan, toplam test ile ortak test arasındaki korelasyonun artma eğiliminde olması daha az hatanın elde edilmesine neden olmuş olabilir. Kim (2014), dış ortak testte sistematik, standart ve toplam hatanın iç ortak teste göre daha az olduğunu bulmuştur. von Davier ve diğerleri (2006), iç ve dış ortak testte KE ile geleneksel eşitleme yöntemlerini karşılaştırdıkları çalışmada iç ve dış ortak testten elde edilen sonuçların benzer olduğunu ifade etmiştir. Bu çalışmada elde edilen sonuçlar Kim (2014)'un çalışmasıyla benzer olduğunda, von Davier ve diğerleri'nin (2006) çalışması ile çelişmektedir. Bu durum von Davier ve diğerleri'nin (2006) çalışmalarının gerçek veri seti üzerinden gerçekleşmesinden kaynaklanıyor olabilir.

**Ortak Madde Oranı:** Budescu (1985), ortak madde oranı arttıkça standart hatanın azaldığını ifade etmiştir. Bu çalışmada bütün eşitleme yöntemlerinde ortak madde oranı arttıkça hatanın (standart ve toplam) azaldığı görülmüştür. Bu bulgular ortak madde oranı arttıkça standart ve toplam hatanın azaldığı sonucunu elde eden çalışmalarla paralellik göstermektedir (Hou, 2007; Kim, 2014; Meng, 2012; Norman Dvorak, 2009; Ricker ve von Davier, 2007; Sinharay ve Holland, 2006b; Wang ve diğerleri, 2008). Norman Dvorak (2009), dış ortak testte ortak madde oranının



ortalama faktör yük, test uzunluğu ve örneklem büyüklüğüne bağlı olduğunu, ortak madde oranının %30 olduğu durumda MTK gerçek puan eşitleme yönteminin daha iyi sonuçlar verdiğini bulmuşlardır. Hou (2007), KE frekans kestirimi ile diğer geleneksel eşitleme yöntemlerini karşılaştırdığı çalışmada, bütün eşitleme yöntemlerinde ortak madde oranı arttıkça standart hata ve toplam hatanın azaldığı; ortak madde oranı iki katına çıktığında toplam hatadaki azalmanın %40 - %50 oranında olduğunu ifade etmiştir. Ayrıca, ortak madde oranı arttıkça ortak madde ile toplam test arasındaki korelasyonun arttığını ve bu durumun da hatanın azalmasına neden olabileceğini belirtmiştir. NEAT deseninde farklı yetenek düzeylerindeki bireylere uygulanan test formlarından elde edilen parametrelerin belirlenen kalibrasyon yöntemiyle aynı ölçek üzerine yerleştirilmesi ortak madde üzerinden gerçekleştirilmektedir. Bu nedenle ortak madde sayısı arttıkça, ortak maddelerden elde edilen bilgi de artmakta, bu durum da hatanın azalmasının neden olmaktadır.

**Ortak Madde Güçlük Dağılımı:** Çalışmadan elde edilen diğer bulgu da midi ortak test koşulunda standart ve toplam hatanın mini ortak test koşuluna göre daha az olmasıdır. Bu bulgu Antal ve diğerleri (2014), Fitzpatrick ve Skorupski (2016), Kim (2014), Liu ve diğerleri (2009), Sinharay ve Holland (2006b, 2007) ile Sinharay ve diğerleri'nin (2012), elde ettiği sonuçlar ile paralellik göstermektedir. Fitzpatrick ve Skorupski (2016), MTK gerçek puan eşitleme yöntemlerinin iç ortak testte midi ortak testin mini ortak test kadar iyi sonuçlar verdiği bulgusuna ulaşmışlardır. Sinharay ve Holland'ın (2006b, 2007), zincirleme ve son tabakalama eşit yüzdelikli eşitleme yöntemlerini karşılaştırdıkları çalışmada grupların yetenek dağılımları arasındaki fark arttıkça midi ortak testin mini ortak teste göre daha düşük standart ve toplam hata verdiğini ve iç ortak testte midi ve semi midi ortak testin, mini ortak test kadar iyi sonuçlar verdiğini, dış ortak testte ise midi ortak testin daha avantajlı olduğu sonucuna varmışlardır. Kim (2014) midi ortak testte sistematik ve toplam hata daha az iken, mini ortak testte standart hatanın daha az olduğunu rapor etmiştir. Liu ve diğerleri (2009, 2011a) gerçek veri ile zincirleme eşitleme yöntemleri ile Tucker ve frekans kestirimi eşitleme yöntemlerini karşılaştırdıkları çalışmada standart ve toplam hata açısından midi ortak testin mini ortak testten daha iyi sonuçlar verdiğini ve ayrıca eğer eşitlemede orta puan ölçeği daha önemli ise midi ortak testin mini ortak teste tercih edilebileceğini, ancak eşitlemede yüksek puan ölçeğinden elde edilen puanlar

önemli ise, midi ve mini ortak test arasındaki fark çok az olsa da testin özelliklerine göre hangi ortak testin kullanılması gerektiğine karar vermeleri gerektiğini ifade etmişlerdir.

Test geliştiriciler içerik ve istatistiksel açıdan toplam test ile aynı özelliklere sahip ortak test geliştirmeyi amaçlamaktadır. Ancak bu özelliklerin sağlanması gerekli değildir (Fitzpatrick & Skorupski, 2016; Sinharay ve Holland, 2006b). Bu gereklilikten dolayı, mini ortak test çok güç ve çok kolay maddeleri de içermekte ve bu durum zayıf madde istatistiklerine sahip maddelerin de testlerde yer almasına neden olmaktadır. Dış ortak testte ise içerik, ortalama güçlük gibi özellikler önemli iken, ortak madde güçlük dağılımının nasıl olduğu o kadar da önemli değildir. Simülasyon ve gerçek uygulamalara ait eşitleme durumlarında midi ortak testin mini ortak testten daha iyi sonuçlar verdiği ortaya çıkarsa, test geliştirme için önemli bir adım olacaktır (Sinharay ve Holland, 2006b). Fitzpatrick ve Skorupski (2016), test geliştiricilerin toplam testle benzer özelliklere sahip ortak testler ile madde güçlük parametresinin daha az değişkenliğe sahip olduğunda yani midi ortak testlerin benzer sonuçlar verdiğini ve hatta bazı durumlarda test formlarının eşitlenmesi için iyi bir seçenek olduğundan emin olabileceklerini belirtmiştir. Bu çalışmada da dış ortak testte çoğu koşulda ve iç ortak testte ise birçok koşulda benzer sonuçlar elde edilmiştir. Ayrıca iç ortak testte gruplar arası yetenek dağılımı benzer olduğunda midi ortak test mini ortak teste yakın ya da az daha yüksek hatalar vermişken, gruplar arası yetenek dağılımı farklı olduğunda midi ortak test daha düşük hata vermiştir. Bunun bir nedeni grupların yetenek dağılımı farklı olduğunda; mini ortak testte her iki grup tarafından çok kolay maddeler doğru cevaplanırken, çok zor maddeler yanlış cevaplanmakta, ancak orta güçlükteki maddeler yetenek düzeyi iyi grup tarafından cevaplanabilirken, düşük yetenek düzeyindeki grup tarafından cevaplanamamasıdır. Yani midi ortak testte maddeler daha kolay olduğu için daha çok birey tarafından cevaplanmakta ve bu durumda ortak testten daha çok bilgi elde edilmesine neden olmaktadır. Bu durum da hatayı azaltmaktadır.

### ***Test Eşitleme Yöntemleri;***

Eşitleme yöntemleri bazında bulgular değerlendirildiğinde; doğrusal eşitleme yöntemlerinin, eşit yüzdeliği eşitleme ve MTK gerçek puan eşitleme yöntemlerine göre daha düşük standart hata verdiği bulgusuna ulaşılmıştır. Yani, h parametresinin

seçiminin standart hata değerlerinin azalmasına neden olmuştur. Elde edilen bu bulgunun Choi (2009), Liou ve diğerleri (1997) ve Mao'nun (2006) bulgularıyla tutarlı olduğu görülmektedir. Choi'nin (2009) yaptığı çalışmada örneklem büyüklüğü, test uzunluğu, bant genişliği ve öndüzgünleştirme parametresi değişkenleriyle Kernel eşitleme ile geleneksel eşitleme yöntemlerini karşılaştırdığı çalışmasında, doğrusal Kernel eşitleme yöntemlerinin eşit yüzdelikli eşitleme yöntemlerine göre daha düşük standart hata verdiğini bulmuştur. Mao (2006), Kernel eşitlemede standart grup deseninde örneklem büyüklüğü, yetenek dağılımı, bant genişliğinin seçimi ve düzgünleştirme parametre derecesi faktörlerine göre standart hata değerlerini karşılaştırmış ve grupların dağılımının normal veya aynı olmasından standart hatanın çok fazla etkilemediğini, bunun da Kernel eşitlemede standart hatanın normallik varsayımı gerektirmemesinden kaynaklandığını belirtmiştir. Liou ve diğerlerinin (1997) yaptıkları çalışmada geniş bant genişliğinde, Gauss Kernel metotun standart hatayı azalttığı sonucuna ulaşmışlardır. Kernel doğrusal eşitleme yöntemlerinin RMSD açısından diğer yöntemlere göre daha yüksek olduğu görülmektedir. Ricker ve von Davier (2007), RMSD hesaplanırken kullanılan kriter eşitleme yönteminin doğrusal olmadığı (eşit yüzdelikli eşitleme gibi) durumlarda, doğrusal eşitleme yöntemlerinin lineer eşitleme fonksiyonlarını kullanmasından dolayı RMSD değerlerinin büyük çıktığını ve kriter eşitleme için doğrusal eşitleme yöntemi kullanılırsa eşitleme sonuçlarına ilişkin hataların daha iyi sonuçlar vereceğini ifade etmişlerdir. Bu çalışmada da kriter eşitleme olarak eşit yüzdelikli eşitleme yöntemi kullanıldığından, toplam hatanın doğrusal eşitleme yöntemlerinin aleyhine elde edilmesine neden olmuş olabilir.

Uç puanlarda MTK gerçek puan eşitleme, KE yöntemlerine göre daha düşük hata vermiştir. Cook ve Eignor (1991), uç puanlarda MTK gerçek puan eşitlemenin KTK'ya dayalı eşitlemeden daha iyi eşitleme sonuçlar verdiğini belirtmiştir. Bu çalışmadan elde edilen bulgu da Cook ve Eignor (1991) ile örtüşmektedir. Kernel eşitleme yöntemlerinin uç puanlarda orta puanlara göre daha yüksek standart hata verdiği bulunmuştur. KE yöntemlerinde örneklem büyüklüğü arttıkça uç puanlardaki hata miktarlarının azaldığı görülmektedir. Mao, von Davier ve Rupp (2006) yaptıkları Kernel eşitleme ile gözlenen puan eşitleme yöntemlerinin NEAT deseninde iç ve dış ortak testte karşılaştırdıkları çalışmada, KE ile diğer eşitleme yöntemleri arasında

sadece düşük puan aralığında geniş farklılıklar bulmuşlardır. Mao (2006) ve Mao, von Davier ve Rupp (2006), Kernel eşitleme yöntemlerinin uç puanlarda yüksek standart hata vermesini yığılmalı puan dağılımlarının süreklileştirilmesinde Gauss Kernel yönteminin kullanmasından kaynaklandığını belirtmişlerdir. Gauss Kernel süreklileştirme metodunda, puan ölçeği  $+\infty$  ile  $-\infty$  arasında yer almakta ve bu durum uç puanlardan daha yüksek standart hatanın ortaya çıkmasına sebep olmaktadır.

Godfrey (2007), Meng (2012) ve Norman Dvroat (2009) tarafından Kernel eşitleme yöntemleri ile MTK gerçek puan eşitleme yöntemlerinin karşılaştırıldığı çalışmalarda, Kernel eşitleme yöntemlerinin, MTK gerçek puan eşitleme yöntemleri kadar iyi sonuçlar verdiği ve hatta orta puan ölçeğinde MTK gerçek puan eşitlemeden daha iyi sonuçlar elde edildiğini belirtilmiştir. Ayrıca özellikle gerçek puan eşitleme yöntemlerinin kullanılmasının uygun olmadığı durumlarda KE yöntemlerinin kullanılabileceğini ifade etmişlerdir. KE yöntemleri ile MTK gerçek puan eşitleme yöntemlerinin çeşitli koşullara göre ele alındığı bu çalışmada da, KE yöntemlerinin bazı koşullarda MTK gerçek puan eşitleme kadar iyi sonuçlar verdiği bulunmuştur. Bu çalışmadan elde edilen bulgular ve önceki çalışmalar ışığında, test formlarının eşitlenmesi sonucunda elde edilen eşitlenmiş puanlar eşitleme yöntemlerine göre farklılık göstermektedir. Testin amacı doğrultusunda, eşitleme yöntemlerinin güçlü ve zayıf yönleri dikkate alınarak, eşitleme yöntemine karar verilmelidir. Seçilecek eşitleme yönteminin diğer eşitleme yöntemlerinden her koşulda iyi sonuç vermeyeceği de unutulmamalıdır.

## 5. SONUÇ ve ÖNERİLER

Bu bölümde araştırmanın bulgularına dayalı olarak ulaşılan sonuçlara ve bu sonuçlardan yola çıkarak sunulabilecek önerilere yer verilmiştir.

### 5.1. Sonuçlar

Bu çalışmada Kernel eşitleme yöntemleri ile MTK gerçek puan eşitleme yöntemi örneklem büyüklüğü, gruplar arası yetenek dağılımı, ortak madde oranı ve ortak madde güçlük dağılımı değişkenleri açısından karşılaştırılmıştır. Kernel eşitleme yöntemlerinden Kernel zincirleme eşit yüzdelikli, Kernel son tabakalama eşit yüzdelikli, Kernel zincirleme doğrusal ve Kernel son tabakalama doğrusal eşitleme kullanılmıştır. Ele alınan koşullar doğrultusunda veriler üretilmiş ve hangi yöntemin daha az hatalı sonuçlar verdiği incelenmiştir. Bu amaçla, eşitlemenin doğruluğu için değerlendirme kriterlerinden SEE ve RMSD kullanılarak aşağıdaki sonuçlara ulaşılmıştır.

Araştırmadan elde edilen sonuçlar, kriter eşitleme olarak eşit yüzdelikli eşitleme yönteminin kullanıldığı durumda elde edilmiştir. Bu doğrultuda;

- Çalışmada ele alınan koşullara göre, örneklem büyüklüğü arttıkça bütün eşitleme yöntemlerinden elde edilen standart hata ve toplam hata değerlerinin azaldığı görülmektedir. Ancak örneklem büyüklüğünün standart hata üzerindeki etkisi, toplam hata üzerindeki etkisine göre daha fazladır. 500 örneklem için en yüksek standart hata değeri MTK gerçek puan eşitleme yönteminde elde edilmişken, örneklem büyüklüğü arttıkça yöntemlerden elde edilen standart hata değerleri birbirine yaklaşmıştır. Toplam hata açısından ise, MTK gerçek puan eşitleme küçük örneklemde en yüksek hata, örneklem büyüklüğü arttıkça düşük hata verdiği sonucuna ulaşılmıştır.
- Gruplar arası yetenek dağılımı farklı olduğunda bütün eşitleme yöntemlerinde hatanın arttığı, ancak hatadaki artışın eşitleme yöntemlerine göre farklılık gösterdiği sonucuna ulaşılmıştır. Gruplar arası yetenek dağılımı benzer olduğunda son tabakalama eşitleme yöntemleri zincirleme eşitleme yöntemlerine göre daha düşük hata vermiştir. Gruplar arası yetenek dağılımı farklı olduğunda ise zincirleme eşitleme yöntemlerinin daha iyi sonuçlar verdiği elde edilmiştir. Ayrıca Kernel eşitleme yöntemlerinde gruplar arası yetenek

dağılımı farklı olduğunda uç puanların, MTK gerçek puan eşitlemede ise puan ölçeğinin orta yüksek kısmında daha yüksek standart hata verdiği sonucuna ulaşılmıştır.

- Bütün eşitleme yöntemlerinde ortak madde oranı arttıkça standart ve toplam hata azalmıştır. Bu azalma miktarı eşitleme yöntemlerine göre farklılık göstermektedir. Özellikle son tabakalama ve MTK gerçek puan eşitleme yöntemlerinde hatanın daha çok azaldığı görülmüştür. Son tabakalama eşitleme yöntemlerinde ortak madde sayısı arttıkça testlerin ortak test üzerindeki eşit dağılım varsayımının sağlanma olasılığının hatanın azalmasında etkili olmuştur.
- Bütün eşitleme yöntemlerinde iç ve dış ortak test koşulunun hata üzerinde etkisi olduğu bulunmuştur. Dış ortak testte, iç ortak teste göre daha düşük standart ve toplam hata elde edilmiştir.
- Ortak madde güçlük dağılımının da standart hata ve toplam hata üzerinde etkisi olduğu bulunmuştur. İç ortak testte, gruplar arası yetenek dağılımı benzer olduğunda mini ve midi ortak test benzer/yakın sonuçlar verirken, gruplar arası yetenek dağılımı farklı olduğunda midi ortak testin daha düşük hata verdiği sonucuna ulaşılmıştır. Dış ortak testte ise midi ortak test, mini ortak teste göre daha kararlı sonuçlar vermiştir.
- Örneklem büyüklüğünün 5000 olduğu koşulda, dış ortak testte bütün ortak madde oranlarında, iç ortak testte ise %20 ve %30 ortak madde oranında MTK gerçek puan eşitlemenin Kernel eşitlemeye göre daha düşük hata verdiği bulunmuştur.
- Kernel eşitlemeye dayalı yöntemleri kendi içinde karşılaştırıldığında, doğrusal eşitleme yöntemlerinin (geniş h) eşit yüzdelikli eşitleme yöntemlerine (ideal h) göre daha düşük standart hata verdiği bulunmuştur. Toplam hata açısından ise eşit yüzdelikli eşitleme yöntemleri daha iyi çıkmıştır.
- Bütün koşullarda Kernel eşitleme yöntemlerinin orta puan ölçeğinde daha düşük standart hata verdiği bulunmuştur. Ancak puan frekansının daha az olduğu uç puanlarda daha yüksek standart hata verdiği sonucuna ulaşılmıştır.

### ***Koşulların ortak etkileri incelendiğinde;***

Gruplar arası yetenek dağılımının benzer olduğu koşulda, ortak madde oranı arttıkça doğrusal eşitleme yöntemlerinin diğer eşitleme yöntemlerine daha yakın eşitleme hatası verdiği bulunmuştur.

- MTK ve eşit yüzdeli eşitleme yöntemlerinin, örneklem büyüklüğünün 5000 olduğu koşulda gruplar arası yetenek dağılımı farklı olduğunda eşitleme hatalarının benzer olduğu sonucuna ulaşılmıştır.
- 1500 ile 5000 örneklem büyüklüklerinden elde edilen standart hataların birbirine yakın olduğu çıkmıştır.
- Dış ortak testte, iç ortak teste göre daha düşük ortalama standart hata elde edilmiştir. Ancak diğer koşulların ortak etkisiyle ortalama standart hatalarındaki bu azalma miktarı çok küçük bulunmuştur.
- Genel olarak örneklem büyüklüğü, gruplar arası yetenek dağılımı, ortak madde tipi, ortak madde oranı ve ortak madde güçlük dağılımının ve bu koşulların etkileşimlerinin eşitleme yöntemlerinin standart hata ve eşitleme hatalarını etkilediği sonucuna ulaşılmıştır.

## **5.2. Öneriler**

Araştırma sonuçlarından yola çıkarak geliştirilen öneriler iki başlık halinde verilmiştir. Bu çalışma farklı eşitleme yöntemlerinin eşitleme sonuçlarını nasıl etkilediği konusunda araştırmacılara fikir verebilir.

### **5.2.1. Uygulayıcılara Yönelik Öneriler**

- Bu çalışmanın sonuçlarına dayanarak, test eşitleme çalışmalarında özellikle uç puanların fazla olduğu ve bireyler hakkında önemli kararların verileceği durumlarda Kernel eşitleme yöntemleri yerine MTK gerçek puan eşitleme yöntemleri kullanılabilir. Ancak MTK gerçek puan eşitlemenin yapılamayacağı koşullarda, Kernel eşitleme yöntemlerinin iyi bir seçenek olduğu görülmüştür.
- Daha kararlı eşitleme kestirimleri için dış ortak testte midi ortak test tercih edilebilir. İç ortak testte ise birçok koşulda midi ve mini ortak test benzer hatta gruplar arası yetenek dağılımının farklı olduğu koşulda midi ortak test daha az hatalı sonuçlar vermiştir. Grupların yetenek dağılımları farklı ise

midi ortak test seçilebilir. Ancak iç ortak testte içerik etkisi (kapsam ve madde istatistikleri vb.) de dikkate alınarak midi ortak testin kullanılmasına karar verilmelidir.

- Denk olmayan gruplarda ortak madde deseninde gruplar arası yetenek dağılımları arasındaki fark küçük ise KE son tabakalama eşitleme, geniş ise KE zincirleme eşitleme yöntemleri kullanılması önerilebilir.
- Bu çalışmada gruplar arası yetenek dağılımı farklı olduğunda, ortalama standart hatadaki bu azalışın örneklem büyüklüklerine göre farklılaştığı görülmüştür. Bu nedenle yetenek dağılımları farklı ise geniş örneklem büyüklüğü seçilmelidir.
- Bu çalışmada dış ortak testte daha küçük eşitleme hatası ve standart hata elde edilmiştir. Dış ortak testte bireyler daha fazla maddeye cevap vereceklerinden yorgunluk ve sıkılma gibi etkiler ortaya çıkabilir. Bu durum da test cevaplama sürecini olumsuz etkileyebilir. Ortak test tipine karar verirken bu koşullar da göz önüne alınmalıdır.
- Hem iç hem de dış ortak testte ortak madde oranı arttıkça eşitleme yöntemlerinin daha iyi performans gösterdiği görülmüştür. Bu nedenle daha doğru eşitleme sonuçları için ortak madde oranının %20'den daha fazla seçilmesi önerilmektedir.

### **5.2.2. İleri Araştırmalara Yönelik Öneriler**

- Bu çalışmada geniş ölçekli testlerdeki test uzunluğu dikkate alınarak test uzunluğu belirlenmiş ve sabit tutulmuştur. Test uzunluğunun eşitleme sonuçlarının doğruluğu ve kesinliği üzerindeki etkisini belirleyebilmek için farklı uzunluktaki testlerle çalışılabilir.
- Bu çalışmada 3 parametrelili lojistik modele göre veriler üretilerek eşitleme yöntemleri karşılaştırılmıştır, bundan sonraki çalışmalarda farklı modellere dayalı veriler üretilerek eşitleme çalışmaları yapılabilir.
- Bu çalışmada örneklem büyüklüğü 500, 1500 ve 5000 olarak alınmıştır, ileriki çalışmalarda daha küçük veya geniş ölçekli testler dikkate alınarak farklı büyüklüklerdeki örneklemle çalışılması önerilebilir. Ayrıca bu çalışmada grupların örneklem büyüklüğü eşit alınmıştır, bundan sonraki çalışmalarda grupların örneklem büyüklüğü farklı alınabilir.



- Bu çalışmada Kernel eşitleme yöntemleri (zincirleme ve son tabakalama) ile MTK gerçek puan eşitlemede ayrı kalibrasyon yöntemlerinden Haebara karşılaştırılmıştır. Daha sonraki çalışmalarda Kernel eşitleme yöntemleriyle geleneksel eşitleme yöntemleri, MTK gerçek puan eşitlemede farklı kalibrasyon ve MTK gözlenen puan eşitleme yöntemlerinin karşılaştırılması önerilebilir.
- Bu çalışmada ortak madde güçlük dağılımı mini ve midi ortak test olarak ele alınmıştır, bundan sonraki çalışmalarda semi midi ortak test koşulu da ele alınarak karşılaştırma yapılabilir.
- Bu çalışmada, Kernel eşitleme yöntemlerinde h parametresinin seçiminde ceza fonksiyonu kullanılmıştır, yapılacak çalışmalarda h parametresinin seçiminde başka fonksiyonlar (çapraz geçерleme, thumb vb.) kullanılarak yapılabilir.
- Bu çalışmada eşitleme yöntemlerini karşılaştırmak için simülasyon verisi kullanılmıştır. Benzer bir çalışma gerçek veri setleri kullanılarak yapılabilir.
- Bu çalışmada denk olmayan gruplarda ortak madde deseni kullanılmıştır. Daha sonraki çalışmalar için denk gruplarda ortak madde deseni kullanılarak sonuçlar karşılaştırılabilir. Ayrıca farklı eşitleme desenlerinde farklı eşitleme yöntemlerini kullanmaları önerilebilir.

## KAYNAKÇA

- Aksekioglu, B. (2017). *Madde tepki kuramına dayalı test eşitleme yöntemlerinin karşılaştırılması: PISA 2012 fen testi örneği* (Yayınlanmamış yüksek lisans tezi). Akdeniz Üniversitesi, Eğitim Bilimleri Enstitüsü. Mersin.
- Albano, A. A. (2010). *Equate: Statistical methods for test equating [Computer software manual]*. Available from <http://CRAN.R-project.org/package=equate> (R package).
- Andersson, B., & Davier, A. A. (2014). Improving the bandwidth selection in Kernel equating. *Journal of Educational Measurement*, 51(3), 223-238.
- Andersson, B., Branberg, K., Wiberg, M. (2013). Performing the Kernel Method of Test Equating with the Package kequate. *Journal of Statistical Software*, 55(6), 1–25.
- Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 508–600). Washington, DC: American Council on Education.
- Angoff, W. H. (1982). Summary and derivation of equating used at ETS. In P. W. Holland & D. B. Rubin (Eds.). *Test equating*, (pp. 55–57). NY: Academic Press.
- Antal, J., Proctor, T. P. & Melican, G.C., (2014). The Effect Of Anchor Test Construction On Scale Drift. *Applied Measurement in Education*, 27: 159–172, 2014.
- Aşiret, S. (2014). *Küçük örneklerde test eşitleme yöntemlerinin çeşitli faktörlere göre incelenmesi*(Yayınlanmamış yüksek lisans tezi). Mersin Üniversitesi, Eğitim Bilimleri Enstitüsü. Mersin.
- Brennan, R. L., Wang, T., Kim, S., & Seol, J. (2009). *Equating Recipes* (CASMA Monograph No. 1). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa. (Available on <http://www.education.uiowa.edu/casma>)
- Bozdağ, S.(2007). *Şans başarısının test eşitlemeye etkisi*(Yayınlanmamış yüksek lisans tezi). Mersin Üniversitesi, Sosyal Bilimler Enstitüsü. Mersin.
- Budescu, D. (1985). Efficiency of linear equating as a function of the length of the anchor test. *Journal of Educational Measurement*, 22, 13-20.
- Choi, S. I. (2009). *A Comparison of Kernel Equating and Traditional Equipercentile Equating Methods and the Parametric Bootstrap Methods for Estimating Standard Errors in Equipercentile Equating*. University of Illinois at Urbana-Champaign.
- Cook, L. L. (2007). Practical problems in equating test scores: A practitioner's perspective. Dorans, N. J., Pommerich, M., & Holland, P. W. (Eds.) *Linking and aligning scores and scales* (pp. 73-88). Springer New York.
- Cook, L. L., & Eignor, D. R. (1991). An NCME instructional module on IRT equating methods. *Educational Measurement: Issues and Practice*, 10(3), 37-45.
- Cook, L. L. & Eignor, D. R. (1983). Practical considerations regarding the use of item response theory to equate tests. In Hambleton, R. K. (Ed.) *Applications of item response theory* (pp.175–195). Vancouver, British Columbia: Educational Research Institute of British Columbia.

- Crocker, L. & Algina J. (1986). *Introduction to classical and modern test theory*. New York: Harcourt Brace Jovanovich College Publishers.
- Cui, Z.,& Kolen, M. J. (2008). Comparison of parametric and nonparametric bootstrap methods for estimating random error in equipercentile equating. *Applied Psychological Measurement*, 32(4), 334-347.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: Guilford Publications.
- Demirus, K.B. (2015). *Ortak maddelerin deęişen madde fonksiyonu gösterip göstermemesi durumunda test eşitlemeye etkisinin farklı yöntemlerle incelenmesi* (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü. Ankara.
- Dorans, N. J. & Holland, P. W. (2000). Population invariance and the equatability of tests: basic theory and the linear case. *Journal of Educational Measurement*, 37(4), 281-306.
- Dorans, N. J.,& Walker, M. E. (2007). Sizing up linkages. *Linking and aligning scores and scales* (pp. 179-198). Springer New York.
- Embretson, S. E. & Reise, S. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Ferrando, P. J. (2004). Kernel-smoothing estimation of item characteristic functions for continuous personality items: An empirical comparison with the linear and the continuous-response models. *Applied psychological measurement*, 28(2), 95-109.
- Fitzpatrick, J.,& Skorupski, W. P. (2016). Equating With Midtests Using IRT. *Journal of Educational Measurement*, 53(2), 172-189.
- Grant, M. C., Zhang, L., & Damiano, M. (2009). An Evaluation of Kernel Equating: Parallel Equating with Classical Methods in the SAT Subject Tests [TM] Program. Research Report. ETS RR-09-06. *Educational Testing Service*.
- Godfrey, K. E. (2007). *A comparison of Kernel equating and IRT true score equating methods* (Unpublished Doctoral Dissertation). The University of North Carolina.
- Gök, B. (2012). *Denk olmayan gruplarda ortak madde deseni kullanılarak madde tepki kuramına dayalı eşitleme yöntemlerinin karşılaştırılması* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Gündüz, T. (2015). *Test eşitlemede Madde Tepki Kuramına dayalı yetenek parametresine yönelik ölçek dönüştürme yöntemlerinin karşılaştırılması* (Yayımlanmamış Yüksek Lisans Tezi). Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- 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*(Unpublished Doctoral Dissertation). Iowa Üniversitesi.
- Hanson, B. A. & Beguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, 26(1), 3-24.

- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston: Academic Publishers Group.
- Hambleton, R. K. & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement, 12*, 38-47.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park: Sage.
- Harris, D. J., & Crouse, J. D. (1993). A study of criteria used in equating. *Applied Measurement in Education, 6*, 195-240.
- Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 187-220). Westport, CT: Praeger.
- Holland, P. W. (2007). A framework and history for score linking. Dorans, N. J., Pommerich, M., & Holland, P. W. (Eds.) *Linking and aligning scores and scales* (pp. 5-30). Springer New York.
- Holland, P. W., & Thayer, D. T. (2000). Univariate and bivariate loglinear models for discrete test score distributions. *Journal of Educational and Behavioral Statistics, 25*, 133-183.
- Holland, P. W., Dorans, N. J., & Petersen, N. S. (2006). 6 Equating Test Scores. *Handbook of statistics, 26*, 169-203.
- Holland, P., von Davier, A., Sinharay, S. & Han, N. (2006). *Testing the untestable assumptions of the chain and post-stratification equating methods for the NEAT design* (ETS RR-06-17). Princeton, NJ: Educational Testing Service.
- Hou, J. (2007). *Effectiveness of the hybrid Levine equipercentile and modified frequency estimation equating methods under the common-item nonequivalent groups design*. (Unpublished doctoral dissertation). University of Iowa: Iowa City, IA.
- İnci, Y. (2014). *Örneklem Büyüklüğünün Test Eşitlemeye Etkisi* (Yayınlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Kabasakal, K. A. (2014). *Değişen Madde Fonksiyonunun Test Eşitlemeye Etkisi* (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü. Ankara.
- Kahraman, H. (2012). *Düzenleştirilmiş puanların eşitleme hatasına etkisi* (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü. Ankara.
- Kan, A. (2010) Test Eşitleme: Aynı davranışı ölçen, farklı madde formlarına sahip testlerin istatistiksel eşitliğinin sınanması. *Eğitimde ve Psikolojide Ölçme Dergisi, 1*.
- Karkee, T. B., & Wright, K. R. (2004). Evaluation of Linking Methods for Placing Three-Parameter Logistic Item Parameter Estimates onto a One-Parameter Scale. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

- Keleciođlu, H.(1994). *Öđrenci Seçme Sınavı Puanlarının Eşitlenmesi Üzerine Bir Çalıřma* (Yayınlanmamıř Doktora Tezi). Hacettepe Üniversitesi. Sosyal Bilimler Enstitüsü. Ankara.
- Keller III, R. R. (2007). *A comparison of item response theory true score equating and item response theory-based local equating.*(Unpublished doctorate thesis). University of Massachusetts.
- Kim, S. H.,& Cohen, A. S. (2002). A comparison of linking and concurrent calibration under the graded response model. *Applied Psychological Measurement*, 26(1), 25-41.
- Kim, H.Y., (2014). *A comparison of smoothing methods for the common item nonequivalent groups design.* (Unpublished Doctoral Dissertation). Iowa Üniversitesi.
- Kim, S.,& Kolen, M. J. (2006). Robustness to format effects of IRT linking methods for mixed-format tests. *Applied Measurement in Education*, 19(4), 357-381.
- Kim, S.H.,& Cohen, A. S. (1998). Comparison of linking and concurrent calibration under item response theory. *Applied Psychological Measurement*, 22(2), 131-43.
- Kim, S.,& Lee, W. (2004). *IRT scale linking methods for mixed-format tests* (ACT Research Report). Iowa City, IA: ACT, Inc.
- Kilmen, S. (2010). *Madde tepki kuramına dayalı test eşitleme yöntemlerinden kestirilen eşitleme hatalarının örneklem büyüklüđü ve yetenek dağılımına göre karşılaştırılması.* (Yayınlanmamıř doktora tezi). Ankara Üniversitesi, Eğitim Bilimler Enstitüsü. Ankara.
- Kolen, M. J. (1988). An NCME instructional module on traditional equating methodology. *Educational Measurement: Issues and Practice*, 7, 29-36.
- Kolen, M.J.(2006). Book Review (*The Kernel method of test equating*). *Psychometrika*, 71(1), 211-214. DOI: 10.1007/s11336-005-1364-0.
- Kolen, M. J.,& Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices*. New York: Springer.
- Kolen, M. J.,& Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices (3 Edtion)*. New York: Springer.
- Kolen, M. J. (2007). Data collection designs and linking procedures. In *Linking and aligning scores and scales* (pp. 31-55). New York: Springer
- Lee, W. C. & Ban, J. C. (2010). Comparison of IRT Linking Procedures. *Applied Measurement in Education*, 23(1), 23-48.
- Lee, Y. S. (2007). A comparison of methods for nonparametric estimation of item characteristic curves for binary items. *Applied Psychological Measurement*, 31, 121-134.
- Lee, Y. H.,& von Davier, A. A. (2009). Equating through alternative kernels. In *Statistical models for test equating, scaling, and linking* (pp. 159-173). Springer New York.
- Liang, T.,& von Davier, A. A. (2014). Cross-validation: An alternative bandwidth-selection method in Kernel equating. *Applied Psychological Measurement*, 38(4), 281-295.

- Liou, M., Cheng, P. E., & Johnson, E. G. (1997). Standard errors of the Kernel equating methods under the common-item design. *Applied Psychological Measurement, 21*(4), 349-369.
- Liu, J., Sinharay, S., Holland, P. W., Feigenbaum, M., & Curley, E. (2009). *The Effects of Different Types of Anchor Tests on Observed Score Equating* (ETS Research Report 09-41). Princeton, NJ: Educational Testing Service.
- Liu, J., Sinharay, S., Holland, P. W., Curley, E., & Feigenbaum, M. (2011a). Test score equating using a mini-version anchor and a midi anchor: A case study using SAT R data. *Journal of Educational Measurement, 48*, 361–379.
- Liu, J., Sinharay, S., Holland, P., Feigenbaum, M., & Curley, E. (2011b). Observed score equating using a mini-version anchor and an anchor with less spread of difficulty: A comparison study. *Educational and Psychological Measurement, 71*, 346-361.
- Liu, J., & Low, A. (2007). *An exploration of Kernel equating using SAT data: Equating to a similar population and to a distant population* (ETS RR-07-17). Princeton, NJ: Educational Testing Service.
- Livingston, S. A. (1993). Small-Sample Equating With Log-Linear Smoothing. *Journal of Educational Measurement, 30*(1), 23-39.
- Livingston, S. A. (2004). *Equating test scores (without IRT)*. Princeton, NJ: Educational Testing Service.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Mao, X. (2006). *An investigation of the accuracy of the estimates of standard errors for the Kernel equating functions*. (Unpublished doctoral thesis). University of Iowa, Iowa City.
- Mao, X., von Davier, A. A., & Rupp, S. (2006). *Comparisons of the Kernel equating method with the traditional equating methods on PRAXISTM data* (ETS RR-06-30). Princeton, NJ: Educational Testing Service.
- Meng, Y. (2012). *Comparison of Kernel Equating and Item Response Theory Equating Methods*. (Unpublished Doctoral Dissertation). University of Massachusetts Amherst.
- Moses, T., & Holland, P. (2007). *Kernel and traditional equipercentile equating with degrees of presmoothing* (ETS Research Report 07-15). Princeton, NJ: Educational Testing Service.
- Moses, T., & Holland, P. (2007). *Kernel and traditional equipercentile equating with degrees of presmoothing* (ETS RR-07-15). Princeton, NJ: Educational Testing Service.
- Mutluer, C. (2013). *Yıl içinde farklı dönemlerde yapılan Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) puanlarına ilişkin bir test eşitleme çalışması*. (Yayımlanmamış Yüksek Lisans Tezi). Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Norman Dvorak, R. K. (2009). *A comparison of Kernel equating to the test characteristic curve methods*. (Unpublished doctoral dissertation). University of Nebraska, Lincoln.

- Nozawa, Y. (2008). *Comparison of parametric and nonparametric IRT equating methods under the common-item nonequivalent groups design*. (Unpublished doctoral dissertation). University of Iowa. Iowa city.
- Öztürk, N., (2010). *Akademik personel ve lisansüstü eğitimi giriş sınavı puanlarının eşitlenmesi üzerine bir çalışma*. (Yayınlanmamış yüksek lisans tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü. Ankara.
- Partchev, I. (2016). *Package 'irtoys'*. (Version 0.2.0) [Çevrim-içi: <https://cran.r-project.org/web/packages/irtoys/irtoys.pdf>, Erişim tarihi: Ekim 2016.]
- Petersen, N. S., Cook, L. L., & Stocking, M. L. (1983). IRT versus conventional equating methods: A comparative study of scale stability. *Journal of Educational Statistics*, 8,137-156.
- Petersen, N.S.,Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming and equating. In R.L. Linn (Ed.), *Educational Measurement* (3rd ed.) (pp. 221-262). New York: American Council on Education.
- Potenza, M. T.,& Dorans, N. J. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. *Applied Psychological Measurement*, 19(1), 23-37.
- Powers, S. J., Hagne, S. L., Wang, W., He, Y., Liu, C., & Kolen, M. J. (2011). Effects of group differences on mixed-format equating. In M. J. Kolen, & W. Lee (Ed.), *Mixed format tests: Psychometric properties with a primary focus on equating (Volume 1)*(CASMA Monograph No. 2.1) (pp. 51-73). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.
- Powers, S. J.,& Kolen, M. J. (2011). Evaluating equating accuracy and assumptions for groups that differ in performance. In M. J. Kolen, & W. Lee (Ed.), *Mixed-format tests: Psychometric properties with a primary focus on equating (Volume 1)*(CASMA Monograph No. 2.1) (pp. 137-176). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.
- Ricker, K.,& von Davier, A. A. (2007). *The impact of anchor test length on equating results in a nonequivalent groups design* (ETS RR-07-44). Princeton, NJ: Educational Testing Service.
- Rizopoulos, D. (2015). *Package 'ltm'*. [<https://cran.r-project.org/web/packages/ltm/ltm.pdf>, Erişim tarihi: Ekim 2016.]
- Qu, Y. (2007). *The effect of weighting in Kernel equating using counter-balanced designs*. (Unpublished doctoral thesis). Michigan State University: East Lansing.
- Sarkar, D. (2017). *Package "lattice"*. [<https://cran.r-project.org/web/packages/lattice/lattice.pdf>, Erişim Tarihi: Mayıs, 2016.]
- Seungho Yang, M. A. (2007). *A comparison of unidimensional and multidimensional rasch models using parameter estimates and fit indices when assumption of unidimensionality is violated*. (Unpublished Doctoral Dissertation). The Ohio State University.

- Sinharay, S., & Holland, P. W. (2006a). *The correlation between the scores of a test and an anchor test* (ETS SR-06-04). Princeton, NJ: Educational Testing Service.
- Sinharay, S., & Holland, P. W. (2006b). *Choice of anchor test in equating* (ETS RR-06-35). Princeton, NJ: Educational Testing Service.
- Sinharay, S., & Holland, P. W. (2007). Is it necessary to make anchor tests mini-forms of the tests being equated or can some restrictions be relaxed? *Journal of Educational Measurement*, 44, 249–275.
- Sinharay, S., Haberman, S., Holland, P. W., & Lewis, C. (2012). *A note on the choice of an anchor test in equating* (ETS Research Report 12-14). Princeton, NJ: Educational Testing Service.
- Skaggs, G., & Lissitz, R. W. (1986). IRT test equating: Relevant issues and a review of recent research. *Review of Educational Research*, 56, 495-529.
- Şahhüseyinoğlu, D. (2005). *İngilizce yeterlik sınavı puanlarının üç farklı eşitleme yöntemine göre karşılaştırılması*. (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi. Sosyal Bilimler Enstitüsü. Ankara.
- Tsai, T. H. (1997). *Estimating minimum sample sizes in random groups equating*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.
- Tong, Y., & Kolen, M. (2005). Assessing equating results on different equating criteria. *Applied Psychological Measurement*, 29 (6), 418-432.
- Uçar, A (2015). Artı değer özelliğine sahip alt testlerde eşitleme yöntemlerinin karşılaştırılması. (Yayınlanmamış yüksek lisans tezi). Mersin Üniversitesi, Eğitim Bilimleri Enstitüsü. Mersin.
- Uysal, İ. (2014). *Madde Tepki Kuramı'na dayalı test eşitleme yöntemlerinin karma modeller üzerinde karşılaştırılması*. (Yayınlanmamış Yüksek Lisans Tezi). Abant İzzet Baysal Üniversitesi, Eğitim Bilimleri Enstitüsü. Bolu.
- van der Linden, W. J. (2006). Book review of A. A. von Davier, P. W. Holland & D. T. Thayer. The kernel method of test equating. New York: Springer. *Journal of Educational Measurement*, 43, 291-294.
- von Davier, A., Holland, P. W., & Thayer, D. T. (2004). *The Kernel method of equating*. New York, NY: Springer.
- von Davier, A. A., Holland, P. W., Livingston, S. A., Casabianca, J., Grant, M. C., & Martin, K. (2006). *An evaluation of the Kernel equating method. A special study with pseudotests constructed from real test data* (Research Report No. RR-06-02). Princeton, NJ: Educational Testing Service.
- von Davier, A.A. (2008). New results on the linear equating methods for the non-equivalent-groups design. *Journal of Educational and Behavioral Statistics*, 33(2), 186-203.
- von Davier, A. A. (2013). Observed-score equating: An overview. *Psychometrika*, 78(4), 605-623.



- Wang, T., Lee, W., Brennan, R. & Kolen, M. (2008). A comparison of the frequency estimation and chained equipercentile methods under the common-item nonequivalent groups design. *Applied Psychological Measurement*, 32, 632–651.
- Weeks, J. P. (2010). Plink: An R package for linking mixed-format tests using IRT-based methods. *Journal of Statistical Software*, 35(12), 1-33.
- Wei, H. (2010, May). *Impact of non-representative anchor items on scale drift*. Paper presented at the annual conference of the National Council on Measurement in Education, Denver, CO.
- Wingersky, M. S., Cook, L. L., & Eignor, D. R. (1987). *Specifying the characteristics of linking items used for item response theory item calibration*. ETS Research Report Series, 1987(1).
- Zhao, Y. (2008). *Approaches for addressing the fit of item response theory models to educational test data*. (Unpublished doctorate thesis). University of Massachusetts.
- Zhu, W. (1998). Test equating: What, why, how?. *Research quarterly for exercise and sport*, 69(1), 11-23.

## EKLER DİZİNİ

# EK 1. ETİK KOMİSYONU ONAY BİLDİRİMİ

Form: 40

## Tez Çalışması Etik Komisyon İzin Muafiyeti Formu

04 / 08 / 2017

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

**Tez Başlığı / Konusu:** Kernel Eşitleme ve Madde Tepki Kuramı Eşitleme Yöntemlerinin Karşılaştırılması

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.


  
Çiğdem AKIN ARIKAN  
(Öğrencinin Adı Soyadı, İmzası)

### Öğrenci Bilgileri

Adı Soyadı	Çiğdem AKIN ARIKAN
Öğrenci No	N11247097
Anabilim Dalı	Eğitim Bilimleri
Programı	Eğitimde Ölçme ve Değerlendirme
Statüsü	<input type="checkbox"/> Yüksek Lisans <input checked="" type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr.

### Danışman Görüşü ve Onayı

Kernel Eşitleme ve Madde Tepki Kuramı Eşitleme Yöntemlerinin Karşılaştırılması" başlıklı tez çalışmasında bilgisayar ortamında simülasyon yöntemi ile veri üretilmiş olup, herhangi bir birey ya da denek üzerinden veri toplanmamıştır.

  
Prof. Dr. Selahattin GELBAL  
(İmza)  
(Danışmanın unvanı, Adı ve Soyadı)

## EK 2. ORJİNALLİK RAPORU



HACETTEPE ÜNİVERSİTESİ  
EĞİTİM BİLİMLERİ ENSTİTÜSÜ  
DOKTORA TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ  
EĞİTİM BİLİMLER ENSTİTÜSÜ  
EĞİTİM BİLİMLERİ ANA BİLİM DALI BAŞKANLIĞI'NA

Tarih: 04/08/2017

Tez Başlığı: Kernel Eşitleme ve Madde Tepki Kuramına Dayalı Eşitleme Yöntemlerinin Karşılaştırılması

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

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Endeksi	Gönderim Numarası
04/08/2017	104	160041	08/06/2016	%3	809517277

Uygulanan filtreler:

- 1- Kaynakça hariç
- 2- Alıntılar dâhil
- 3- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

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

Gereğini saygılarımla arz ederim.

04/08/2017

Adı Soyadı: Çiğdem AKIN ARIKAN

Öğrenci No: N11247097

Anabilim Dalı: Eğitim Bilimleri

Programı: Eğitimde Ölçme ve Değerlendirme

Statüsü:  Y.Lisans  Doktora  Bütünleşik Dr.

### DANIŞMAN ONAYI

Prof. Dr. Selahattin GELBAL



HACETTEPE UNIVERSITY  
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES  
THESIS/DISSERTATION ORIGINALITY REPORT

HACETTEPE UNIVERSITY  
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES  
TO THE DEPARTMENT OF EDUCATIONAL SCIENCES

Date: 08/04/2017

Thesis Title : Comparison of Kernel Equating and Item Response Theory Equating Methods

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

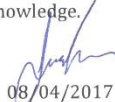
Time Submitted	Page Count	Character Count	Date of Thesis Defence	Similarity Index	Submission ID
08/04 /2017	104	160041	06/08/2017	3%	809517277

Filtering options applied:

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

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

I respectfully submit this for approval.

  
08/04/2017

**Name Surname:** Çiğdem AKIN ARIKAN

**Student No:** N11247097

**Department:** Educational Sciences

**Program:** Educational Measurement and Evaluation

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

**ADVISOR APPROVAL**



Prof. Dr. Selahattin GELBAL

### EK 3. VERİ SETLERİNİN ÜRETİLMESİNDE KULLANILAN R KODU

# %20 ortak madde oranı için mini ortak test kullanıldığında 500 örneklem büyüklüğü için X ve Y formlarının elde edilmesine ilişkin kod#####

```
library(irtoys)
library(xlsx)

replication <- 1
N <- 500
X <- 40
na <- 10

x.a <- runif(X, 0.5, 2)
x.b <- rnorm(X,0,1)
x.c <- runif(X,0.2,0.3)
y.a <- runif(X, 0.5, 2)
y.b <- rnorm(X,0,1)
y.c <- runif(X,0.2,0.3)
a.a <- runif(na, 0.5, 2)
a.b <- rnorm(na,0,1)
a.c <- runif(na,0.2,0.3)

x.person <- vector("list",replication)
x.item <- vector("list",replication)
x.data <- vector("list",replication)
for(i in 1:replication) {
  x.person[[i]] <- rnorm(N,0,1)
  x.item[[i]] <- cbind(x.a,x.b,x.c)
}
for(i in 1:replication) {
  x.data[[i]] <- sim(ip=x.item[[i]], x=x.person[[i]])
}

a.item <- vector("list",replication)
a.x.data <- vector("list",replication)
for(i in 1:replication) {
  a.item[[i]] <- cbind(a.a,a.b,a.c)
}
library(irtoys)
for(i in 1:replication) {
  a.x.data[[i]] <- sim(ip=a.item[[i]], x=x.person[[i]])
}
total.x <- vector("list",replication)

for(i in 1:replication) {
  total.x[[i]] <- cbind(x.data[[i]], a.x.data[[i]])
}
```

```

y.person <- vector("list",replication)
y.item <- vector("list",replication)
y.data <- vector("list",replication)
for(i in 1:replication) {
y.person[[i]] <- rnorm(N,0.05,1)
y.item[[i]] <- cbind(y.a,y.b,y.c)
}
library(irtoys)
for(i in 1:replication) {
y.data[[i]] <- sim(ip=y.item[[i]], x=y.person[[i]])
}
a.y.data <- vector("list",replication)
library(irtoys)
for(i in 1:replication) {
a.y.data[[i]] <- sim(ip=a.item[[i]], x=y.person[[i]])
}
total.y<- vector("list",replication)
for(i in 1:replication) {
total.y[[i]] <- cbind(y.data[[i]], a.y.data[[i]])
}

for(i in 1:replication) {

write.xlsx(total.x[[i]] , "C:/Users/ıgdem/Desktop/xdata.xlsx", sheetName="1")

write.xlsx(total.y[[i]], "C:/Users/ıgdem/Desktop/ydata.xlsx", sheetName="1")

}

```

#### EK 4. ORTAK MADDELERE İLİŞKİN BETİMSEL İSTATİSTİKLER

Ortak Maddelere İlişkin Betimsel İstatistikler				
Ortak Madde Güçlük Dağılımı	Ortak Test Tipi	Ortak Madde Oranı (%)	Ortalama	Standart Sapma
Mini	İç	20	0.10	0.89
		30	0.0	0.91
		40	-0.11	0.92
	dış	20	0.01	0.90
		30	0.11	0.89
		40	0.01	0.91
Midi	İç	20	0.09	0.17
		30	0.02	0.19
		40	-0.12	0.21
	dış	20	0.00	0.18
		30	0.10	0.19
		40	0.00	0.19



**EK 5. 500 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN RMSD DEĞERLERİ**

	Dış Ortak Test			İç Ortak Test		
	%20	%30	%40	%20	%30	%40
<u>mini test/yb</u>						
Z1	2.26	2.19	2.16	2.32	2.30	2.26
ST1	2.24	2.18	2.14	2.31	2.28	2.24
Z2	2.42	2.37	2.34	2.46	2.44	2.41
ST2	2.41	2.36	2.33	2.46	2.43	2.39
MTK	2.24	2.18	2.15	2.32	2.30	2.24
<u>midi test/yb</u>						
Z1	2.25	2.17	2.14	2.31	2.29	2.25
ST1	2.22	2.16	2.12	2.31	2.29	2.24
Z2	2.41	2.38	2.35	2.45	2.43	2.41
ST2	2.39	2.36	2.33	2.45	2.41	2.40
MTK	2.22	2.17	2.13	2.32	2.30	2.23
<u>mini test/yf</u>						
Z1	2.32	2.30	2.27	2.34	2.32	2.28
ST1	2.34	2.32	2.28	2.34	2.33	2.29
Z2	2.46	2.43	2.39	2.52	2.47	2.42
ST2	2.48	2.44	2.40	2.54	2.48	2.43
MTK	2.25	2.24	2.18	2.42	2.40	2.32
<u>midi test /yf</u>						
Z1	2.31	2.28	2.25	2.33	2.31	2.27
ST1	2.33	2.30	2.26	2.35	2.33	2.29
Z2	2.44	2.41	2.38	2.50	2.44	2.39
ST2	2.46	2.42	2.39	2.52	2.45	2.41
MTK	2.24	2.24	2.15	2.41	2.39	2.31

Z1: Zincirleme Eşit Yüzdelikli Eşitleme

ST1: Son Tabakalama Eşit Yüzdelikli Eşitleme

Z2: Zincirleme Doğrusal Eşitleme

ST2: Son Tabakalama Doğrusal Eşitleme

MTK: Madde Tepki Kuramına Dayalı Eşitleme

yb: Gruplar arası yetenek dağılımının benzer olduğu durum

yf: Gruplar arası yetenek dağılımının farklı olduğu durum

**EK 6. 1500 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN RMSD DEĞERLERİ**

mini test/yb	Dış Ortak Test			İç Ortak Test		
	20%	30%	40%	20%	30%	40%
Z1	2.14	2.12	2.10	2.18	2.15	2.13
ST1	2.13	2.11	2.09	2.17	2.11	2.09
Z2	2.28	2.26	2.23	2.31	2.30	2.27
ST2	2.27	2.23	2.20	2.31	2.29	2.25
MTK	2.14	2.11	2.09	2.18	2.17	2.16
<b>midi test/yb</b>						
Z1	2.14	2.11	2.09	2.18	2.16	2.13
ST1	2.13	2.11	2.09	2.18	2.13	2.09
Z2	2.27	2.25	2.22	2.32	2.29	2.26
ST2	2.26	2.22	2.20	2.31	2.28	2.23
MTK	2.13	2.11	2.09	2.18	2.17	2.16
<b>mini test /yf</b>						
Z1	2.17	2.12	2.10	2.31	2.28	2.26
ST1	2.17	2.13	2.11	2.31	2.27	2.25
Z2	2.40	2.35	2.33	2.39	2.36	2.32
ST2	2.43	2.40	2.35	2.42	2.40	2.35
MTK	2.17	2.12	2.10	2.28	2.22	2.18
<b>midi test /yf</b>						
Z1	2.17	2.12	2.10	2.28	2.25	2.22
ST1	2.17	2.13	2.11	2.29	2.26	2.24
Z2	2.31	2.28	2.26	2.37	2.35	2.30
ST2	2.33	2.30	2.26	2.40	2.36	2.32
MTK	2.16	2.12	2.10	2.25	2.21	2.17

Z1: Zincirleme Eşit Yüzdelikli Eşitleme

ST1: Son Tabakalama Eşit Yüzdelikli Eşitleme

Z2: Zincirleme Doğrusal Eşitleme

ST2: Son Tabakalama Doğrusal Eşitleme

MTK: Madde Tepki Kuramına Dayalı Eşitleme

yb: Gruplar arası yetenek dağılımının benzer olduğu durum

yf: Gruplar arası yetenek dağılımının farklı olduğu durum

**EK 7. 5000 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN RMSD DEĞERLERİ**

mini test/yb	Dış Ortak Test			İç Ortak Test		
	20%	30%	40%	20%	30%	40%
Z1	1.89	1.86	1.84	1.98	1.97	1.94
ST1	1.87	1.85	1.84	1.97	1.96	1.92
Z2	1.98	1.96	1.94	2.06	2.03	1.99
ST2	1.98	1.93	1.89	2.07	2.02	1.96
MTK	1.87	1.86	1.83	1.97	1.95	1.93
<b>midi test/yb</b>						
Z1	1.87	1.85	1.83	1.96	1.95	1.93
ST1	1.87	1.84	1.83	1.97	1.95	1.92
Z2	1.97	1.94	1.92	2.05	2.01	1.99
ST2	1.96	1.91	1.87	2.04	1.99	1.96
MTK	1.87	1.85	1.82	1.96	1.94	1.93
<b>mini test /yf</b>						
Z1	1.90	1.88	1.87	1.98	1.98	1.95
ST1	1.92	1.91	1.87	2.00	1.98	1.95
Z2	2.06	2.03	2.01	2.13	2.11	2.07
ST2	2.09	2.06	2.03	2.18	2.13	2.10
MTK	1.90	1.88	1.86	1.98	1.98	1.95
<b>midi test /yf</b>						
Z1	1.89	1.87	1.85	1.98	1.97	1.94
ST1	1.90	1.89	1.86	1.99	1.97	1.95
Z2	2.02	2.01	1.99	2.10	2.06	2.04
ST2	2.05	2.03	2.00	2.13	2.09	2.06
MTK	1.88	1.87	1.85	1.97	1.96	1.95

Z1: Zincirleme Eşit Yüzdelikli Eşitleme

ST1: Son Tabakalama Eşit Yüzdelikli Eşitleme

Z2: Zincirleme Doğrusal Eşitleme

ST2: Son Tabakalama Doğrusal Eşitleme

MTK: Madde Tepki Kuramına Dayalı Eşitleme

yb: Gruplar arası yetenek dağılımının benzer olduğu durum

yf: Gruplar arası yetenek dağılımının farklı olduğu durum

**EK 8. 500 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN SEE DEĞERLERİ**

mini test/yb	Dış Ortak Test			İç Ortak Test		
	20%	30%	40%	20%	30%	40%
Z1	0.723	0.706	0.682	0.769	0.748	0.721
ST1	0.665	0.653	0.632	0.732	0.698	0.661
Z2	0.670	0.627	0.603	0.684	0.661	0.646
ST2	0.645	0.602	0.581	0.683	0.654	0.632
MTK	0.762	0.718	0.686	0.806	0.783	0.762
<b>midi test/yb</b>						
Z1	0.671	0.652	0.629	0.785	0.756	0.725
ST1	0.649	0.634	0.608	0.737	0.715	0.705
Z2	0.622	0.594	0.584	0.685	0.668	0.651
ST2	0.618	0.586	0.561	0.689	0.659	0.643
MTK	0.710	0.671	0.635	0.803	0.776	0.746
<b>mini test/yf</b>						
Z1	0.942	0.905	0.891	0.980	0.964	0.912
ST1	0.945	0.920	0.905	0.995	0.976	0.963
Z2	0.929	0.889	0.846	0.957	0.896	0.871
ST2	0.931	0.895	0.852	0.974	0.937	0.880
MTK	0.963	0.925	0.910	0.998	0.990	0.975
<b>midi test/yf</b>						
Z1	0.874	0.841	0.809	0.942	0.879	0.872
ST1	0.901	0.872	0.856	0.980	0.960	0.934
Z2	0.852	0.816	0.779	0.903	0.864	0.823
ST2	0.861	0.827	0.793	0.923	0.876	0.846
MTK	0.913	0.896	0.871	0.991	0.984	0.962

Z1: Zincirleme Eşit Yüzdelikli Eşitleme

ST1: Son Tabakalama Eşit Yüzdelikli Eşitleme

Z2: Zincirleme Doğrusal Eşitleme

ST2: Son Tabakalama Doğrusal Eşitleme

MTK: Madde Tepki Kuramına Dayalı Eşitleme

yb: Gruplar arası yetenek dağılımının benzer olduğu durum

yf: Gruplar arası yetenek dağılımının farklı olduğu durum

**EK 9. 1500 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN SEE DEĞERLERİ**

Mini/yb	Dış Ortak Test			İç Ortak Test		
	20%	30%	40%	20%	30%	40%
Z1	0.472	0.436	0.425	0.483	0.463	0.429
ST1	0.459	0.424	0.419	0.462	0.445	0.394
Z2	0.393	0.358	0.325	0.406	0.372	0.350
ST2	0.380	0.356	0.319	0.382	0.359	0.323
MTK	0.471	0.435	0.423	0.480	0.455	0.425
<b>Midi/yb</b>						
Z1	0.457	0.423	0.402	0.466	0.443	0.406
ST1	0.425	0.414	0.396	0.433	0.431	0.379
Z2	0.365	0.338	0.313	0.383	0.368	0.330
ST2	0.350	0.323	0.302	0.362	0.348	0.304
MTK	0.446	0.430	0.394	0.461	0.440	0.402
<b>Mini/yf</b>						
Z1	0.499	0.490	0.472	0.503	0.504	0.496
ST1	0.503	0.492	0.447	0.517	0.520	0.506
Z2	0.390	0.343	0.327	0.397	0.369	0.356
ST2	0.395	0.357	0.327	0.419	0.381	0.364
MTK	0.505	0.485	0.452	0.515	0.505	0.464
<b>Midi/yf</b>						
Z1	0.467	0.452	0.444	0.473	0.464	0.451
ST1	0.469	0.453	0.436	0.473	0.472	0.458
Z2	0.352	0.331	0.315	0.377	0.339	0.316
ST2	0.356	0.335	0.313	0.379	0.341	0.324
MTK	0.474	0.451	0.423	0.475	0.455	0.424

Z1: Zincirleme Eşit Yüzdelikli Eşitleme

ST1: Son Tabakalama Eşit Yüzdelikli Eşitleme

Z2: Zincirleme Doğrusal Eşitleme

ST2: Son Tabakalama Doğrusal Eşitleme

MTK: Madde Tepki Kuramına Dayalı Eşitleme

yb: Gruplar arası yetenek dağılımının benzer olduğu durum

yf: Gruplar arası yetenek dağılımının farklı olduğu durum

**EK 10. 5000 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN SEE DEĞERLERİ**

mini/yb	Dış Ortak Test			İç Ortak Test		
	20%	30%	40%	20%	30%	40%
Z1	0.240	0.237	0.222	0.246	0.239	0.222
ST1	0.211	0.210	0.195	0.212	0.212	0.196
Z2	0.180	0.174	0.151	0.182	0.177	0.155
ST2	0.169	0.168	0.136	0.170	0.169	0.141
MTK	0.225	0.202	0.195	0.228	0.204	0.199
<b>midi/yb</b>						
Z1	0.230	0.221	0.210	0.234	0.228	0.220
ST1	0.206	0.198	0.188	0.208	0.199	0.192
Z2	0.159	0.153	0.142	0.163	0.156	0.152
ST2	0.148	0.146	0.133	0.157	0.151	0.143
MTK	0.223	0.199	0.186	0.226	0.206	0.199
<b>mini/yf</b>						
Z1	0.246	0.233	0.229	0.247	0.240	0.235
ST1	0.216	0.214	0.212	0.220	0.217	0.213
Z2	0.194	0.187	0.147	0.198	0.194	0.168
ST2	0.172	0.170	0.137	0.173	0.185	0.159
MTK	0.281	0.231	0.222	0.288	0.245	0.229
<b>midi/yf</b>						
Z1	0.242	0.251	0.247	0.246	0.239	0.233
ST1	0.218	0.226	0.211	0.219	0.235	0.219
Z2	0.168	0.159	0.145	0.175	0.181	0.168
ST2	0.171	0.169	0.146	0.164	0.172	0.161
MTK	0.270	0.225	0.210	0.276	0.234	0.213

Z1: Zincirleme Eşit Yüzdelikli Eşitleme

ST1: Son Tabakalama Eşit Yüzdelikli Eşitleme

Z2: Zincirleme Doğrusal Eşitleme

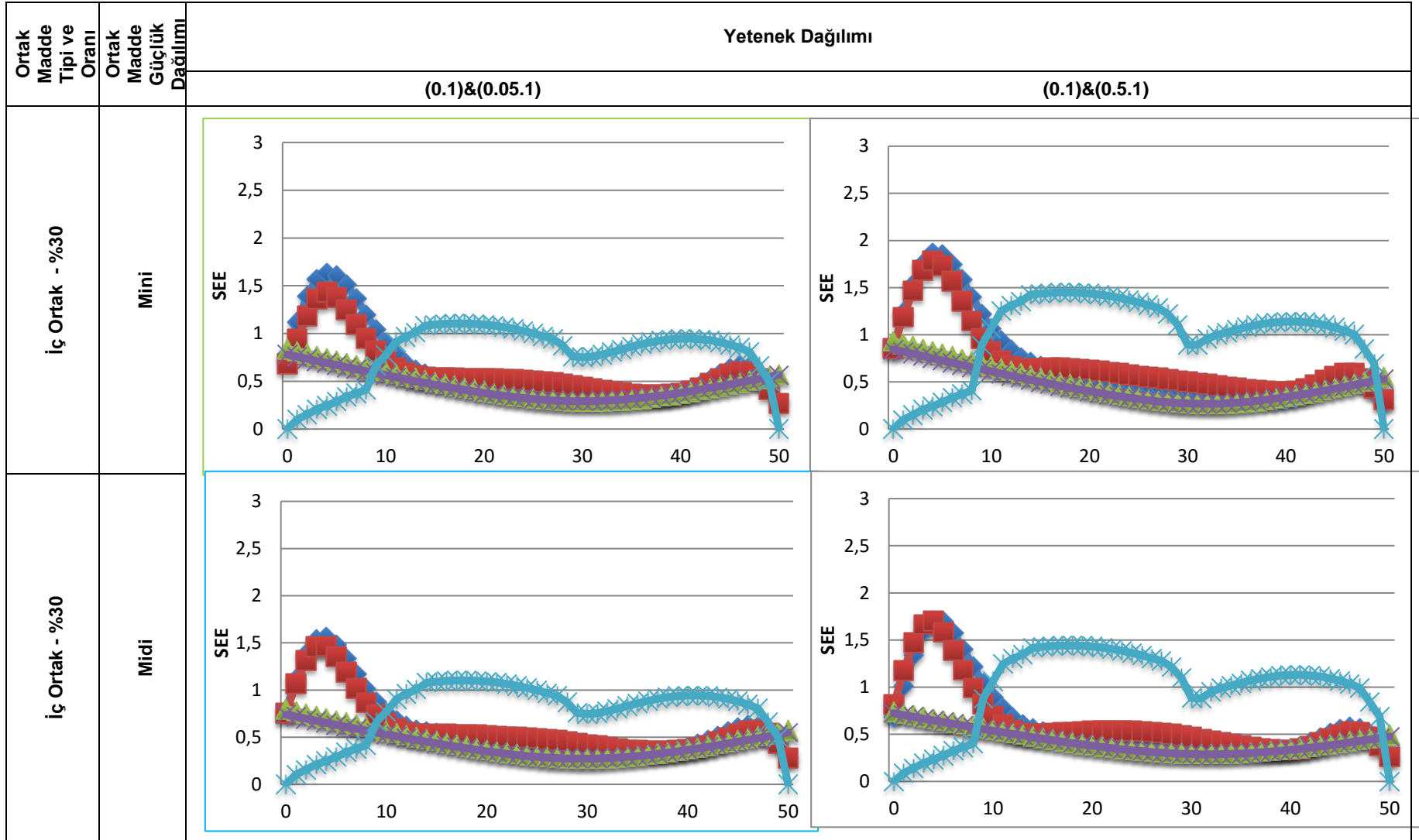
ST2: Son Tabakalama Doğrusal Eşitleme

MTK: Madde Tepki Kuramına Dayalı Eşitleme

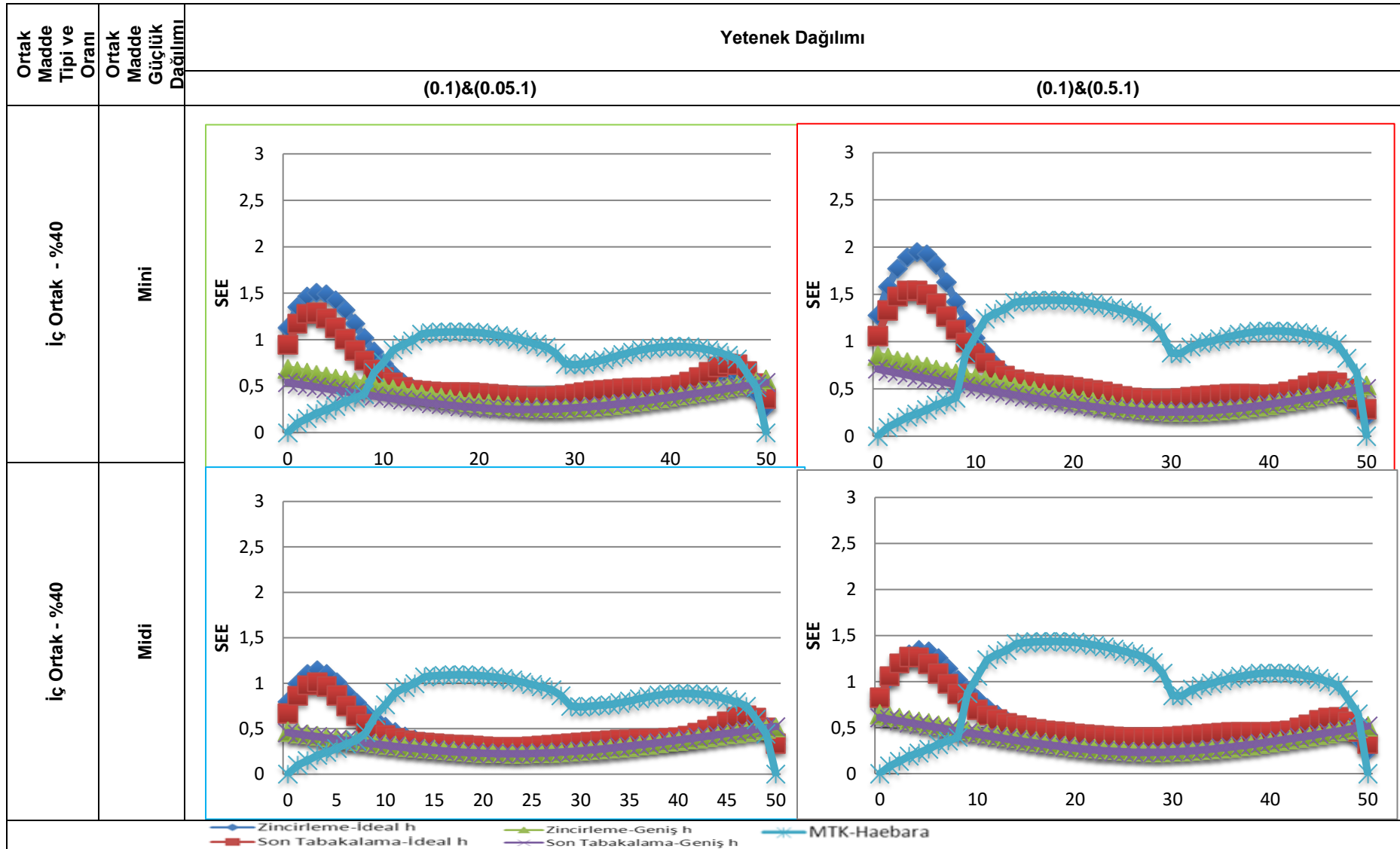
yb: Gruplar arası yetenek dağılımının benzer olduğu durum

yf: Gruplar arası yetenek dağılımının farklı olduğu durum

## EK 11. PUAN ÖLÇEĞİ BOYUNCA 500 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN SEE DEĞERLERİNE İLİŞKİN GRAFİKLER

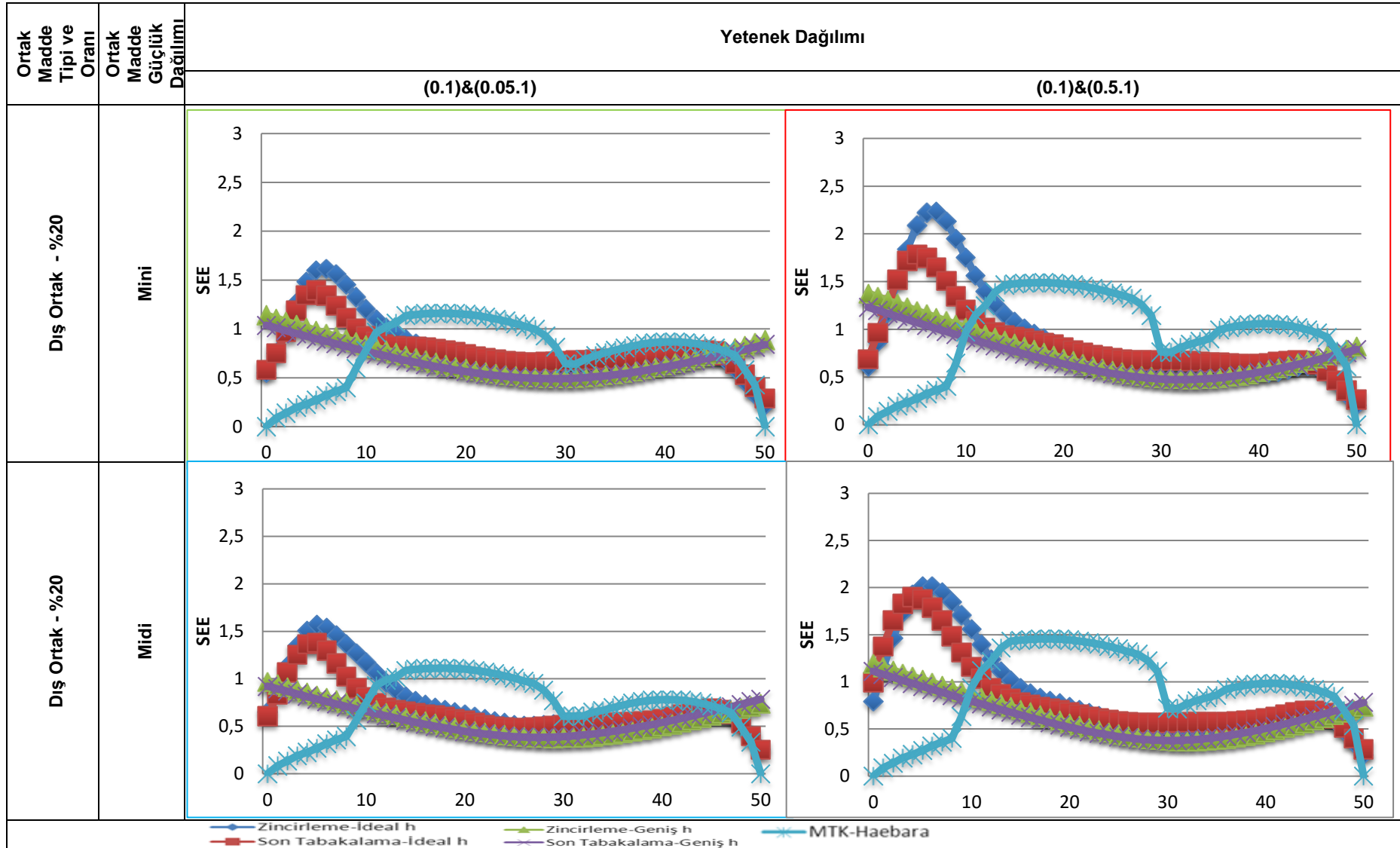


Şekil1: 500 Örneklem İçin Ortak Madde Oranının %30 olduğu Koşulda Elde Edilen SEE Değerleri

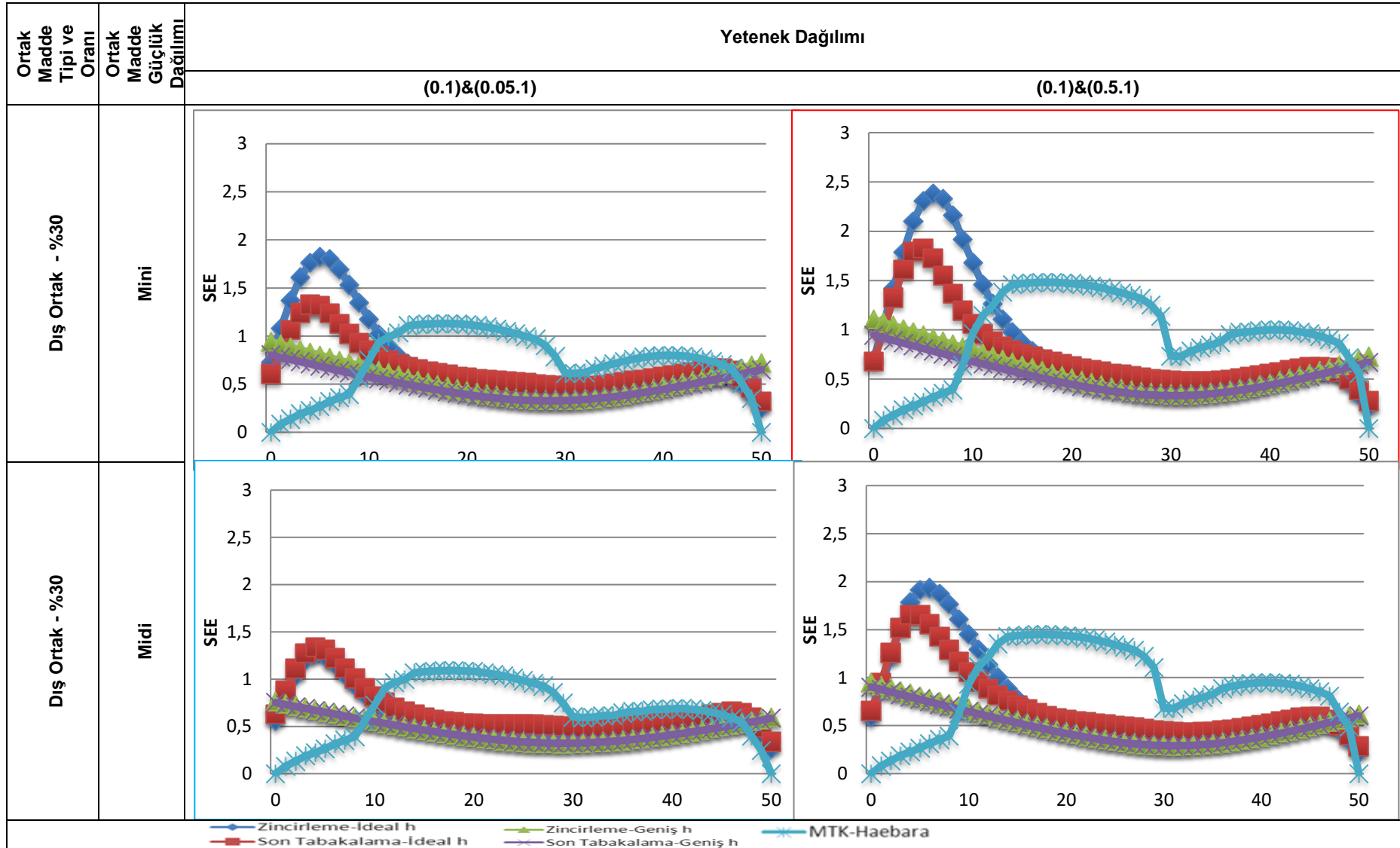


Şekil 2: 500 Örneklem için Ortak Madde Oranının %40 olduğu Koşulda Elde Edilen SEE Değerleri

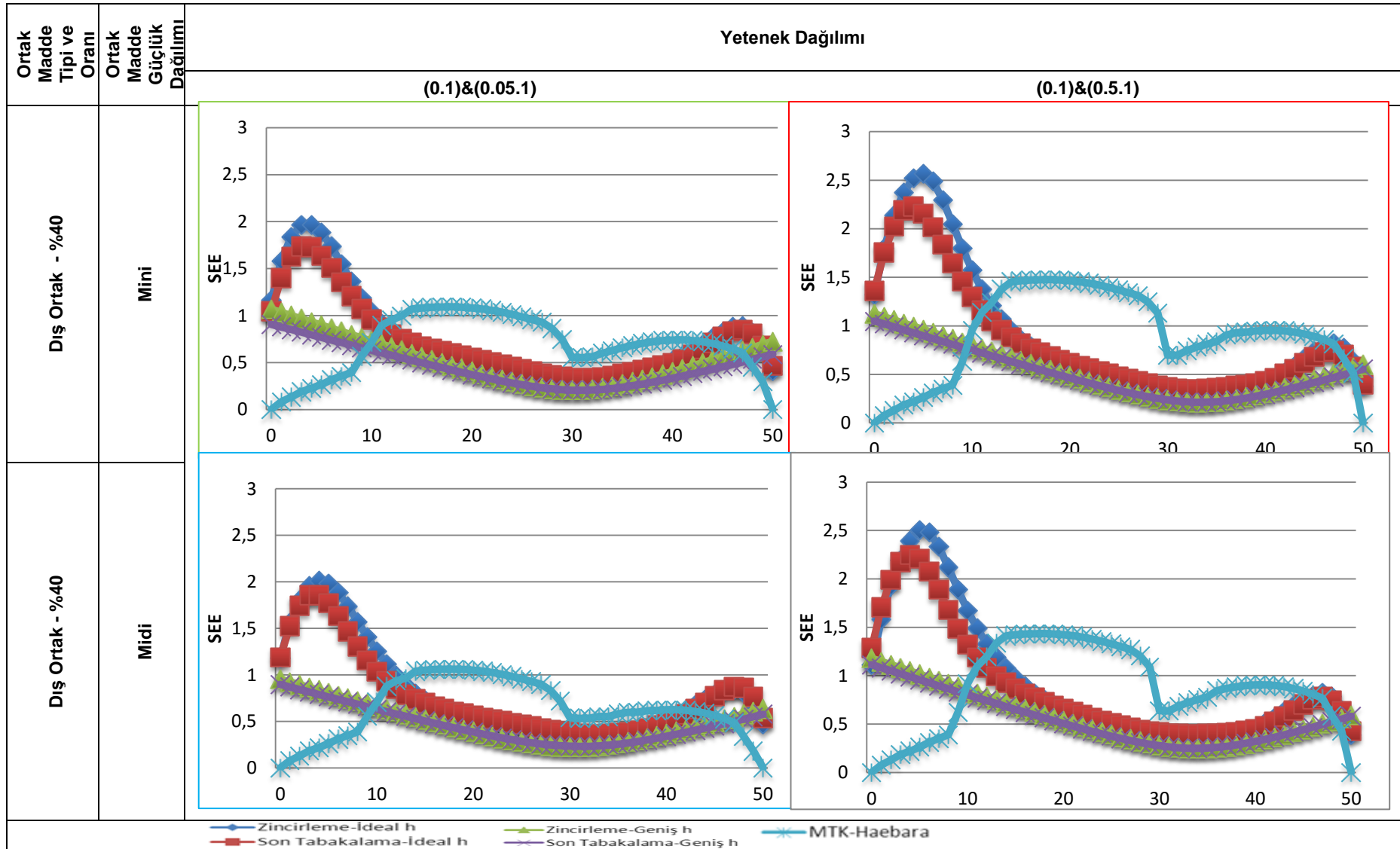




Şekil 3: 500 Örneklem için Ortak Madde Oranının %20 olduğu Koşulda Elde Edilen SEE Değerleri

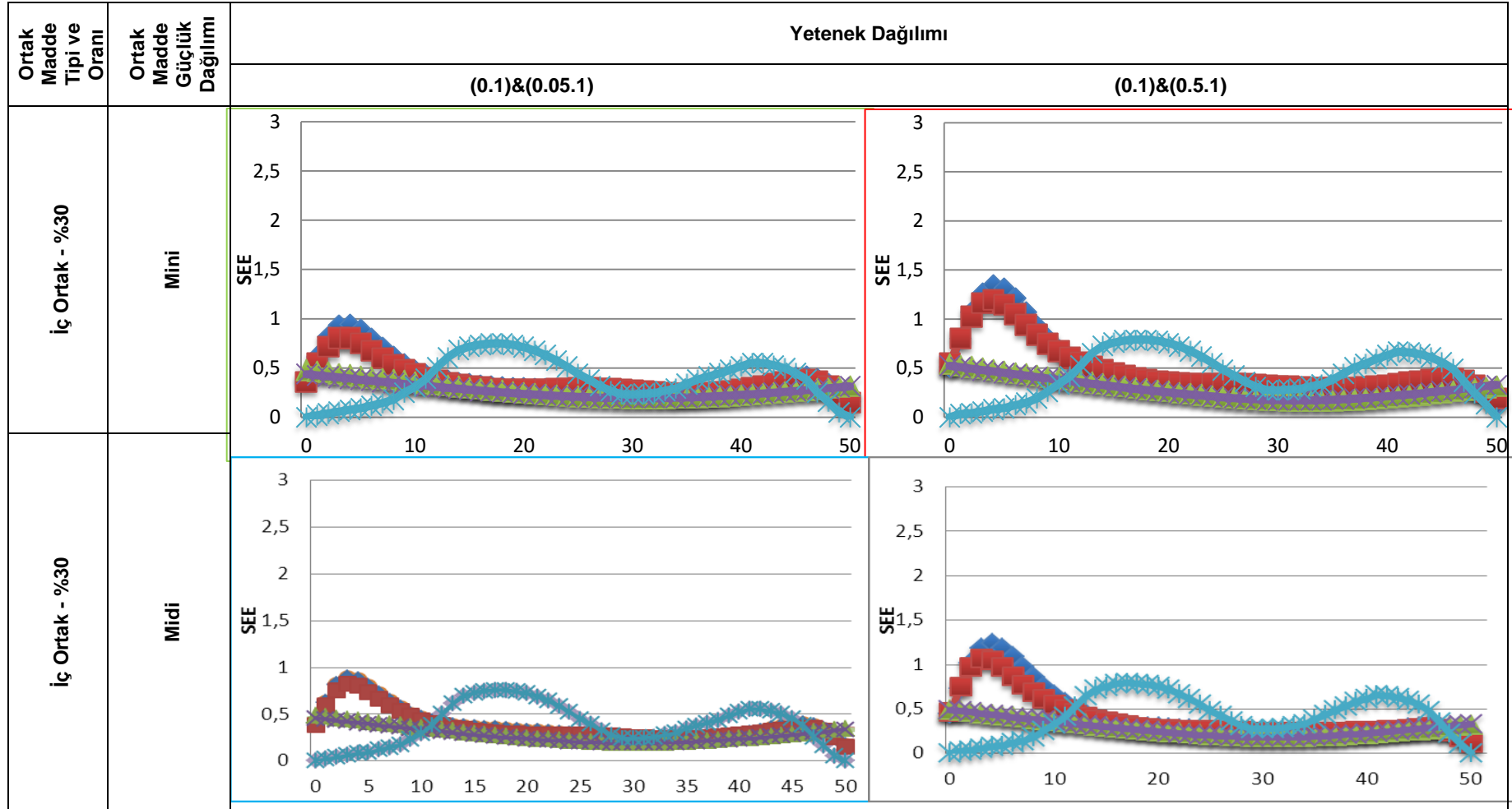


Şekil 4: 500 Örneklem İçin Ortak Madde Oranının %30 olduğu Koşulda Elde Edilen SEE Değerleri

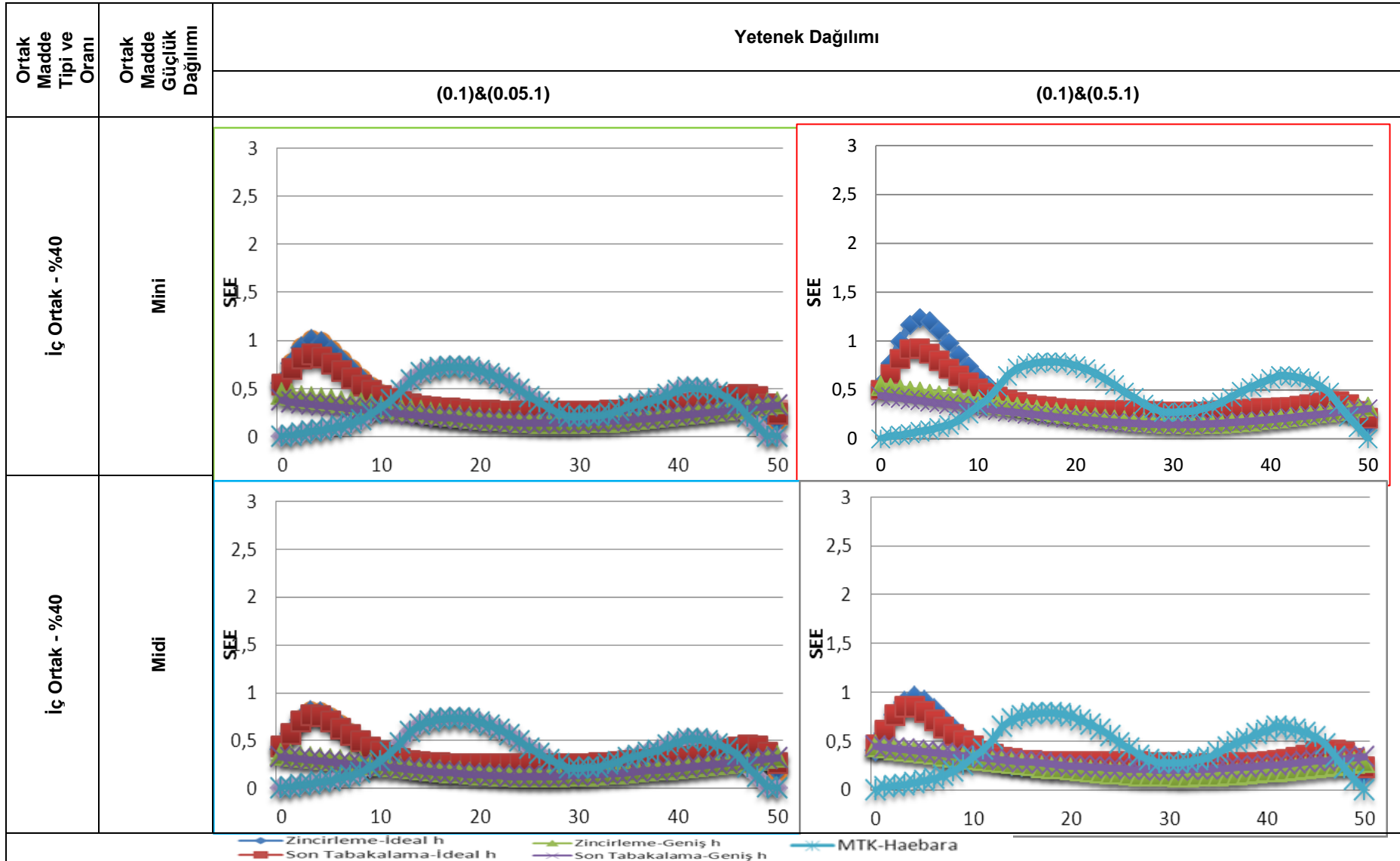


Şekil 5: 500 Örneklem İçin Ortak Madde Oranının %40 olduğu Koşulda Elde Edilen SEE Değerleri

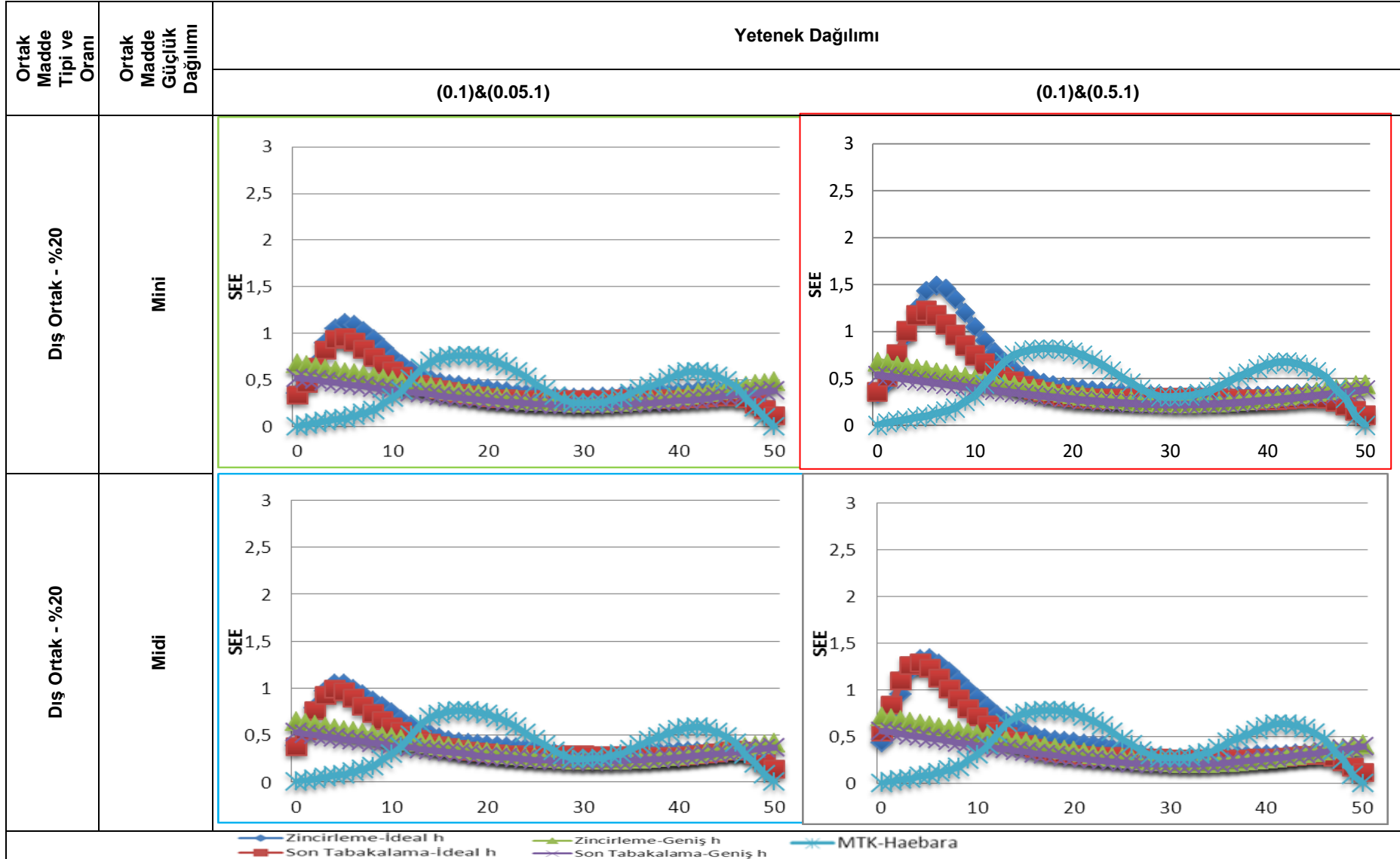
## EK 12. PUAN ÖLÇEĞİ BOYUNCA 1500 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN SEE DEĞERLERİNE İLİŞKİN GRAFİKLER



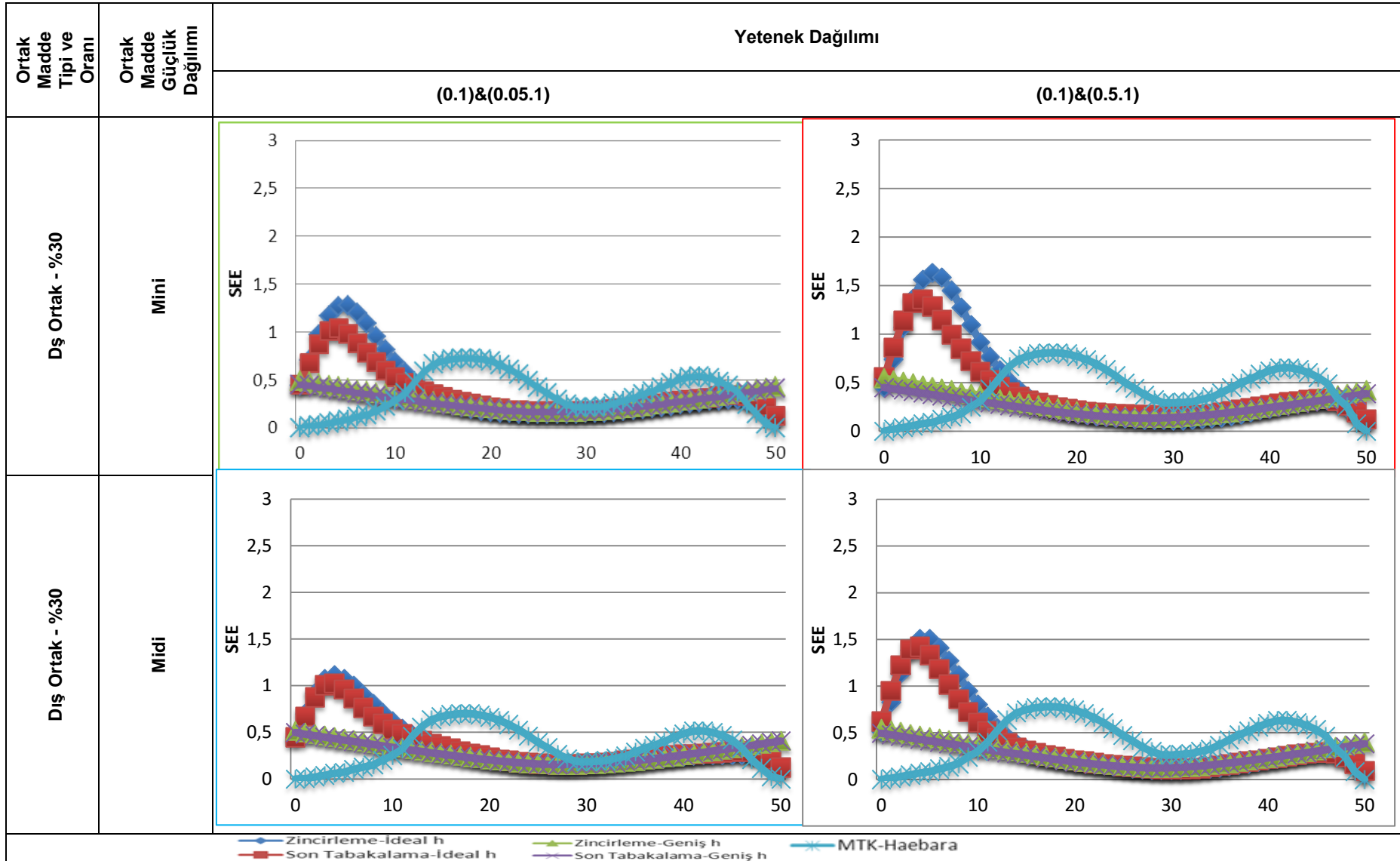
Şekil 1: 1500 Örneklem İçin Ortak Madde Oranının %30 olduğu Koşulda Elde Edilen SEE Değerleri



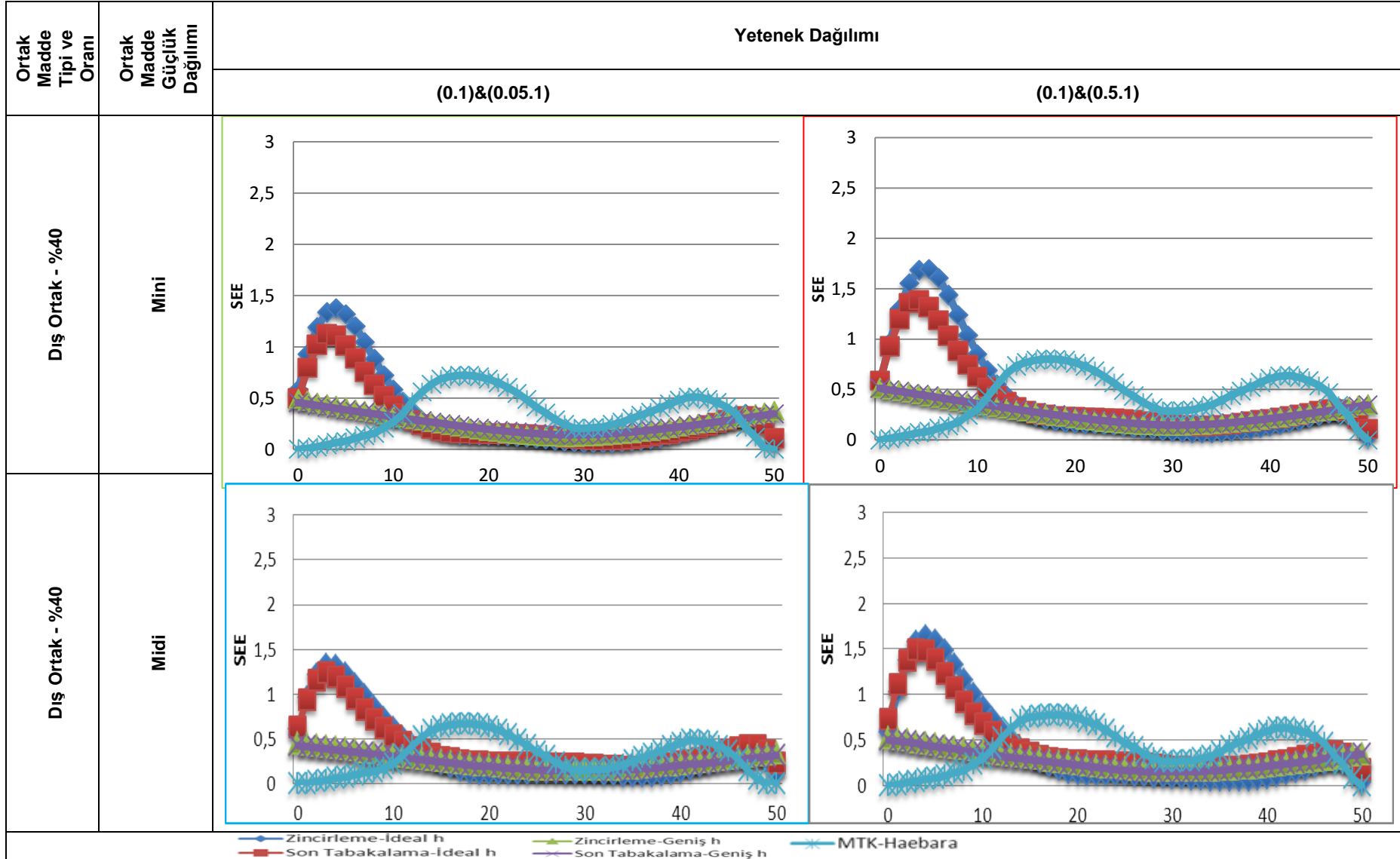
Şekil 2: 1500 Örneklem İçin Ortak Madde Oranının %40 olduğu Koşulda Elde Edilen SEE Değerleri



Şekil 3: 1500 Örneklem İçin Ortak Madde Oranının %20 olduğu Koşulda Elde Edilen SEE Değerleri



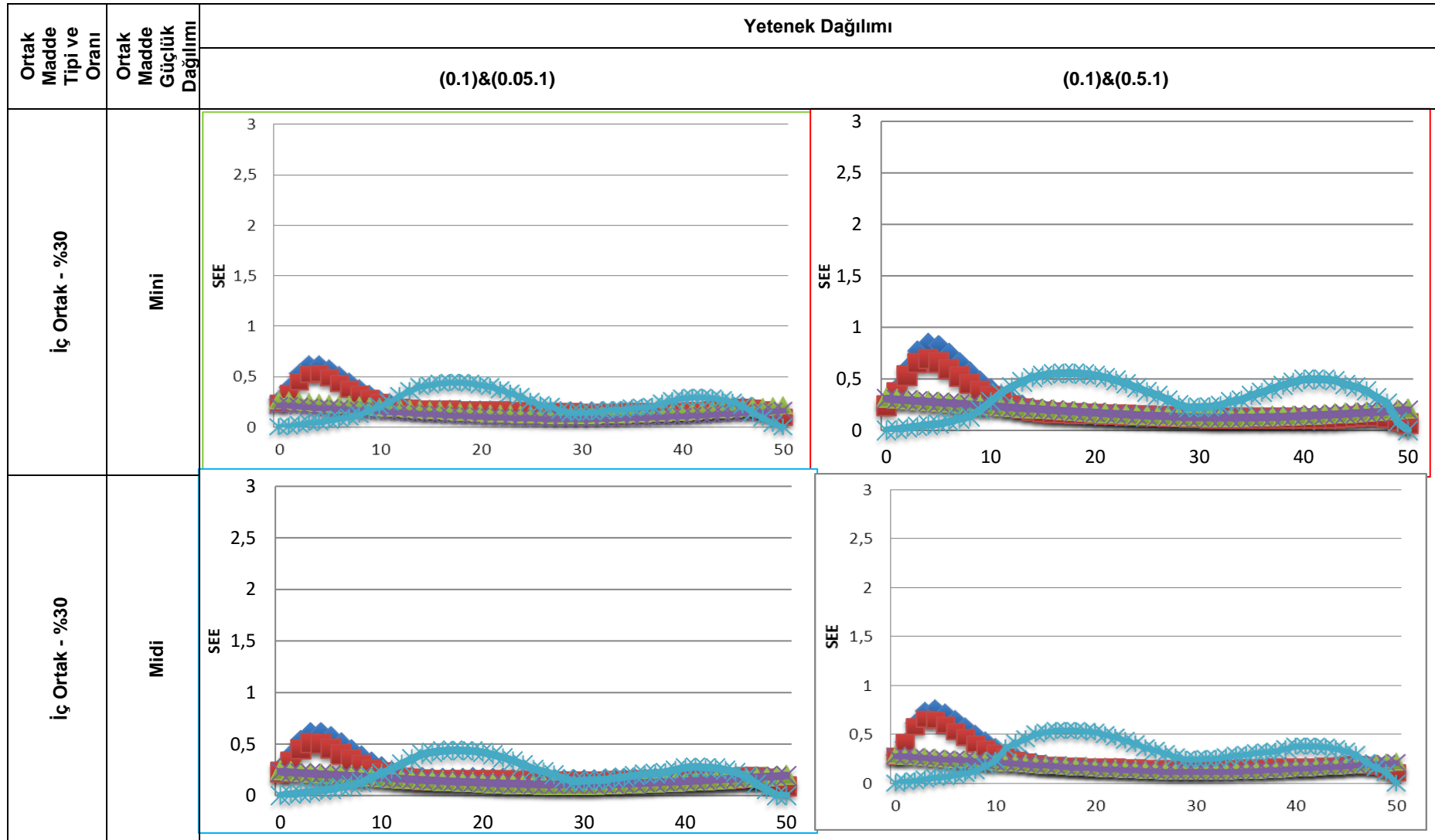
Şekil 4: 1500 Örneklem için Ortak Madde Oranının %30 olduğu Koşulda Elde Edilen SEE Değerleri



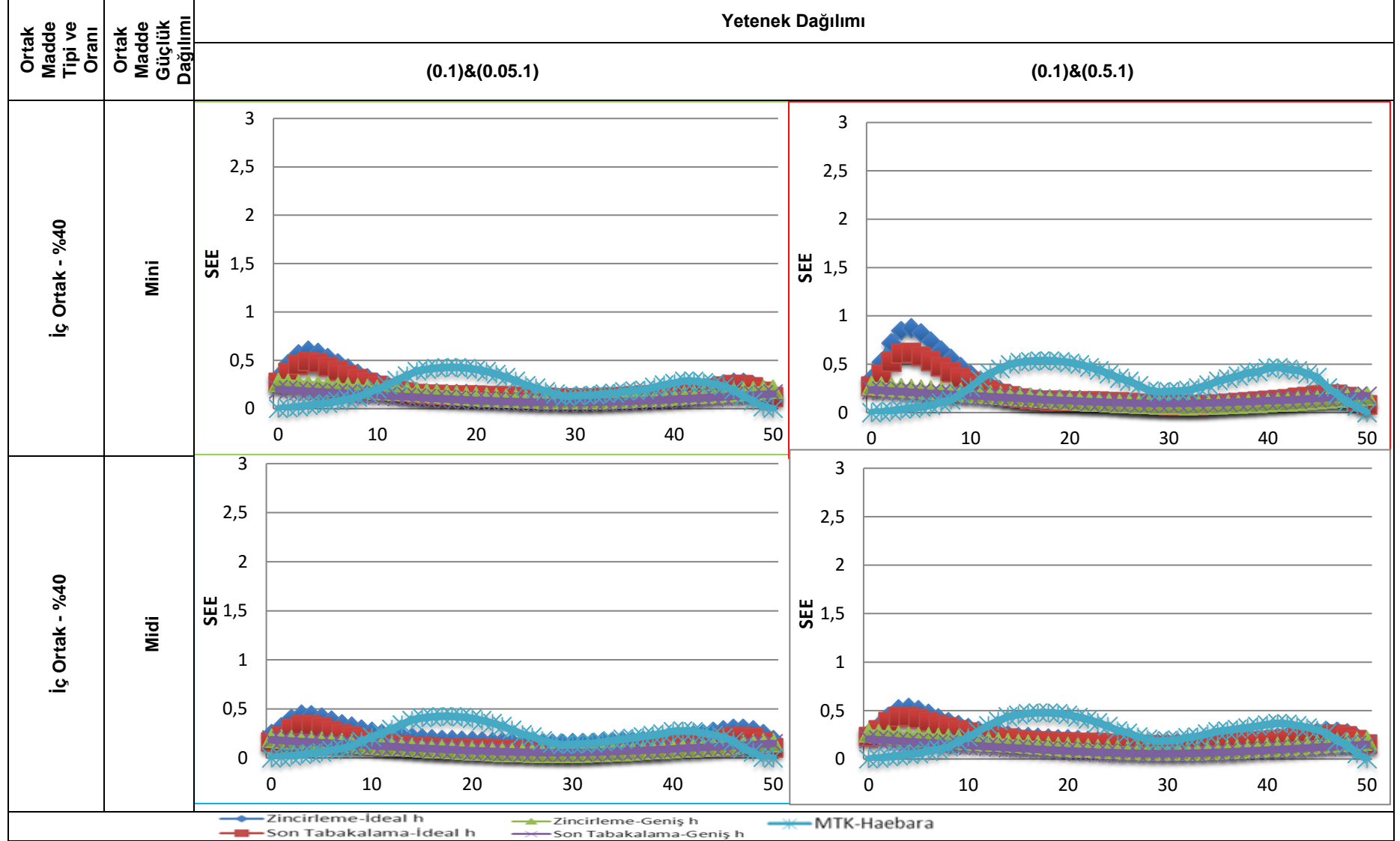
Şekil 5: 1500 Örneklem İçin Ortak Madde Oranının %40 olduğu Koşulda Elde Edilen SEE Değerleri



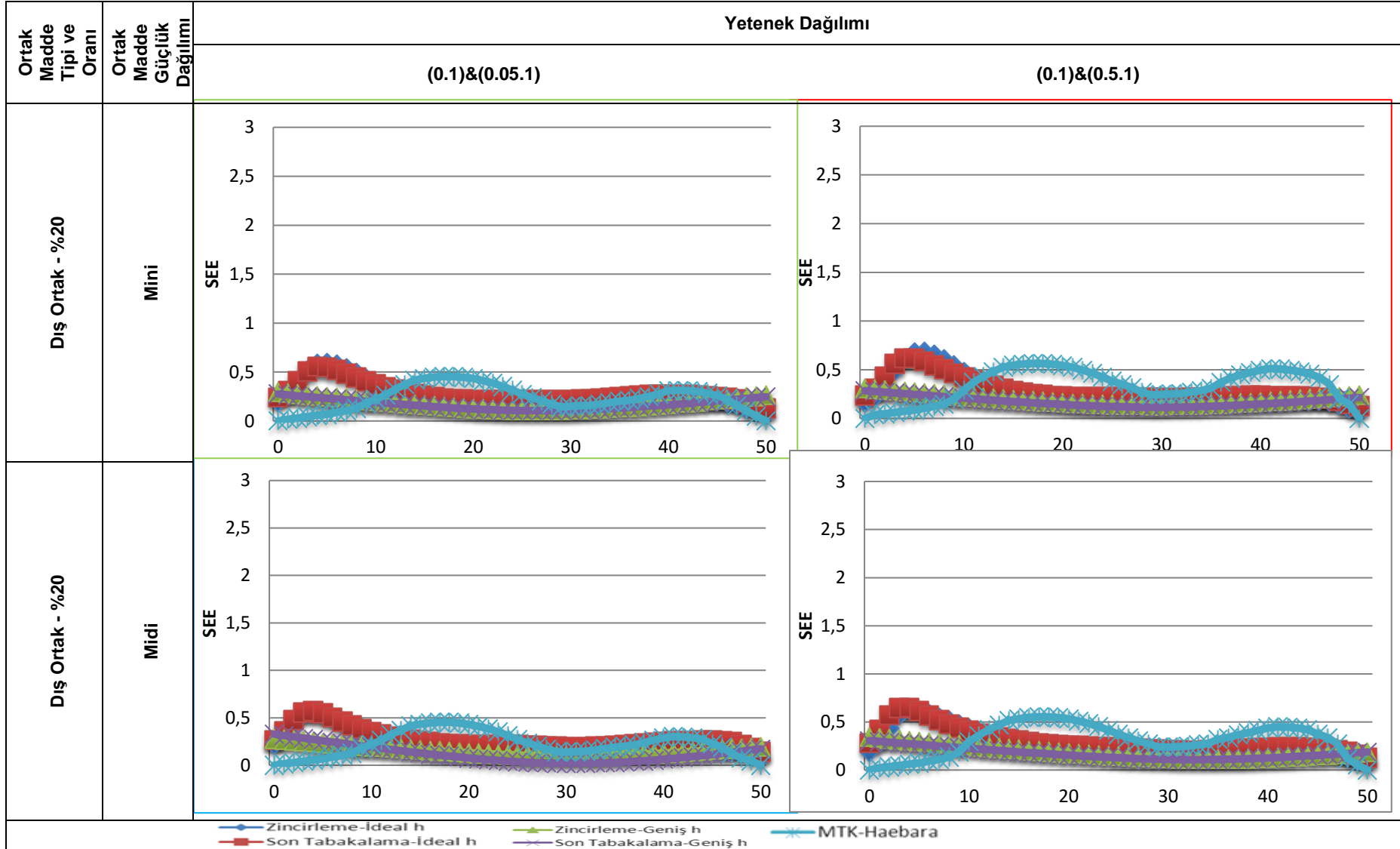
## EK 13. PUAN ÖLÇEĞİ BOYUNCA 5000 ÖRNEKLEM İÇİN KOŞULLARA GÖRE ELDE EDİLEN SEE DEĞERLERİNE İLİŞKİN GRAFİKLER



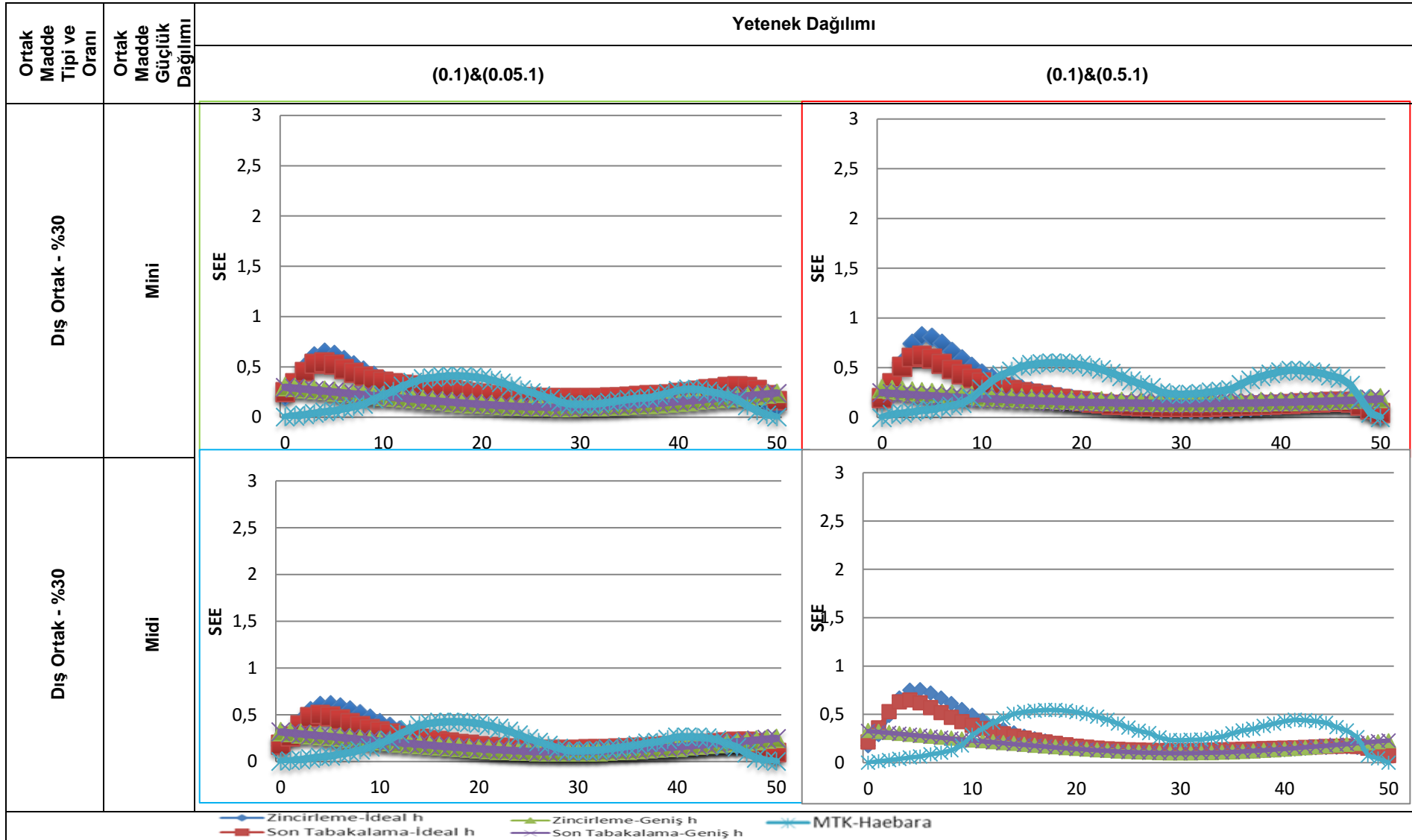
Şekil 1: 5000 Örneklem İçin Ortak Madde Oranının %30 olduğu Koşulda Elde Edilen SEE Değerleri



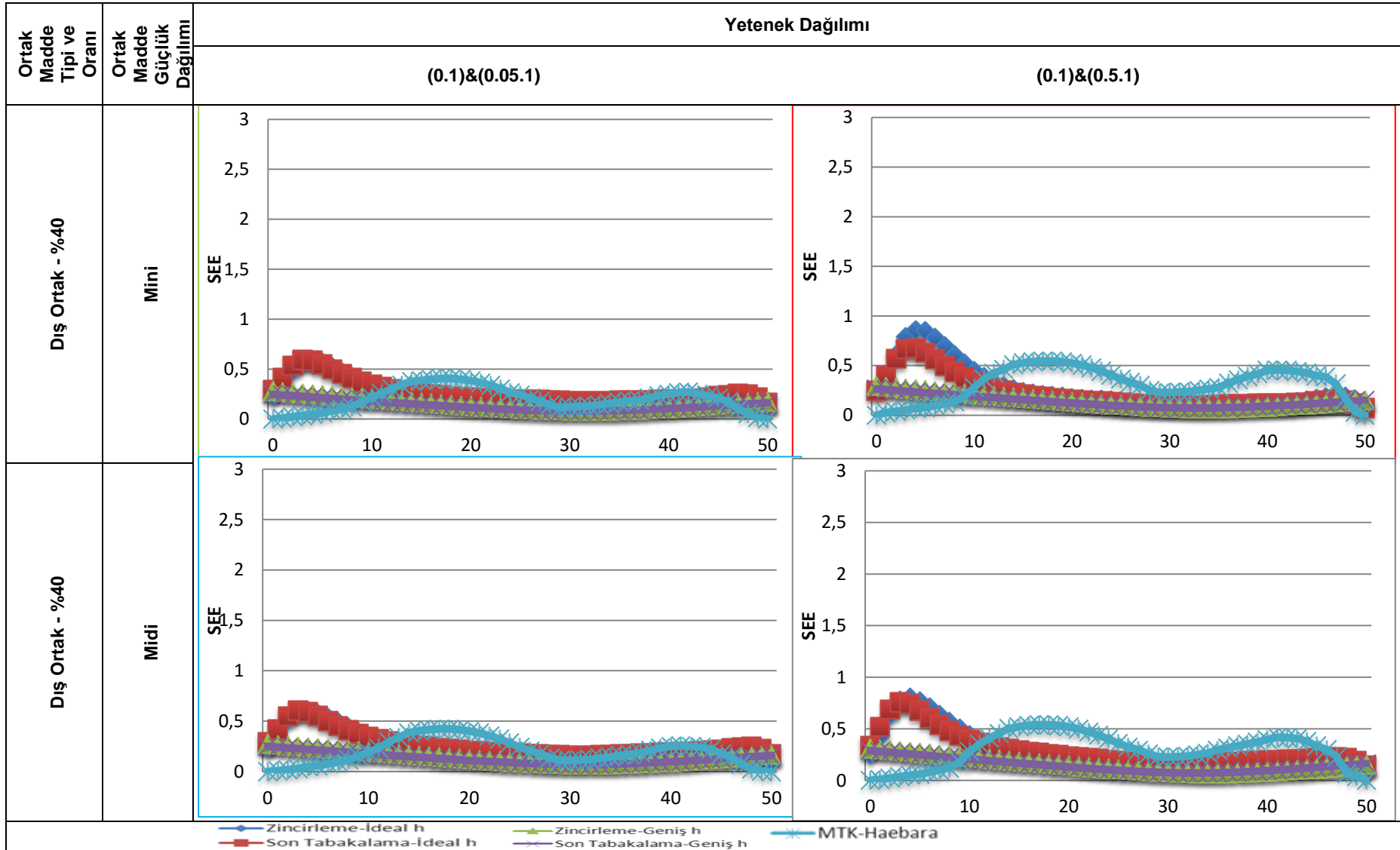
Şekil 2: 5000 Örneklem İçin Ortak Madde Oranının %40 olduğu Koşulda Elde Edilen SEE Değerleri



Şekil 3: 5000 Örneklem İçin Ortak Madde Oranının %20 olduğu Koşulda Elde Edilen SEE Değerleri



Şekil 4: 5000 Örneklem İçin Ortak Madde Oranının %30 olduğu Koşulda Elde Edilen SEE Değerleri



Şekil 5: 5000 Örneklem İçin Ortak Madde Oranının %40 olduğu Koşulda Elde Edilen SEE Değerleri

## ÖZGEÇMİŞ

### Kişisel Bilgiler

<i>Adı Soyadı</i>	Çiğdem AKIN ARIKAN
<i>Doğum Yeri</i>	Tunceli
<i>Doğum Tarihi</i>	20/01/1983

### Eğitim Durumu

<i>Lise</i>	Tunceli Atatürk Lisesi	2001
<i>Lisans</i>	Marmara Üniversitesi	2006
<i>Yüksek Lisans</i>	Abant İzzet Baysal Üniversitesi	2010
<i>Yabancı Dil</i>	İngilizce: Okuma (Çok iyi). Yazma (Orta). Konuşma (Orta)	

### İletişim

<i>e-Posta Adresi</i>	<a href="mailto:akincgdm@gmail.com">akincgdm@gmail.com</a>
-----------------------	--

<i>Jüri Tarihi</i>	08.06.2016
--------------------	------------