



Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü

Eđitim Bilimleri Ana Bilim Dalı

Eđitimde Ölçme ve Deđerlendirme Bilim Dalı

**DENK OLMAYAN GRUPLARDA ORTAK MADDE DESENİ
KULLANILARAK MADDE TEPKİ KURAMINA DAYALI
EŐİTLEME YÖNTEMLERİNİN KARŐILAŐTIRILMASI**

Bilge GÖK

Doktora Tezi

Ankara, 2012

DENK OLMAYAN GRUPLARDA ORTAK MADDE DESENİ
KULLANILARAK MADDE TEPKİ KURAMINA DAYALI
EŞİTLEME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Bilge GÖK

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Doktora Tezi

Ankara, 2012

KABUL VE ONAY

Bilge GÖK tarafından hazırlanan “Denk Olmayan Gruplarda Ortak Madde Deseni Kullanılarak Madde Tepki Kuramına Dayalı Eşitleme Yöntemlerinin Karşılaştırılması” başlıklı bu çalışma, 14.12.2012 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Doktora tezi olarak kabul edilmiştir.



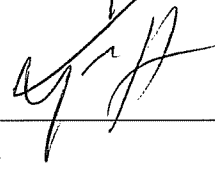
Prof. Dr. Selahattin GELBAL (Başkan)




Doç. Dr. Hülya KELEÇIOĞLU (Danışman)



Doç. Dr. Adnan KAN



Doç. Dr. Nuri DOĞAN



Yrd. Doç. Dr. Burcu ATAR

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylım.

Prof. Dr. Yusuf ÇELİK

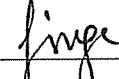
Enstitü Müdürü

BİLDİRİM

Hazırladığım tezin tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin kâğıt ve elektronik kopyalarının Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- Tezimin tamamı her yerden erişime açılabilir.
- Tezim sadece Hacettepe Üniversitesi yerleşkelerinden erişime açılabilir.
- Tezimin yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

14.12.2012



Bilge Gök

TEŞEKKÜR

Uzun yıllar süren yüksek lisans ve doktora öğrenimim boyunca ilgi ve desteğini benden hiçbir zaman esirgemeyen, her çıkmaza düştüğümde bana yol gösteren, motive edici ve cesaret verici sözleriyle her zaman yanımda olduğunu hissettiren kendisinden her anlamda çok şey öğrendiğim tez danışmanım, canım hocam Doç. Dr. Hülya KELECİOĞLU'na;

Yüksek lisans öğreniminden bu yana akademik yaşantımda kendimi geliştirmemde büyük payı olan değerli hocalarım Prof. Dr. Selahattin GELBAL'a, Doç. Dr. Nuri DOĞAN'a;

Tezimi okuyarak görüşleri ve eleştirileriyle çalışmama katkıda bulunan jüri üyesi sayın hocalarım Doç. Dr. Adnan KAN ve Yrd. Doç. Dr. Burcu ATAR'a;

Bugünlere gelmemde en büyük paya sahip olan ve benden sevgi ve desteklerini hiçbir zaman esirgemeyen sevgili annem ve babama;

Sabrı, anlayışı, özverisi ile her zaman yanımda olan, beni destekleyen ve beni hiçbir zaman yalnız bırakmayan canım eşime;

Tez çalışmam sürecinde yardım ve desteğini benden esirgemeyen, bu tezin oluşmasında büyük katkısı olan, hem akademik anlamda hem de manevi olarak yanımda olan ve umutsuzluğa kapıldığım anlarda güven verici sözleriyle beni destekleyen canım arkadaşım Kübra ATALAY'a;

Bu zorlu ve yorucu süreçte katkı ve yardımları için arkadaşlarım Ayşe YOLCU ve Mehmet KATRANCI'ya;

Bu süreçte adını sayamadığım her zaman bana destek olan ve yanımda olan tüm hocalarıma ve arkadaşlarıma;

NE KADAR TEŞEKKÜR ETSEM AZDIR....

ÖZET

GÖK, Bilge. *Denk Olmayan Gruplarda Ortak Madde Deseni Kullanılarak Madde Tepki Kuramına Dayalı Eşitleme Yöntemlerinin Karşılaştırılması*, Doktora Tezi, Ankara, 2012.

Araştırmada farklı koşullara (örneklem büyüklüğü, yetenek dağılımı, test uzunluğu ve model türü) göre türetilen test formlarını madde tepki kuramına dayalı kestirim yöntemlerini (ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord) kullanarak eşitlemek ve bu yöntemlerden elde edilen sonuçları karşılaştırmak amaçlanmıştır. Bununla birlikte araştırmada hangi koşullarda hangi yöntemin daha iyi sonuç verdiği de araştırılmıştır. Araştırma iki ve üç parametrelili lojistik modele uyumlu iki kategorili simülatif veriler kullanılarak yürütülmüştür. Bu araştırmada iki test formunu eşitleyebilmek için “denk olmayan gruplarda ortak madde/test (NEAT) deseni” kullanılmıştır. Verilerin türetilmesinde WINGEN3 programından yararlanılmış ve araştırmada kullanılan 36 koşulun her biri için 50 tekrar yapılmıştır. Madde ve yetenek parametrelerinin kestirilmesi PARSCALE 4.1 ile ayrı kalibrasyon için test eşitleme ve ölçekleme IRTEQ ile yapılmıştır. Araştırmada madde ve yetenek parametreleri expected a posteriori (EAP) ile kestirilmiştir. Araştırmada elde edilen sonuçlar, eşitleme yanlılığı (BIAS) ve eşitleme hatası (RMSE) ölçütlerine göre değerlendirilmiştir.

Araştırmadan elde edilen sonuçlar doğrultusunda, koşullar genel olarak değerlendirildiğinde en iyi eşitlemelerin 3000 kişilik örneklem, 80 maddelik testler, benzer yetenek dağılımına sahip gruplar, 2PLM ve SL yöntemi kullanılarak elde edilebileceği ifade edilebilir. Ayrıca araştırmada ele alınan koşullar doğrultusunda, büyük örneklem ile daha uzun testler kullanıldığında ve benzer yetenek dağılımına sahip gruplarda yöntemlerin daha az hatalı ve yanlı olduğu sonucuna da ulaşılmıştır.

Anahtar Sözcükler

Test eşitleme, madde tepki kuramı, eşitleme hatası, eşitleme yanlılığı.

ABSTRACT

GÖK, Bilge. *Comparison of IRT Equating Methods Using the Common-Item Nonequivalent Groups Design*, Ph. D. Dissertation, Ankara, 2012.

The purpose of this research was to equate the test forms which were constructed in different conditions (sample size, ability distribution, length of test and model type) through scaling methods based on item response theory (mean-mean, mean-sigma and Stocking Lord) and to compare the results obtained from these methods. In addition, it was also investigated that which method gave better results under different conditions. The research was conducted with using dichotomous simulated data which was consistent with two and three parameter logistic model. In order to equate two test forms “the common-item nonequivalent groups” was used in this research. WINGEN3 program was utilized for data generation and 50 replication were done for 36 different condition used in this research. PARSCALE 4.1 was utilized for the prediction of item and ability parameters and IRTEQ was utilized for test equating and scaling in separate calibration. Item and ability parameters were predicted through expected a posteriori (EAP) in this research. The results obtained from this simulation study were evaluated based on equating bias (BIAS) and equating error (RMSE) criterions.

The results revealed that, when the conditions evaluated generally, the best equating occurred in 3000-subjects samples, 80-item tests, groups have similar ability distribution, using 2PLM and SL methods. Moreover, the results indicated that methods had less biased and less equating errors when large sample sizes together with long tests were used in groups which had similar ability distributions under the conditions considered in this research.

Key Words

Test equating, item response theory, equating error, equating bias.

İÇİNDEKİLER

KABUL VE ONAY	İ
BİLDİRİM	İİ
TEŞEKKÜR	İİİ
ÖZET.....	İV
ABSTRACT	V
İÇİNDEKİLER	VI
TABLolar DİZİNİ	İX
ŞEKİLLER DİZİNİ	X
BÖLÜM I.....	1
GİRİŞ	1
1.1. Problem Durumu	1
1. 2. Test Eşitleme.....	2
1.2.1. Yatay Eşitleme	4
1.2.2. Dikey Eşitleme (Ölçekleme)	5
1.3. Eşitleme Koşulları	5
1.3.1. Eşitlik	5
1.3.2. Gruplar Arası Değişmezlik	6
1.3.3. Simetri	6
1.3.4. Aynı Yapıyı Ölçme	7
1.4. Eşitleme Desenleri	8
1.4.1. Tek Grup Deseni	8
1.4.2. Random Grup Deseni	9
1.4.3. Denk Olmayan Gruplarda Ortak Test/Madde Deseni (NEAT Design)	10
1.5. Eşitleme Yöntemleri.....	11
1.5.1. Klasik Test Kuramına (KTK) Dayalı Eşitleme Yöntemleri.....	13
1.5.1.1. Eşit Yüzdelikli (Equipercntile) Eşitleme.....	13
1.5.1.2. Doğrusal (Linear) Eşitleme	14
1.5.1.3. Ortalama (Mean) Eşitleme	14
1.5.2. Madde Tepki Kuramına Dayalı Eşitleme.....	14
1.5.3. Madde Tepki Kuramına Dayalı Eşitleme Yöntemleri	18

1.5.3.1. Rasch Modeli	18
1.5.3.2. İki Parametrelı Lojistik Model	19
1.5.3.3. Üç Parametrelı Lojistik Model	19
1.5.4. Madde Tepki Kuramına Dayalı Ölçek Dönüştürme (Kalibrasyon) Yöntemleri	20
1.5.4.1. Ayrı Kalibrasyon/Kestirim (Separate Calibration/Estimation)	21
1.5.4.1.1. Ortalama-Ortalama ve Ortalama-Standart Sapma Yöntemleri	23
1.5.4.1.2. Karakteristik Eğri Dönüştürme Yöntemleri	24
1.5.4.2. Eş zamanlı Kalibrasyon/Kestirim (Concurrent Calibration/Estimation) ...	26
1.6. Eşitleme Hatası	27
1.7. Araştırmanın Önemi ve Amacı	28
1.8. Problem Cümlesi	31
1.9. Alt Problemler	31
1.10. Sınırlılıklar	33
1.11. Kısaltmalar	34
1.12. İlgili Araştırmalar	34
BÖLÜM II	40
YÖNTEM	40
2.1. Araştırmanın Türü	40
2.2. Veri Toplama Deseni	40
2.3. Simülasyon Faktörleri ve Koşulları	41
2.4. Verilerin Türetilmesi (Simülasyon Çalışması)	45
2.5. Araştırmada Kullanılan Eşitleme Yöntemleri	47
2.6. Eşitleme Sürecinin Uygulanması	49
2.7. Verilerin Analizi ve Değerlendirme Ölçütü	50
BÖLÜM III	54
BULGULAR VE YORUM	54
3.1. Alt problem 1'e ilişkin bulgular ve yorumlar	54
3.2. Alt problem 2'ye ilişkin bulgular ve yorumlar	60

3.3. Alt problem 3'e ilişkin bulgular ve yorumlar.....	64
3.4. Alt problem 4'e ilişkin bulgular ve yorumlar.....	65
3.5. Alt problem 5'e ilişkin bulgular ve yorumlar.....	68
3.6. Alt problem 6'ya ilişkin bulgular ve yorumlar	72
3.7. Alt problem 7'ye ilişkin bulgular ve yorumlar.....	72
3.8. Alt problem 8'e ilişkin bulgular ve yorumlar.....	76
3.9. Alt problem 9'a ilişkin bulgular ve yorumlar	79
BÖLÜM IV	82
SONUÇ VE ÖNERİLER.....	82
4.1. SONUÇLAR	82
4.2. ÖNERİLER	86
KAYNAKÇA	89
EKLER.....	108

TABLOLAR DİZİNİ

Tablo 1.1. Tek grup desenine ilişkin desen tablosu	9
Tablo 1.2. Random grup desenine ilişkin desen tablosu	9
Tablo 1.3. Denk olmayan gruplarda ortak madde/test desenine ilişkin desen tablosu ...	10
Tablo 2.1. İki ve üç parametrelili lojistik modele göre türetilen madde sayısı	44
Tablo 2.2. Araştırmada kullanılan simülasyon koşulları	44
Tablo 3.1. 3PLM ile kestirilen madde parametrelerinin simülasyon koşullarına göre yanlılık (BIAS) ve eşitleme hataları (RMSE)	55
Tablo 3.2. 2PLM ile kestirilen madde parametrelerinin simülasyon koşullarına göre yanlılık (BIAS) ve eşitleme hataları (RMSE)	60
Tablo 3.3. 3PLM ile kestirilen yetenek parametrelerinin simülasyon koşullarına göre yanlılık (BIAS) ve eşitleme hataları (RMSE)	65
Tablo 3.4. 2PLM ile kestirilen yetenek parametrelerinin simülasyon koşullarına göre yanlılık (BIAS) ve eşitleme hataları (RMSE)	69
Tablo 3.5. 3PLM ile yapılan gerçek puan eşitlemenin simülasyon koşullarına göre yanlılık (BIAS) ve eşitleme hataları (RMSE)	73
Tablo 3.6. 2PLM ile gerçek puan eşitlemesinin simülasyon koşullarına göre yanlılık (BIAS) ve eşitleme hataları (RMSE)	76
Tablo 3.7. Madde ve yetenek parametreleri ile gerçek puan eşitlemesi için en az hatalı sonuç veren yöntemler.....	80
Tablo 3.8. Yöntemlerin madde ve yetenek parametreleri ile gerçek puan eşitlemesi durumlarında en kararlardan en hatalı sonuçlar verene doğru sıralanışı.....	81

ŞEKİLLER DİZİNİ

Şekil 2.1. Araştırmada kullanılan eşitleme deseni	41
Şekil 2.2. MTK gerçek puan eşitleme örneği.....	48

BÖLÜM I

GİRİŞ

Bu bölümde problem durumu, araştırmanın önemi ve amacı, problem cümlesi, alt problemler, sınırlılıklar, kısaltmalar ve ilgili araştırmalara yer verilmiştir.

1.1. PROBLEM DURUMU

Standartlaştırılmış koşullarda testler uygulamanın temel amacı, cevaplayıcıların yeteneklerini mümkün olduğunca objektif ve doğru bir şekilde ölçmektir. Test puanları, herhangi bir okula veya işe yerleşmeye çalışan bireylerin yeteneklerinin ölçülmesi gibi amaçlar için kullanılır. Bireyleri okula ya da işe yerleştirmek amacıyla genellikle büyük ölçekli ve merkezi sınavlar uygulanır. Bu sınavların sonuçlarına dayalı olarak gerçekleştirilen yerleştirmeler de her yıl ya da yılda birkaç kez yapıldığından, bu sınavlar da belirli aralıklarla uygulanır. Bir işe ya da okula yerleştirmek için farklı zamanlarda yapılan sınavların amacı aynı olduğundan, her sınavda sorulan soruların güvenliğini sağlamak büyük bir sorun oluşturur. Bunun için aynı amaçla uygulanan testlerin pek çok formu geliştirilir. Aynı amaca yönelik olarak uygulanan sınavlarda farklı soruların sorulması, sınava giren adaylara bazı formların kolay, bazı formların zor gelmesine yol açabilir. Bir cevaplayıcının diğer bir cevaplayıcıdan daha zor bir test formu almasını engellemek ve aynı testin pek çok formu üzerinde karşılaştırılabilir puanlar elde etmek için test formlarının eşitlenmesine ihtiyaç duyulur (Cook & Eignor, 1991). Örneğin, aynı özelliği ölçen iki adet standartlaştırılmış test (A ve B) ve bu testlerin uygulandığı iki farklı grup (grup 1 ve grup 2) olduğunu varsayalım. Grup 1’de yer alan öğrenciler yalnızca A Testini, grup 2’de yer alan öğrenciler ise yalnızca B testini almış olsunlar. Grup 1 için A testine ilişkin ortalama puan 84; grup 2 için B testine ilişkin ortalama puan 80 olsun ve t testi sonucuna göre iki grubun ortalama puanları arasında anlamlı bir farklılık elde edilsin. Bu durumda, bu iki test ölçtüğü özellik açısından karşılaştırıldığında, grup 1’de yer alan öğrenciler grup 2’de yer alan öğrencilerden daha başarılıdır şeklinde bir sonuca varılabilir mi? İşte bu soruya verilecek cevap “test eşitleme” kavramının önemini ortaya koymaktadır. Bir seçme

sınavında kesme puanının üzerinde puan alarak başarılı olan bir adaya aynı amaçla yapılan sınavın başka formu uygulandığında söz konusu aday, kesme puanının altında puan alabilir. Bu nedenle, her yıl tekrarlanan aynı amaca yönelik bu tür sınavların paralel olduğu varsayılan formlarının eşitlenmesi gerekir. Aynı özelliği ölçen farklı testlerden veya test formlarından elde edilen puanların karşılaştırılabilir olup olmadığının belirlenmesi, eğitimde önemli bir sorun teşkil etmektedir. Bu nedenle, test eşitlemeye en genel anlamda, aynı örtük özelliği ölçen testlerin farklı formlarından elde edilen puanları karşılaştırmak amacıyla ihtiyaç duyulur (Petersen, Marco & Stewart, 1982; Tsai, 1997; Wolkowitz, 2008).

Farklı test formlarından elde edilen puanlar karşılaştırılmak istendiğinde bu puanların denk olması gerekir (Dorans, 1990). Bir testin farklı formlarından alınan puanlar çoğunlukla aynı ölçekte yer almadığı için bu puanların birbirleriyle karşılaştırılmaları mümkün olmaz. Aynı teste ait farklı formların ya da farklı yıllarda aynı amaca yönelik olarak yapılan farklı testlerin, kapsam ve istatistiksel özellikler bakımından mümkün olduğunca denk olması sağlanmalıdır (Beguin, 2000; Tsai, Hanson, Kolen & Forsyth, 2001; Kolen & Brennan, 2004). Sınavlar aynı amaçla yapılmış olsa bile sınava giren öğrenciler sınavın farklı formlarından farklı puanlar alabilmekte ve böylece sınava giren öğrenciler aldıkları formlara göre avantajlı olmakta ya da dezavantajlı duruma düşmektedirler. Bu durumda, farklı formları alan cevaplayıcılar için yetenekteki farklılıklar formun güçlüğündeki farklılıklarla karıştırılabilir (Albano, 2010). Bu nedenle, bir testin daha kolay veya zor bir formunu alan cevaplayıcılar tarafından karşılaşılan avantaj veya dezavantajı ortadan kaldırmak için puanların eşitlenmesi gerekir (Thorndike, 1971; Harris, 2003).

1. 2. TEST EŞİTLEME

Test eşitleme, bir testin en az iki formu olduğu durumlarda kullanılır ve farklı formları alan cevaplayıcıları birbirleriyle karşılaştırır (Kolen & Brennan, 1995; Mohandas, 1998; Felan, 2002). Farklı uygulamalar veya farklı test formları karşısında öğrencileri karşılaştırma ihtiyacı bu testlerin aynı ölçeğe yerleştirilmesini gerektirir. Test eşitleme, iki veya daha çok testten alınan puanlar arasındaki ilişkiyi ortaya çıkaran istatistiksel bir

tekniktir. Puanlar arasındaki ilişkilere dayanarak, testler ortak bir ölçeğe yerleştirilir (Chu & Kamata, 2003). Diğer bir deyişle, iki veya daha fazla testi ortak bir ölçeğe yerleştirmek için kullanılacak istatistiksel yöntem “test eşitleme” olarak adlandırılır ve eşitleme sonucunda formlardan elde edilen puanlar birbirlerinin yerine kullanılabilir (Hambleton & Swaminathan, 1985; Holland & Dorans, 2006; Kolen & Brennan, 2004). Başarılı bir eşitlemeden sonra, örneğin eşitlenmiş bir test formundan 26 puan alan cevaplayıcıların eşitlenmiş farklı test formlarından 26 puan alan diğer cevaplayıcılarla aynı başarı düzeyinde olduğu düşünülür (Kolen & Brennan, 1995). Puanların eşitlenmesi, X formundan Y formuna ham puanları dönüştürmek için eşitleme fonksiyonunun kullanılmasıyla gerçekleştirilebilir (Holland & Dorans, 2006).

Eşitleme için bir formun birim sistemini başka bir formun birim sistemine dönüştürmek gerekir ve bu dönüştürmeden sonra formların eş değer olacağı belirtilir (Angoff, 1971; Akt. Woldbeck, 1998). Angoff (1971) bu sayısal dönüştürmeyi Fahrenheit derece ile santigrat dereceyi karşılaştırmak için kullanılan dönüştürme formüllerine benzetmiştir (Akt. Buras, 1996). Böylece bir formdan elde edilen puanlar diğer bir formdan elde edilen puanlar ile doğrudan karşılaştırılabilir. Test geliştirme uzmanları içerik ve istatistiksel özellikler açısından benzer test formları oluşturmak için çabalasa da, bu test formları genellikle güçlük düzeyleri açısından farklılık gösterir. Güçlük düzeylerindeki bu farklılıkları açıklayabilmek için test formlarını eşitlemek gerekir (Dongyang, 2009). Genellikle büyük ölçekli sınavlarda içerik ve güçlük düzeyleri bakımından benzer olan farklı test formları kullanılır. Pek çok test formuna gereksinim duyulduğu bir durumda farklı formlardan elde edilen puanların karşılaştırılabilir olmasından emin olmak gerekir (Dismer-Davis, 1988; Norman-Dvorak, 2009). Eğer eşitleme tam anlamıyla yerine getirilirse, bir cevaplayıcı tarafından alınan belirli bir test formunun, cevaplayıcının test puanı üzerinde sistematik bir etkiye sahip olması beklenmez ve elde edilen puanlar rahatlıkla karşılaştırılabilir (Nozawa, 2008). Böylece test sonuçlarına göre doğru kararlar alınabilir.

Test eşitleme büyük ölçekli test uygulamalarında önemli bir rol oynar. Test eşitleme ile ilgili araştırmalar incelendiğinde, testleri eşitlemenin önemli ve gerekli olduğu vurgulanmakla birlikte, “Her durumda testleri eşitlemek gerekli midir?” sorusuna verilecek cevabın irdelenmesi gerekir. Eşitlemenin gerekli ya da uygun olup olmadığına

bakıldıktan sonra testler eşitlenmelidir. İki test formundan elde edilen puanların dağılımı çok benzer olmadığında eşitleme uygun olmaz; puan dağılımları çok benzer olduğunda ise eşitleme gereksizdir (Harris & Crouse, 1993). Bu yüzden eşitleme, sadece sürecin uygun veya gerekli olduğuna dair kanıt toplandıktan sonra yürütülmelidir (Chulu, 2006). Eşitleme süreci en doğru eşitleme ilişkisini ortaya çıkaracak şekilde tasarlanmalıdır. Eşitleme süreci hatalı sonuçlara yol açarsa, farklı formlardan elde edilen puanlar karşılaştırılmaz. Bu durumda puanların değerlendirileceği ortak bir metrik de söz konusu değildir (Kolen & Brennan, 2004).

Test eşitleme yatay (horizontal) ve dikey (vertical) eşitleme olmak üzere iki kısımda incelenmektedir (Cook & Eignor, 1983). Bu araştırmada test formlarını eşitlemek için yatay (horizontal) eşitleme yöntemi kullanılmıştır.

1.2.1. Yatay Eşitleme

Yatay eşitleme, test güvenliğini sağlamak için bir testin pek çok formuna gereksinim duyulduğunda kullanılır. Yatay eşitlemede amaç, benzer kapsam ve psikometrik özelliğe sahip test formlarından doğrudan karşılaştırılabilir test puanları elde etmektir. Bu tür eşitleme yaygın olarak kullanılan eşitleme türüdür ve genellikle büyük ölçekli standartlaştırılmış testlerin alternatif formlarını eşitlemek için kullanılır (Skaggs & Lissitz, 1982, 1986).

Benzer güçlük düzeylerindeki testlerin uygulandığı gruplarda benzer yetenek dağılımları da gözlenirse, bu testler yatay eşitleme ile eşitlenebilir. Aynı sınıf düzeyinde farklı yıllarda uygulanan matematik testlerini eşitlemek veya bir sınavın benzer yetenek düzeyindeki gruplara uygulanan formlarını eşitlemek, bu tür eşitlemeye örnek olarak gösterilebilir. Eğer test puanlarının dağılımları benzer değilse gruplar arası farklılık çeşitli yollarla kontrol edilerek eşitleme yapılır. Yatay eşitlemede formların güçlük ve kapsam açısından birbirine paralel olduğu varsayılır (Kolen, 1988). Bir testin formları güvenilirlik, güçlük ve kapsam bakımından büyük farklılık gösterirse ve cevaplayıcıların yetenek dağılımları da birbirinden farklıysa eşitleme yöntemleri işlevlerini yerine getiremez.

1.2.2. Dikey Eşitleme (Ölçekleme)

Dikey eşitleme (ölçekleme), farklı başarı düzeyindeki bireylerin yeteneklerinin karşılaştırılmasını amaçlayan farklı güçlük düzeylerindeki birden fazla testle yapılır. Testi alan bireylerin yetenek dağılımları da farklıdır. Bu tür eşitlemede aynı bilgi veya becerileri ölçen fakat güçlükleri farklı olarak tasarlanan iki testten elde edilen puanlar eşitlenir (Felan, 2002). Farklı sınıf düzeylerindeki ilköğretim öğrencilerinin okuma yeteneğinin eşitlenmesi dikey eşitlemeye bir örnektir. Kolen (1994), dikey eşitlemenin farklı sınıf düzeyindeki öğrenciler arasında karşılaştırmalar yapılmasına ve zaman içerisinde öğrencilerin gelişimlerinin karşılaştırılmasına imkân verdiğini ifade eder. Dikey eşitleme, yatay eşitlemeden daha karmaşık bir süreç gerektirir çünkü dikey eşitlemenin amacı, iki farklı cevaplayıcı grubu ve iki farklı yetenek düzeyini karşılaştırmaktır. Bu tür eşitleme genellikle başarı testlerinde kullanılır (Skaggs & Lissitz, 1982, 1986).

1.3. EŞİTLEME KOŞULLARI

Test geliştirme uzmanları, test eşitleme için karşılanması gereken koşullar konusunda farklı görüşlere sahiptir. Petersen, Kolen ve Hoover (1989)'a göre, iki testin eşitlenebilmesi için şu koşulların yerine getirilmesi gerekir: Eşitlik, gruplar arası değişmezlik, simetri ve aynı yapıyı ölçme. Dorans ve Holland (2000) ile Lord (1980)'a göre eşitlemenin yapılabilmesi için karşılanması gereken koşullar; aynı yapıyı ölçme, güvenilirliklerin eşit olması, simetri, eşitlik ve gruplar arası değişmezliktir. Hambleton ve Swaminathan (1985) ise eşitleme koşullarını eşitlik, simetri, gruplar arası değişmezlik ve tek boyutluluk olarak sıralamışlardır. Bu araştırmada Petersen, Kolen ve Hoover (1989) tarafından ifade edilen koşullar temel alınmıştır ve ayrıntılı olarak bu koşullar üzerinde durulmuştur.

1.3.1. Eşitlik

Lord (1980), “eşitlik” kavramını şu şekilde açıklamıştır: “Eğer her θ yetenek düzeyindeki cevaplayıcılar için X ve Y testleri denkse, bu cevaplayıcıların X veya Y

testini almaları, herhangi bir fark oluşturmamalıdır". Lord (1977, 1980)'un ifade ettiği gibi eşitlik koşulu pek çok özelliğe sahiptir (Akt. Hambleton & Swaminathan, 1985):

1. Farklı yetenek ya da becerileri ölçen testler eşitlenemez.
2. Eşit güvenilirliğe sahip olmayan testlere ilişkin ham puanlar eşitlenemez.
3. Farklı güçlük düzeylerine sahip testlere ilişkin ham puanlar (dikey eşitleme durumlarında) eşitlenemez (doğrusal olmayan ilişki).
4. Bir X testinin puanlarının θ yetenek düzeyindeki koşullu frekans dağılımı (X testine ilişkin θ puanın $f[x | \theta]$ fonksiyonu), Y testinin dönüştürülmüş x (y) puanına ilişkin $f[x(y) | \theta]$ koşullu frekans dağılımı ile aynı olmalıdır.
5. Paralel olmayan testler eşitlenemez; çünkü paralel olmayan testlerin hata puanları dağılımları farklıdır ve güvenilirlikleri de eşit değildir.

Eşitlik kavramı, yukarıda da açıklandığı üzere test puanlarını eşitlemede büyük bir role sahiptir. Bu koşul, test puanları eşitlenmeden önce karşılanması gereken koşullardan sadece biridir.

1.3.2. Gruplar Arası Değişmezlik

Gruplar arası değişmezlik, test eşitlemenin eşitleme sürecinde yer alan cevaplayıcılardan bağımsız olması ve eşitlemeden türetilen dönüşümün benzer tüm koşullara uygulanmasıdır. Alt popülasyonlar karşısında eşitleme ilişkilerinin denkliğini kontrol etmek, gruplar arası değişmezlik gereksinimini değerlendirmenin güvenilir bir yoludur.

1.3.3. Simetri

Simetrik olan eşitleme ilişkilerini gerektiren simetri özelliği (Lord, 1980), X formundan alınan bir puanı Y formuna dönüştürmede kullanılan fonksiyon, Y formundan alınan bir puanı X formuna dönüştürmek için kullanılan fonksiyonun tersinin alınmasını gerektirir. Eşitleme, ister X formundan Y formuna, isterse Y formundan X formuna yapılsın, eşitlenmiş puanların aynı olması anlamındadır. Kısaca bu özellik dönüşümün tersine çevrilebilir olmasını ifade eder.

1.3.4. Aynı Yapıyı Ölçme

Bu özellik, söz konusu olan iki aracın aynı özelliği (örtük özellik, yetenek veya beceri) ölçmesini ifade eder. Farklı içeriğe sahip ve farklı özelliği ölçen testler eşitlenemez. Örneğin, bir sözel yetenek testinden elde edilen puanlar sayısal yetenek testinden elde edilen puanlara eşitlenemez.

Bununla birlikte eşitlemede tek boyutlu Madde Tepki Kuramı (MTK) kullanıldığında, yukarıdaki koşullara eşitlenecek testlerin tek boyutlu olması koşulu da eklenmelidir (Hambleton & Swaminathan, 1985). Uygulamada, yukarıda sözü edilen dört durumun birlikte karşılanması pek mümkün değildir. Bundan dolayı Petersen ve diğerleri (1989), eşitlemede hangi koşulların karşılanması ve hangi yöntemlerin kullanılması gerektiği ile ilgili olarak uzmanlar arasında bir anlaşmazlık olduğunu ifade etmişlerdir. Angoff (1971) eşitleme için sadece testte yer alan maddelerin aynı yapıyı ölçmesi ve gruplar arası değişmezlik koşullarının karşılanmasına ihtiyaç olduğunu ifade eder. Dorans (1990) ise bir testi diğerine eşitlemek için her iki testin de aynı yapıyı ölçmesi gerektiğini ve her iki testin de tek boyutlu maddeleri içermesinin zorunlu olduğunu ifade eder. Reckase, Ackerman ve Carlson (1988) tarafından yapılan bir çalışmada ise, madde havuzundan veya madde bankasından seçilen maddelerin MTK modelleri için tek boyutluluk koşulunu karşılaması gerektiği ortaya konmuştur. Bir testin tek boyutlu olduğu ne zaman söylenebilir? Lord ve Novick (1968)'e göre boyut, yerel bağımsızlık varsayımını karşılamak için gereksinim duyulan yeteneklerin toplam sayısını tanımlar. Nandakumar (1994) ise, yerel bağımsızlık varsayımını karşılamak üzere maddelere verilen cevaplar için sadece bir yeteneğe gereksinim duyulursa, bu maddelerin tek boyutlu olarak ifade edildiğini ileri sürer.

Eşitlenecek testler, yukarıda ifade edilen eşitleme koşullarını karşılamalıdır. Eşitleme koşulları göz önüne alınarak eşitlemenin uygun olup olmadığı belirlendikten sonra, eşitleme çalışmaları için şu basamaklar izlenmelidir (Kolen & Brennan, 2004):

- Eşitlemenin amacına karar vermek
- Alternatif formlar oluşturmak
- Veri toplama desenini (eşitleme desenini) seçmek ve uygulamak
- Eşitlemenin operasyonel tanımlarını seçmek

- İstatistiksel kestirim yöntemlerini seçmek
- Eşitleme sonuçlarını değerlendirmek.

1.4. EŞİTLEME DESENLERİ

Eşitlemenin amacı belirlenip formlar oluşturulduktan sonra veri toplama desenine karar verilir. Eşitlemede veri toplama deseni “eşitleme deseni” olarak ifade edilir (Kolen & Brennan, 2004; Mao, von Davier & Rupp, 2006). Eşitleme desenleri, her bir cevaplayıcının aldığı test formlarının sayısı ile grup sayısına göre farklılık gösterir ve bu desenler cevaplayıcıların yetenekleri arasındaki farkları çeşitli yollarla kontrol etmeye çalışır (Holland, Dorans & Petersen, 2007). Uygun bir eşitleme deseninin kullanılması eşitlemenin en önemli basamaklarından biridir (Holland & Dorans, 2006). Literatürde eşitleme desenlerinin çeşitli sınıflamalarının yapıldığı görülmektedir (Angoff, 1971; Crocker & Algina, 1986; Kolen & Brennan, 2004). Bu çalışmada Crocker ve Algina (1986) ile Kolen ve Brennan (2004)’ın üçlü sınıflaması ele alınmıştır:

- Tek grup deseni (Single group design)
- Random grup deseni (Random group design)
- Denk olmayan gruplarda ortak madde/test deseni (Non-equivalent groups anchor test design-NEAT).

Bu araştırmada “denk olmayan gruplarda ortak madde/test deseni” kullanılmıştır.

1.4.1. Tek Grup Deseni

En basit test eşitleme desendir. Eşitlenecek iki test aynı cevaplayıcı grubuna verilir. Aynı cevaplayıcıların her iki testi de alması nedeniyle testlerin zorluk düzeyleri cevaplayıcıların yetenek düzeylerinden etkilenmez. Ne var ki aynı gruba iki uygulama yapmanın etkisi eşitleme sürecini etkileyebilir. Bir başka deyişle, testlerin veriliş sırası bir hata kaynağı oluşturabilir. Bundan sakınmak için formların sırası dengelenir (counterbalancing). Dengelemenin yolu, tek grup deseniindeki bireylerin yarısına eşitlenecek testlerden X formu önce verilirken, diğer yarısına da Y formunun önce verilmesidir. Böylelikle testlerin veriliş sırasından kaynaklanacak hatalar (sıra etkisi) giderilmiş olur (Kolen & Brennan, 2004). Bu desenin dezavantajlarından biri yorgunluk

nedeniyle ikinci formdan elde edilecek performansın düşük olmasıdır. Bu desen küçük örneklemelerde doğru bir eşitleme sağlar. Yorgunluk ve sıra etkisi dengelenerek kontrol edilirse, bu deseni kullanmanın yararı random grup deseninden daha küçük örneklem gerektirmesidir. Bu desen, sıra etkisinden dolayı uygulamada nadiren kullanılan bir desendir (Kolen & Brennan, 2004). Tek grup deseni Tablo 1.1’de gösterilmiştir:

Tablo 1.1. Tek grup desenine ilişkin desen tablosu

Örneklem	Form X	Form Y
G1	√	√

1.4.2. Random Grup Deseni

Bu desende gruplar random olarak ikiye ayrılır. Gruplara her iki test de uygulanır fakat her gruba sırasıyla farklı bir test uygulanır. Grup 1 X_1 testini aldıktan sonra Y_2 testi, grup 2 de Y_1 testini aldıktan sonra X_2 testi uygulanır. Bu sarmal süreç, X ve Y formunu alan random eşdeğer grupların karşılaştırılmasına imkân sağlamaktadır. Bu desende gruplardan gelen farklılık azaltılarak eşitlemenin hassasiyeti artırılır. Bunu sağlamak için bu desende daha büyük gruplara gereksinim duyulur (Zhu, 1998). Bu desende aynı gruba iki uygulama yapmaktan kaynaklanan etkiler ortadan kalkmaktadır. Ne var ki, gruplar aynı olmadığı için yetenek dağılımındaki farklılıklar, eşitleme sürecinde bilinmeyen bir yanlılık ortaya çıkarabilir. Bu desende grupların iki formdaki performans düzeyleri arasındaki fark, formların güçlük düzeyleri arasındaki farkı göstermektedir (Kolen ve Brennan, 2004). Bu desen Tablo 1.2’de gösterilmiştir:

Tablo 1.2. Random grup desenine ilişkin desen tablosu

Örneklem	Form X_1	Form X_2	Form Y_1	Form Y_2
RG1	√			√
RG2		√	√	

1.4.3. Denk Olmayan Gruplarda Ortak Test/Madde Deseni (NEAT Design)

Ortak test/madde deseni, alanyazında en çok kullanılan desendir (Kolen & Brennan, 2004; Chulu, 2006; Keller III, 2007; Norman-Dvorak, 2009; Kim, Walker, Mchale, 2010). Bu desen, denk olmayan gruplara uygulanır ve ortak maddeler üzerinde grupların performansını karşılaştırarak iki grup arasındaki eşitleme ilişkisini ortaya çıkarmada kullanılır. Cevaplayıcılar eşitlenecek formlardan sadece birini yanıtlar. Bu desen denk olmayan iki gruba uygulandığı için denk olmayan gruplarda ortak madde ya da ortak test deseni (NEAT) olarak adlandırılır (von Davier, Holland & Thayer, 2004). Ortak test iki biçimde düzenlenebilir. Eşitlenecek her bir test cevaplayıcının puanının bir parçası olarak ortak madde seti içerirse buna “iç ortak test” adı verilir. Eşitlenecek testlerle beraber eş zamanlı olarak iki gruba da başka bir ortak test uygulanır ve bu test cevaplayıcının puanının bir parçası olmazsa “dış ortak test” adını alır (Crocker & Algina, 1986). Ortak maddeler her bir test formunu alan grupların yeteneğindeki farklılıkları düzeltmede kullanılır (Angoff, 1971).

Bu desende, ortak maddeler iki cevaplayıcı grubu arasındaki farklılıkları değerlendirmede kullanılır. Ortak maddeler doğru bir şekilde seçildiği sürece bu desen tek grup deseni veya denk grup desenindeki problemleri azaltır (Holland & Dorans, 2006; Hambleton, Swaminathan & Rogers, 1991). Bu desen Tablo 1.3’te gösterilmiştir:

Tablo 1.3. Denk olmayan gruplarda ortak test/madde desenine ilişkin desen tablosu

Örneklem	Form X	Form Y	Ortak Test
G1	√	√	√
G2			√

Denk olmayan gruplarda ortak test deseni, sıklıkla aynı formun güvenlik nedenlerinden dolayı tekrar uygulanamadığı test koşullarında kullanılır ve bu desene çoğunlukla başarı testlerini kullanan büyük ölçekli sınavlarda rastlanır (Kolen & Brennan, 2004). Ortak maddeler, eşitleme fonksiyonunu belirlemede önemli bir rol oynamaktadır. Bu bakımdan, eşitleme çalışmaları yürütülürken ortak maddelerin özelliklerine dikkat edilmelidir. Genellikle ortak maddeler güçlük ve içerik bakımından tüm testi iyi temsil etmelidir. Bununla birlikte ortak maddeler her bir test formunda aynı yerde yer almalıdır ve farklı formlarda yer alan ortak maddeler aynı olmalıdır (Kelime değişikliği

yapılmamalı veya birbirinin alternatifi olan maddeler kullanılmamalıdır.) (Kolen & Brennan, 2004).

Ortak maddelerin sayısı eşitleme sonuçlarını etkilemektedir. Wright ve Stone (1979) her biri 60 maddeden oluşan X ve Y testleri arasında bağlantı kurmak için 10-20 maddeye ihtiyaç duyulduğunu öne sürmüşlerdir. Hambleton ve diğerleri (1991) ortak maddeler için gereksinim duyulan madde sayısının, testteki madde sayısının yaklaşık olarak %20-25'i arasında olması gerektiğini ifade etmişlerdir. Reckase (1979), ortak maddeler için gereksinim duyulan madde sayısının örneklem sayısı 300'den fazla olduğunda 5-15 madde arasında olduğunu rapor etmiştir (Akt. Skaggs & Lissitz, 1986). Angoff (1971) ise, ortak testin tüm testteki madde sayısının %20'si kadar olması gerektiğini ifade etmiştir. Araştırmalarda ortak maddelerin sayısının minimum 15 olması gerektiği ileri sürülmekle birlikte (McKinley ve Reckase, 1981), tek boyutluluk varsayımı karşılandığında daha az maddenin kabul edilebilir olduğu da ifade edilmektedir (Kolen & Brennan, 2004). Ayrıca yapılan araştırmalar, ortak maddelerin sayısının artmasının eşitleme hatasını azalttığını göstermiştir (Kolen & Brennan, 2004).

1.5. EŞİTLEME YÖNTEMLERİ

Eşitleme yöntemlerinin pek çoğu, farklı ihtiyacı karşılamak için geliştirilmiştir. Yöntemlerin her biri farklı teori veya varsayımlara dayalıdır. Alanyazında eşitleme yöntemlerinin farklı şekillerde sınıflandırıldığı görülmektedir. Bu sınıflamalardan en yaygın kullanılanı, yöntemlerin geleneksel eşitleme yöntemleri ve Madde Tepki Kuramına dayalı eşitleme yöntemleri olarak sınıflandırılmasıdır. İki yöntemin dayandığı kuramsal temel, varsayımları ve matematiksel fonksiyonları büyük ölçüde birbirinden farklıdır. Geleneksel eşitleme yöntemleri, gözlenen puanların gerçek puan ve hatadan oluştuğunu ifade eden klasik test kuramına dayalıdır (Hambleton & Jones, 1993). Bu yöntemler; eşit yüzdellikli eşitleme (equipercentile), doğrusal (linear) eşitleme ve ortalama (mean) eşitleme olmak üzere üçe ayrılır (Barnard, 1996). MTK'yi kullanarak formları eşitlemenin pek çok yararı olmasına rağmen, geleneksel eşitleme yöntemleri daha az varsayım gerektiren daha basit ve daha pratik bir yol olabilir (Kolen & Brennan, 2004; Livingston, 2004). Bununla birlikte MTK ve geleneksel eşitleme yöntemlerini

karşılaştıran çalışmalar incelendiğinde, bazı araştırmalarda MTK eşitleme yöntemlerinin daha kararlı sonuçlar verdiği bulunurken (Han, Kolen & Pohlman, 1997; Yang & Houang, 1996); bazılarında ise MTK ve geleneksel eşitleme yöntemlerinin benzer sonuçlar verdiği görülmüştür (Harris & Kolen, 1986; von Davier & Wilson, 2008).

Test eşitleme yöntemlerinin bir diğer sınıflaması ise farklı psikometrik modellere dayalı olarak eşitlemenin gözlenen ve gerçek puana dayalı olarak yapılmasıdır. Gözlenen puan eşitlemede, eşitleme fonksiyonları hesaplanırken gerçek puanlar doğrudan göz önüne alınmaz. Gözlenen puan eşitlemenin amacı, paralel formlardaki puan dağılımlarının özelliklerinin belirli bir cevaplayıcı grubu için denk olmasını sağlamaktır (Kolen & Brennan, 2004). Diğer taraftan gerçek puan eşitleme gözlenen puan eşitlemeden daha karmaşıktır. Gerçek puan eşitleme yönteminde, gerçek puanın gözlenen puan ile hatadan oluştuğu varsayılır. Bu tür eşitlemenin amacı, bir formdan elde edilen gerçek puanı diğer formdan elde edilen gerçek puana eşitlemektir.

Gözlenen puan eşitleme yöntemleri eşit yüzdelli eşitleme, doğrusal eşitleme, ortalama eşitleme ve kernel eşitleme yöntemleridir. Gerçek puan eşitleme yöntemleri ise MTK'ye dayalı eşitleme yöntemleridir. Gözlenen puan eşitleme yöntemleri alanyazında yaygın olarak kullanılır; çünkü daha basittir ve diğer yöntemlerden daha önce ortaya çıkmıştır (Kolen & Brennan, 2004). Bununla birlikte Kolen (1981) tarafından yapılan araştırmada gerçek puan eşitleme yöntemlerinin gözlenen puan eşitleme yöntemlerinden daha kararlı sonuçlar verdiği tespit edildiği için, bu araştırmada gerçek puan eşitleme yöntemleri kullanılmıştır.

Eşitleme yöntemlerine ilişkin sınıflamalardan bir diğeri de yöntemlerin parametrik ve nonparametrik olmak üzere iki kısımda ele alınmasıdır. Buna göre, gerçek puan eşitleme yöntemleri parametrik yöntemler, gözlenen puan eşitleme yöntemleri de nonparametrik yöntemler olarak ifade edilir. Parametrik MTK'nin varsayımları uygulamada her zaman karşılanmadığından, parametrik yöntemlere alternatif olarak nonparametrik yöntemler geliştirilmiştir.

Eşitleme yönteminin ve modelinin seçimi; eşitlemenin amacına, modelin doğruluğu ve uygunluğuna, cevaplayıcıların özelliklerine ve test verisine bağlıdır (Yang & Houang,

1996). Bu arařtırmada madde tepki kuramına dayalı yöntemler kullanıldıđından ayrıntılı olarak bu yöntemler üzerinde durulmaktadır.

1.5.1. Klasik Test Kuramına (KTK) Dayalı Eřitleme Yöntemleri

Klasik test kuramına (KTK) dayalı eřitleme yöntemleri gruba bađlıdır; yani bilgi ve beceri düzeyindeki farklılıklardan dolayı bir cevaplayıcı grubundan diđerine deđişebilir (Woldbeck, 1998). Aynı zamanda bu yöntemde bireyin yeteneđi kendisine uygulanan maddelere bađımlıdır. Bir birey aynı özelliđi ölçen farklı güçlükteki testlerden farklı puanlar alır. Bu durumda farklı güçlükteki testler arasında karşılařtırma yapılması mümkün olmaz. Klasik test kuramına dayalı eřitleme yöntemleri; eřit yüzdelikli (equipercentile) eřitleme, dođrusal (linear) eřitleme ve ortalama (mean) eřitleme olmak üzere üçe ayrılır.

1.5.1.1. Eřit Yüzdelikli (Equipercentile) Eřitleme

Eřit yüzdelikli eřitleme yöntemi, bir dađılımda aynı yüzdelik sıraya denk gelen puanların belirlenmesine dayalı bir yöntemdir. Angoff (1982)'a göre, X ve Y testlerinde yüzdelik sırası iki grupta da eřit olan puanlar denk olarak kabul edilebilir. Eřit yüzdelikli eřitlemede, eřitlenecek puanların ilk olarak yığılmalı frekansları belirlenir ve aynı yüzdelik sıraya gelen puanlar eřit kabul edilir; ancak dađılımın düzensizliđinden dolayı, aynı yüzdelik sıraya sahip puanlar genellikle eřit olmaz. Bu problemi gidermek için dađılımı sürekli hâle getirmek gerekir. Bunun için düzgünleřtirme (smoothing) yöntemleri kullanılabilir. Eřit yüzdelikli eřitleme, dönüřtürülmüř puan dađılımlarının aynı olmasını sađladıđı hâlde ham puan kullanıldıđında eřitleme için gerekli olan Őartları karşılamamaktadır. Bu nedenle dođrusal olmayan dönüřtürmeye ihtiyaç duyulur. Bunun sonucunda da hangi testin alındıđı cevaplayıcılar için farklılık yaratacaktır. Bu nedenle, bu yöntemde eřitlik kořulu karşılanmaz. Bu yöntem kullanılarak ham puanların eřitlenmesindeki bir diđer problem de eřitleme sürecinin gruba bađımlı olmasıdır (Hambleton & Swaminathan, 1985).

1.5.1.2. Doğrusal (Linear) Eşitleme

Bu yöntem standart puanlara dayalı bir yöntemdir. Doğrusal eşitlemede, aynı standart puanlara karşılık gelen puanların eşit olduğu kabul edilir (Angoff, 1971). Doğrusal eşitlemede ortalama ve standart sapma açısından farklılık gösteren iki test eşitlenir. Bu yöntem, ortalama ve standart sapmalar hariç ham puan dağılımları eşit olduğu zaman uygun bir eşitlemedir. Diğer durumlarda eşit yüzdelikli eşitleme çok daha uygun olabilir (Crocker & Algina, 1986).

1.5.1.3. Ortalama (Mean) Eşitleme

Ortalama eşitlemede, bir formun ortalaması diğer formun ortalamasına denk olacak şekilde dönüştürülür. Ortalama eşitlemede, eşitlenecek olan testlerin güçlükleri arasında fark olduğu fakat bu farkın puan ölçeği boyunca sabit olduğu kabul edilmektedir. Bu yöntemde eşitlenecek olan testlerdeki puanların ortalamaya olan uzaklıkları eşit kabul edilir. Küçük örneklem için, ortalama eşitleme diğer yöntemlere nazaran daha güçlü bir yöntemdir (Kolen & Brennan, 1995, 2004; Barnard, 1996; Felan, 2002).

KTK'ye dayalı üç eşitleme yöntemi birbiriyle karşılaştırıldığında, eğer test formları aynı standart sapmaya sahipse ortalama eşitleme ve doğrusal eşitleme aynı sonuçları verir. Bununla birlikte eşit yüzdelikli eşitleme çoğunlukla daha büyük örneklem ve çok daha karmaşık hesaplamalar gerektirir (Zhu, 1998). Testlerin dağılımları çok benzer olduğunda, eşit yüzdelikli eşitleme kullanmak uygun bir yöntemdir. Eşit yüzdelikli eşitlemenin diğer bir avantajı, varsayımların sayısının doğrusal eşitlemeden daha az olmasıdır.

1.5.2. Madde Tepki Kuramına Dayalı Eşitleme

Madde tepki kuramı, eğitimde ve psikolojide kullanılan testler için pek çok yararlı uygulamaya sahip güçlü bir yöntemdir. MTK, “örtük özellikler” veya “madde karakteristik eğrisi” teorisi olarak da bilinir. Bu teori, farklı yetenek düzeyindeki

cevaplayıcıların ölçülen bir özellik için bir maddeye verebilecekleri cevabı tanımlayan matematiksel bir model sunar (Cook & Eignor, 1991).

Madde tepki kuramı, testleri eşitleme ve ölçeklemede yaygın olarak kullanılan bir yöntemdir. MTK'nin ortaya çıkışından itibaren test eşitlemeye duyulan ilgi gittikçe artmaktadır. MTK'ye dayalı eşitlemede, iki test formundaki puanlar arasında matematiksel bir ilişki modellenir. Bu ilişki ortak bir ölçeğe yerleştirilen her bir formdan elde edilen madde parametre kestirimlerine dayalıdır (Norman-Dvorak, 2009; Skaggs & Lissitz, 1986; Dongyang, 2009). MTK'ye dayalı eşitleme yöntemleri, cevaplayıcıların yetenekleri ve maddeyi doğru cevaplama olasılıkları arasındaki ilişkiyi tanımlayan matematiksel bir fonksiyonun olduğu varsayımını temel alır (Dorans, 1990; Hambleton & Swaminathan, 1985; Kolen & Brennan, 1995, 2004).

Geleneksel yöntemlerle eşitleme yapıldığında eşitlik, simetri ve değişmezlik özellikleri sağlanamadığından ötürü sonuçlar çok anlamlı olmamaktadır; ancak model veri uyumu sağlanmış ise madde tepki kuramı ile eşitleme yapmak bu sorunları çözümlenmektedir (Kolen, 1981). Bilindiği üzere, bu kurama göre θ yetenek düzeyi sabitken bireylerin maddelere verdiği cevaplar bağımsızdır. Bu nedenle, madde parametreleri bilindiğinde, bir kişinin θ yetenek düzeyi maddelere verdiği cevaplardan etkilenmez. Bunun sonucunda da bir cevaplayıcının zor ya da kolay bir test alması bir problem teşkil etmez ve bireylerin yetenek kestirimleri örnekleme bağı olarak karşılaştırılabilir. Böylece MTK çerçevesinden bakıldığında, madde parametreleri bilinen testlerle uygulama yapıldığında, test puanlarının eşitlenmesine gerek duyulmaz. Bu hem yatay hem de dikey eşitleme için geçerlidir (Hambleton & Swaminathan, 1985).

Birçok psikometrist tarafından belirtildiği üzere, puanların yetenek metriği üzerinde raporlanması, yorumlanmasını güçleştirebilir. Bu tür durumlarda ölçeklenmiş puanlar kullanılabilir. Alternatif olarak, yetenek puanları test puan metriğine dönüştürülebilir. Özetlemek gerekirse, MTK bağlamında eğer madde parametreleri biliniyor ise eşitleme ihtiyacı oluşmaz. Madde ve yetenek parametreleri bilinmiyor ise madde tepki fonksiyonu yetenek ya da madde parametrelerinin doğrusal bir dönüşümüne bağlı olarak değişmez. Bu nedenle, hem yetenek parametresi hem de madde güçlük parametresi için keyfi bir metrik seçmek gerekir. Madde tepki kuramı modellerinin

değişmezlik özelliği söz konusudur. Bir başka deyişle, cevaplayıcıların yetenek düzeyleri maddelerden, madde parametreleri de cevaplayıcı grubundan etkilenmez (Hambleton ve diğerleri, 1991; Han, 2008; Zhu, 1998) fakat farklı iki gruba uygulanmış maddelerin madde parametreleri farklı gözükabilir. Bu görüntü uyumsuzluğu ise θ 'ya ilişkin metriğin keyfi olarak sabitlenmesinden kaynaklanmaktadır. Oysaki iki gruba ilişkin madde parametreleri ve yetenek parametreleri arasında doğrusal bir ilişki mevcuttur (Sukin & Keller, 2008). Eşitleme yapılabilmesi için parametrelere ilişkin metrikler şu şekilde sabitlenebilir: Tek grup deseninde, her bir test için yetenek parametrelerine ilişkin metrik sabitlenir. Bu durumda eşitlemeye gerek kalmaz. Ortak test deseninde, her bir gruptaki yetenek için metrik sabitlenir. Random grup deseninde ise, yetenek ya da madde parametrelerine ilişkin metrik sabitlenir (Hambleton & Swaminathan, 1985).

MTK ile eşitleme, uygun eşitleme deseninin seçilmesiyle başlar. MTK ile eşitlemede, verilerin elde edildiği grupların yetenek düzeyleri genellikle birbirinden farklı olduğu için en çok ortak test deseni kullanılır. Bir sonraki basamak, uygun MTK modeline karar vermektir. En çok kullanılan MTK modelleri iki kategorili puanlanan testler için bir, iki ve üç parametrelili lojistik modellerdir. Model veri uyumunu test etmek için uyum ölçüleri kullanılabilir. Parametrik model varsayımlarının sağlanmadığı durumlarda ise nonparametrik eşitleme yöntemlerinin tercih edilmesi uygun olacaktır. Seçilen MTK modeli ile madde ve yetenek parametreleri kestirilir, madde ve yetenek parametreleri için ortak bir metrik oluşturulur. Son olarak ise test puanlarının raporlanacağı ölçeğe karar verilir. Test puanları θ yetenek cinsinden raporlanıyor ise işlem bitirilir. Kestirilen gerçek puanlar cinsinden rapor ediliyor ise gerçek puanlar farklı yetenek düzeylerine göre kestirilmeli ve grafikleştirilmelidir. Buradan iki teste ilişkin gerçek puanlar eşitlenebilir. Gözlenen puanlar eşitlenecek ise teorik olarak, koşullu gözlenen puan dağılımları seçilen cevaplayıcı örneklemeine bağlı yetenekler için gerçekleştirilir. Marjinal gözlenen puan dağılımları hesaplanır. Elde edilen tablo ya da grafikten gerçek gözlenen puanlar eşitlenir (Felan, 2002; Hambleton & Swaminathan, 1985; Zhu, 1998; Cook & Eignor, 1991). Özetlemek gerekirse, kullanılan eşitleme yönteminin türüne bağlı olmaksızın eşitleme çalışmaları iki kısımda ele alınır. Bunlar, eşitleme deseni ve uygun ölçek dönüşümünü gerçekleştirmek için kullanılan yöntemlerdir.

MTK yöntemleri, test eşitleme yöntemlerinin önemli bir parçasını oluşturur (Kolen & Brennan, 2004). MTK uygulamaları pek çok istatistiksel varsayım üzerine kuruludur. Temel varsayımlardan biri tek boyutluluk, diğeri ise yerel bağımsızlıktır. Tek boyutluluk, bir testte yer alan maddelerin sadece bir örtük özelliği (θ) ölçmesi, yerel bağımsızlık ise bir cevaplayıcının bir maddeye verdiği cevabın diğer maddelere verdiği cevaptan bağımsız olmasıdır (Hambleton ve diğerleri, 1991). Üçüncü temel varsayım ise her bir maddeye doğru cevap verme olasılığının normal ogive veya lojistik fonksiyon gibi parametrik fonksiyonlarla açıklanabilmesidir. Bu ilişkiyi açıklamada kullanılan pek çok model vardır.

MTK, karmaşık psikometrik problemlere oldukça iyi çözümler getirmesine rağmen, çözümün doğruluğu bu istatistiksel varsayımların ne kadar karşılandığına bağlıdır. Bazı varsayımlar ciddi ölçüde ihlal edilirse, parametrik MTK kullanmak önemli düzeyde yanlışlıklar ortaya çıkarabilir ve yanlış kararlara sebep olabilir. İstatistiksel varsayımların ihlalden sakınmak için iki alternatif yaklaşım vardır. İlk yaklaşım, çok boyutlu MTK modellerini kullanmaktır (McDonald, 1997; Reckase, 1997). Çok boyutlu MTK, birden fazla örtük değişkeni ele alarak tek boyutluluk ihlalden kaçınır. İkinci yaklaşım ise nonparametrik MTK modellerini kullanmaktır (Mokken, 1971; Ramsay, 1991).

Madde tepki kuramına dayalı eşitleme, önceleri parametrik modeller için geliştirilirken, son zamanlarda nonparametrik modeller için de kullanılmaya başlanmıştır. Bu kapsamda MTK'ye dayalı eşitleme, yukarıda da ifade edildiği gibi varsayımların karşılanıp karşılanmamasına bağlı olarak parametrik ve nonparametrik yöntemler olmak üzere iki kısımda incelenir. Ancak, nonparametrik yöntemler karşısındaki tartışmalar, parametrik modellerin nonparametrik modellerden çok daha kararlı sonuçlar verdiği yönündedir (Lei, Dunbar & Kolen, 2004) ve parametrik modeller çok daha güçlü varsayımlara sahiptir (Ansley & Forsyth, 1985). Bu nedenle, çalışmada sadece parametrik MTK'ye dayalı eşitleme yöntemleri kullanılmış ve ayrıntılı olarak bu yöntemler üzerinde durulmuştur.

MTK'ye dayalı eşitleme yöntemleri iki kategorili ve çok kategorili puanlanan maddeler için geliştirilmiştir. MTK modellerinin çoğu iç içe geçmiştir fakat her bir modelin çalışma biçimi ve bu modellere dayalı yöntemler birbirinden farklıdır (de Koning,

Sijtsma & Hamers, 2002). Bu arařtırmada iki kategorili puanlanan maddeler için geliřtirilen MTK yöntemleri kullanıldıđından, sadece bu yöntemlerden bahsedilmiřtir.

1.5.3. Madde Tepki Kuramına Dayalı Eřitleme Yöntemleri

Klasik test kuramına dayalı yöntemlerin tersine MTK modelleri, test formlarını eřitlemede her bir madde için non-lineer eğrileri kullanır (Lord, 1980). Bu eğriler “madde karakteristik eğrileri” olarak adlandırılır. Madde karakteristik eğrisi, belirli bir yetenek düzeyindeki cevaplayıcının beklenen test puanını belirlemede kullanılır ve maddenin doğru cevaplanma olasılığı ile yetenek düzeyi arasındaki ilişkiyi gösterir. Madde karakteristik eğrisinin hangi matematiksel fonksiyonla gösterileceğini belirlemek için farklı modeller ortaya çıkmıřtır. Bu modeller; Rasch modeli, iki parametrelili ve üç parametrelili modeller olmak üzere üçe ayrılır (Skaggs & Lissitz, 1986; Kolen & Brennan, 1995, 2004; Keller III, 2007; Woldbeck, 1998).

1.5.3.1. Rasch Modeli

En basit MTK modeli Rasch modelidir. Bu model “bir parametrelili lojistik model (1PLM)” olarak da bilinir. Rasch modelinde bağımlı deđiřken, bireyin belirli bir maddeyi doğru cevaplama olasılığıdır. Bağımsız deđiřkenler ise bireyin yetenek düzeyi (θ) ve madde güçlük düzeyi (b_i)’dir. Güçlük parametresinin ve yetenek düzeyinin uygulamada -3 ile 3 arasında deđerler aldığı görölmektedir. Bu model, kestirilecek madde parametresinin sadece bir tane olduđu modeldir. Rasch modeli, řans başarısının 0 olması ve her maddenin eřit derecede ayırıcılık gücüne sahip olması řeklinde iki ek varsayıma sahiptir. Bu model ile θ yetenek düzeyindeki bir cevaplayıcının maddeyi doğru cevaplama olasılığı řu řekilde ifade edilir:

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

$P_i\theta$: θ yeteneđindeki bir bireyin i maddesini doğru olarak cevaplama olasılığı

b_i : i maddesinin güçlük indeksi

D: ölçek sabiti (1,7)

Rasch modeli en az örneklem büyüklüğü gereksinimine sahip model olduğu için küçük örneklemelerde de kullanılabilir (Kolen & Brennan, 1995; 2004).

1.5.3.2. İki Parametrelili Lojistik Model

İki parametrelili lojistik modelde (2PLM), modele ayırt edicilik parametresi de dâhil edilmiştir. Bu parametre genel olarak 0 ile 2 arasında değerler alabilir (Hambleton ve diğerleri, 1991). Yetenek düzeyi ile madde güçlüğü arasındaki fark, yüksek ayırt ediciliğe sahip maddelerin cevaplanma olasılığını daha fazla etkilemektedir. İki parametrelili model aşağıdaki eşitlik yardımıyla kestirilir:

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

$P_i\theta$: θ yeteneğindeki bir bireyin i maddesini doğru olarak cevaplama olasılığı

b_i : i maddesinin güçlük indeksi

a_i : i maddesinin ayırtıcılık gücü indeksi

D: ölçek sabiti (1,7)

Bir birey aynı toplam puana sahip iki madde açısından, farklı tahminî yetenek düzeyine sahip olabilir. Yüksek ayırt ediciliğe sahip bir maddedeki başarı ve düşük ayırt ediciliğe sahip bir maddedeki başarısızlık en yüksek yetenek düzeyi tahminini verebilir. Bu nedenle, iki parametrelili modeldeki maddeler yetenek düzeylerinin tahmininde eşit ağırlığa sahip değildirler (Embretson & Reise, 2000).

1.5.3.3. Üç Parametrelili Lojistik Model

Üç parametrelili model (3PLM), madde karakteristik eğrisini tanımlayan fonksiyona ek olarak bir de şans başarısı parametresinin eklenmesiyle oluşturulmuş bir modeldir. Üç parametrelili lojistik modelin matematiksel ifadesi aşağıdaki gibidir:

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

$P_i\theta$: θ yeteneğindeki bir bireyin i maddesini doğru olarak cevaplama olasılığı

b_i : i maddesinin güçlük indeksi

a_i : i maddesinin ayırtıcılık gücü indeksi

c_i i maddesinin şans parametresi

D: ölçek sabiti (1,7)

Yukarıdaki eşitlikte verilen tüm parametreler iki parametrelili lojistik modelde olduğu gibidir ancak bu modelde fazladan c_i parametresi vardır. Bu parametreye “pseudo (şans düzeyi) parametresi” denir. Bu parametre madde karakteristik eğrisinde sıfır olmayan düşük asimptotunu gösterir ve bireyin bir maddeyi doğru olarak cevaplamasının en düşük olasılığına karşılık gelir. Bu parametre şansla doğru cevabı bulmanın mümkün olduğu çoktan seçmeli testlerde önem taşımaktadır ve 0 ile 1 arasında değerler alabilir (Hambleton & Swaminathan, 1985). Kuramsal olarak elde edilen veriler için ve önemli kararlara dayanak teşkil edecek test uygulamalarında üç parametrelili model; bir ve iki parametrelili modelden daha uyumlu sonuçlar verir (Keller III, 2007). Harris (1993) üç parametrelili model için büyük örneklemelere gereksinim duyulduğunu ifade etmiştir (Form başına yaklaşık olarak 1500 kişi). Kolen (1981), Slinde ve Linn (1979) yaptıkları araştırmalarda güçlükleri farklı olan testleri eşitlemede üç parametrelili modelin Rasch modeline göre daha tutarlı sonuçlar verdiğini bulmuşlardır.

1.5.4. Madde Tepki Kuramına Dayalı Ölçek Dönüştürme (Kalibrasyon) Yöntemleri

Uygun olan MTK yöntemi seçildikten sonra, öncelikle MTK parametrelerinin kestirilmesine ihtiyaç duyulur. Yetenek kestiriminin bir önemli özelliği, cevaplayıcıların doğru olarak cevapladığı madde sayısından ziyade (Rasch modeli dışında) madde cevaplarının örüntüsüne bağlı olmasıdır. Bir başka deyişle, eğer doğru olarak cevaplanan maddelerin bazıları farklıysa, doğru cevap sayısı (ham puanları) eşit olan cevaplayıcıların kestirilen θ 'larının da farklı olma olasılığı vardır. MTK'de böyle bir puanlamanın kullanılması yetenek kestirimlerin doğruluğunu artırır.

Tek grup veya denk grup deseni kullanılarak eşitleme yapıldığında, test formları aynı ölçekte olduğu için ek bir ölçeklemeye gereksinim duyulmaz. Denk olmayan gruplara dayalı eşitleme yönteminde ise, gruplar farklı olduğu için farklı test formlarından elde edilen parametreler aynı ölçekte olmayacaktır. Bu yüzden iki test formunu aynı ölçeğe yerleştirmek için doğrusal bir dönüştürme yapılmalıdır (Han, 2008; Kolen & Brennan, 1995). Madde parametrelerini kestirmek için kullanılan bilgisayar programları, analiz

edilen veriler için θ ölçeğini ortalamayı 0, standart sapmayı 1 olacak şekilde tanımlar. Bu durumda gruplar yetenek açısından farklı olmasına rağmen her bir grup için yetenekler ortalama 0, standart sapma 1 olacak şekilde ölçeklenir. Bunun sonucunda da MTK ölçeğinin dönüştürülmesine ihtiyaç duyulur. Böylece doğrusal bir eşitleme MTK parametre kestirimlerini aynı ölçeğe dönüştürmede kullanılabilir. Dönüştürmeden sonra ortak bir ölçek üzerinde yer alan iki grubun yetenek puanlarının ortalama ve standart sapmalarının farklılık göstermesi beklenir. Bazen “kalibre edilmiş/kestirilmiş” olarak da ifade edilen dönüştürülmüş parametre kestirimleri form X ve form Y’den elde edilen ham puanlar arasındaki denklikleri oluşturmada ve daha sonra da puanları ölçeklemede kullanılabilir (Kolen & Brennan, 1995, 2004). Farklı gruplardan elde edilen madde parametrelerini aynı ölçeğe dönüştürmenin “ayrı kalibrasyon/kestirim (separate calibration/estimation)” ve “eş zamanlı kalibrasyon/kestirim (concurrent calibration/estimation)” olmak üzere iki yolu vardır (Kolen & Brennan, 1995, 2004; Nozawa, 2008; Cao, 2008).

1.5.4.1. Ayır Kalibrasyon/Kestirim (Separate Calibration/Estimation)

Bu yöntem kullanıldığında denk olmayan gruplardan elde edilen madde parametreleri lineer olarak ilişkili fakat farklı bir ölçekte olacaktır (Hambleton & Murray, 1983; Akt. Hu, Rogers, Vukmiroviç, 2008). Bu yüzden eşitleme yapmadan önce madde parametrelerini aynı ölçeğe yerleştirmek gereklidir. Bu amaçla denk olmayan gruplarda ortak madde deseninde yer alan ortak maddelerin madde parametreleri ölçek dönüştürmede kullanılır. Sonuç olarak, iki denk olmayan grubun ortak madde parametrelerinin doğrusal bir ilişkiye sahip olması beklenir (Hu ve diğerleri, 2008). Bu yöntemde iki farklı test formu için parametreler ayrı ayrı kestirilir. Kestirilen farklı gruplara ait parametreler MTK’deki ölçek belirsizliğinden dolayı eşit olmayabilir. Elde edilen parametreler farklı ölçekte olduğundan karşılaştırılmaz. Bu karşılaştırmayı sağlamak amacıyla, ortak maddelerin a ve b parametrelerine dayalı olarak elde edilen A ve B sabitleri, bir formdaki θ değerini diğer forma dönüştürür. Ayır kalibrasyon kullanıldığında A ve B sabitleri için, Test I’daki i kişisi için θ değerinin test J’deki karşılığı şu şekilde ifade edilir:

$$\theta_{ji} = A\theta_{li} + B$$

θ_{ji} ve θ_{li} : i kişisinin J ve I testine ait yetenek düzeyi

A: Eşitleme denkleminin eğimi

B: Eşitleme denkleminin sabiti.

Benzer şekilde iki testin madde parametreleri de dönüştürülür. Test I ve Test J'de yer alan madde parametrelerinin dönüşümü şu şekilde ifade edilir:

$$a_{Jj} = \frac{a_{Ij}}{A}$$

$$b_{Jj} = Ab_{Ij} + B$$

$$c_{Jj} = c_{Ij}$$

a_{Ij} , b_{Ij} ve c_{Ij} : j maddesi için I testinden (referans formdan) elde edilen madde parametreleri

a_{Jj} , b_{Jj} ve c_{Jj} : j maddesi için J testinden elde edilen sırasıyla madde ayırtıcılık, madde güçlük ve şans parametreleri (Dönüştürülmüş madde parametreleri)

A: Eşitleme denkleminin eğimi

B: Eşitleme denkleminin sabiti

Yukarıdaki eşitlikte de görüldüğü üzere, c parametresi ölçek dönüşümünden bağımsızdır (Kolen & Brennan, 2004).

Bu kestirim sürecinde madde parametreleri iki farklı cevaplayıcı grubuna verilen X ve Y formları için kestirildiğinde, yetenek parametreleri her iki grup için ortalaması 0, standart sapması 1 olacak şekilde ölçeklenir (Li, 2009). İki kategorili MTK modelleri için ortak madde deseninde ayrı kalibrasyonda ölçek dönüşümü için yaygın olarak kullanılan yöntemler şunlardır:

1. Moment yöntemleri
 - a. Ortalama-ortalama (mean-mean) yöntemi (Loyd & Hoover, 1980),
 - b. Ortalama-standart sapma (mean-sigma) yöntemi (Marco, 1977)
2. Karakteristik eğri dönüştürme (characteristic curve transformation) yöntemleri (Haebara, 1980; Stocking & Lord, 1983).

1.5.4.1.1. Ortalama-Ortalama ve Ortalama-Standart Sapma Yöntemleri

Ortak madde deseninde ölçek dönüştürmenin en basit yolu, ortak maddelerin madde parametrelerinin ortalaması ve standart sapmasını kullanmaktır. Dönüştürmeden sonra madde parametre kestirimleri “kalibre edilmiş” olarak adlandırılır. Her iki yöntem de ölçekleme sabitlerini hesaplamada kullanılır. Marco (1977) tarafından tanımlanan ortalama-standart sapma yönteminde, A ve B eşitleme katsayılarını kestirmek için ortak maddelerden elde edilen madde güçlük parametrelerinin ortalaması ve standart sapması kullanılır. A ve B katsayılarını kestirmek için kullanılan eşitlikler aşağıda verilmiştir:

$$A = \frac{\sigma(b_j)}{\sigma(b_I)} \quad B = \mu(b_j) - A\mu(b_I)$$

$\sigma(b_I)$: I testindeki ortak maddelerden elde edilen b parametrelerine ait standart sapma

$\sigma(b_j)$: J testindeki ortak maddelerden elde edilen b parametrelerine ait standart sapma

$\mu(b_j)$: J testindeki ortak maddelerden elde edilen b parametrelerine ait ortalama

$\mu(b_I)$: I testindeki ortak maddelerden elde edilen b parametrelerine ait ortalama

Loyd ve Hoover (1980) tarafından tanımlanan ortalama-ortalama yönteminde ise A ve B eşitleme katsayılarını kestirmek için ortak maddelerden elde edilen madde ayırıcılık parametrelerinin ortalaması kullanılır. Bu yöntem için katsayılar şu şekilde hesaplanır:

$$A = \frac{\mu(a_I)}{\mu(a_j)}$$

$\mu(a_I)$: I testindeki ortak maddelerden elde edilen a parametrelerine ait ortalama

$\mu(a_j)$: J testindeki ortak maddelerden elde edilen a parametrelerine ait ortalama.

Yukarıda da ifade edildiği gibi iki yöntem, kestirim için farklı parametreleri kullandığından farklı sonuçlar ortaya koyar. Bazen ortalama-standart sapma yöntemi ortalama-ortalama yöntemine tercih edilir; çünkü b parametresinin kestirimi a parametresinin kestiriminden daha karardır. Bununla birlikte Baker ve Al-karni (1991) ve Ogasawara (2000) ortalama-ortalama yönteminin daha tercih edilebilir ve daha kararlı olduğunu çünkü ortalamaların standart sapmalardan daha kararlı olduğunu vurgulamışlardır. Diğer taraftan da Kolen ve Brennan (2004) ortalama-standart sapma yönteminin bazı durumlarda daha tercih edilebilir olduğunu belirtmektedirler. Bu

nedenle, alanyazında her iki yöntemle eşitleme yapılması ve iki yöntemden elde edilen ham/ölçek puanı dönüşümlerinin karşılaştırılması önerilmektedir.

1.5.4.1.2. Karakteristik Eğri Dönüştürme Yöntemleri

Ortalama-ortalama ve ortalama-standart sapma yöntemleri, madde karakteristik eğrileri benzer ancak parametreleri farklı olan maddeler için hatalı sonuçlara yol açabilir (Norman-Dvorak, 2009). I ve J testlerindeki kestirimlerde b parametreleri arasındaki farkın büyük ancak madde karakteristik eğrilerinin benzer olduğu bir madde için, ortalama-standart sapma yöntemi, b parametresi kestirimleri arasındaki farklılıktan etkilenir. Bu problem, ölçek dönüştürme yöntemleri madde parametre kestirimlerinin tümünü aynı anda dikkate almadığı için oluşur. Bu probleme cevap olarak Haebara (1980), madde parametrelerinin tümünü aynı anda göz önüne alan bir yöntem ortaya atmıştır. Daha sonra Stocking ve Lord (1983), Haebara'nın yöntemine benzer bir yöntem geliştirmiştir.

Bu yöntemler, test karakteristik eğrilerine dayalı eşitleme katsayılarını belirleyen gerçek puan eşitleme yöntemleridir (Baker & Al-Karni, 1991). Karakteristik eğri dönüştürme yöntemleri, ortak maddelerin madde ya da test karakteristik eğrileri arasındaki farkı azaltmak için geliştirilmiştir ve sadece NEAT deseniyle toplanan verilere uygulanabilir. Madde karakteristik eğrileri arasındaki farkı açıklamak için Haebara (1980) tarafından kullanılan fonksiyon, belirli bir yetenekteki cevaplayıcılar için her bir maddenin madde karakteristik eğrileri arasındaki farkın karelerinin toplamıdır. Stocking ve Lord (1983) tarafından kullanılan kayıp fonksiyon ise belirli bir yetenekteki cevaplayıcılar için her bir maddenin madde karakteristik eğrileri arasındaki farkın toplamının karesidir. İki yöntem arasındaki tek fark, kayıp fonksiyonun hesaplanmasında ortaya çıkar. İki yöntem genellikle benzer kestirimler sağlar ve özellikle madde ayırt ediciliklerinin dönüştürülmesinde daha iyi sonuçlar verir.

Haebara (1980) yöntemi ortak maddelerin madde karakteristik eğrileri arasındaki farkı azaltır ve bu yöntemde kayıp fonksiyonun matematiksel ifadesi şu şekildedir:

$$L(\theta_i) = \sum_{j=1}^m \left[p_{ij}(\theta_i, a_{1Lj}, b_{1Lj}, c_{1Lj}) - p_{ij}(\theta_i, a_{2Lj}^*, b_{2Lj}^*, c_{2Lj}) \right]^2$$

p_{ij} : j. cevaplayıcı ve i. madde için madde karakteristik fonksiyonu (j.cevaplayıcının i.maddeyi doğru cevaplama olasılığı)

$a_{1Lj}, b_{1Lj}, c_{1Lj}$: form 1'deki j. ortak madde için sırasıyla madde ayırıcılık, güçlük ve şans parametreleri

$a_{2Lj}, b_{2Lj}, c_{2Lj}$: form 2'deki j. ortak madde için sırasıyla madde ayırıcılık, güçlük ve şans parametreleri

m: ortak maddelerin sayısı

$$a_{2Lj}^* = \frac{a_{2Lj}}{A}$$

$$b_{2Lj}^* = Ab_{2Lj} + B$$

Yukarıdaki eşitlikler yardımıyla form 2'deki ortak maddeler yeniden ölçeklenir ve daha sonra kayıp fonksiyon eşitliğinde kullanılır. Kayıp fonksiyon tüm cevaplayıcılar için hesaplanır.

Stocking ve Lord (1983) yöntemi ortak maddelerin madde karakteristik eğrilerinin toplamı arasındaki farkı azaltır. Yaygın olarak kullanılan eşitleme yöntemlerden biri olan Stocking ve Lord karakteristik eğri yönteminde kayıp fonksiyon şu şekilde hesaplanır:

$$L(\theta_i) = \left[\sum_{j=1}^m p_{ij}(\theta_i, a_{X1j}, b_{X1j}, c_{X1j}) - \sum_{j=1}^m p_{ij}(\theta_i, a_{Y2j}^*, b_{Y2j}^*, c_{Y2j}) \right]^2$$

Bu alanda yapılan çalışmalar, karakteristik eğri yöntemlerinin ortalama-ortalama ve ortalama-standart sapma yöntemlerinden daha iyi olduğunu ve daha kararlı sonuçlar üretme eğiliminde olduğunu ortaya koymuştur (Baker & Al-Karni, 1991; Way & Tang, 1991; Hanson & Beguin, 2002; Kolen & Brennan, 2004). Bununla birlikte karakteristik eğri yöntemleri, hesaplama bakımından oldukça karmaşık olduğu ve ortalama-ortalama ve ortalama-standart sapma yöntemlerinden hesaplanan ölçekleme katsayılarındaki uygunluk daha iyi olduğu için ortalama-ortalama, ortalama-standart sapma yöntemleri hâlâ yaygın olarak kullanılmaktadır.

1.5.4.2. Eş zamanlı Kalibrasyon/Kestirim (Concurrent Calibration/Estimation)

Aynı ölçek üzerine maddeleri yerleştirmek için kullanılan diğer bir yöntem, eş zamanlı kalibrasyondur (Chu & Kamata, 2000). Bu yöntem iki formdan elde edilen madde parametrelerini birlikte kestirir. Ortak maddelerin iki test formunda aynı madde parametrelerine sahip olduğu varsayılır. Yetenek dağılımındaki farklılıklar göz önüne alındığından kestirilen madde parametreleri aynı ölçek üzerindedir (Nozawa, 2008; Turhan, 2006). Bu yüzden bu yöntemde fazladan bir dönüştürmeye ve A, B sabitlerinin elde edilmesine ihtiyaç yoktur. Hanson ve Beguin (2002) parametrik model varsayımları karşılandığında eş zamanlı kestirimin ayrı kestirimden daha doğru sonuçlar verdiğini ortaya koymuşlardır.

Yapılan bu araştırmada sadece ayrı kalibrasyon yöntemi kullanılmıştır. Bu yöntem ile iki farklı madde parametresi kestirilmekte ve böylece olası madde problemlerini tanımlamaya yardımcı olmaktadır (Hanson & Beguin, 2002; Kolen & Brennan, 2004). Ayrıca eş zamanlı kestirim yöntemi ayrı kestirim yöntemlerine göre program kullanımı açısından daha fazla yük getirir ve performans açısından da bazı problemlere neden olabilir. Bu nedenle bu araştırma, değişik koşullar altında ayrı kestirim yöntemlerinin performansını ortaya çıkarmaya odaklanmıştır.

Bu araştırmada ele alınan eşitleme yöntemlerinin kendilerine özgü avantaj ve dezavantajları olmakla birlikte, araştırmacılar yaptıkları çalışmalarda hem veri analizi için hem de sonuçları karşılaştırmak için farklı türdeki eşitleme yöntemlerini bir arada kullanırlar. Bu kapsamda araştırmada, MTK'ye dayalı eşitleme yöntemleri kullanılarak bu yöntemlerden elde edilen sonuçlar karşılaştırılmıştır. Bu konuda yurt dışında yapılmış araştırmalara rastlanmakla birlikte, Türkiye'de bu yöntemleri karşılaştıran araştırma sayısı yok denecek kadar azdır. Karşılaştırmalar yapılırken, Türkiye'de yapılan büyük ölçekli sınavların koşulları da dikkate alınmıştır. Bu yönüyle de araştırmanın, Türkiye'deki sınavlar üzerinde yapılacak eşitlemeler için hangi yöntemlerin daha uygun olacağını belirlenmesi konusunda uzmanlara bilgi vermesi beklenmektedir.

1.6. EŞİTLEME HATASI

Eşitleme hataları, eşitleme doğruluğunu tanımlamak için kullanılır. Hataların miktarı, eşitleme yöntemine ve desenine bağlıdır. Random ve sistematik olmak üzere iki tür eşitleme hatası vardır (Kolen, 1988; Kolen & Brennan, 1995; Felan, 2002). Random eşitleme hatası, cevaplayıcı örnekleminde kaynaklanan bir hatadır. Bir başka deyişle, örneklem büyüklüğü random eşitleme hatasının önemli bir kaynağıdır. Random hata, eşitlemenin standart hatasıyla tanımlanır. Kavramsal olarak eşitlemenin standart hatası, eşitlenmiş puanların standart sapmasıdır. Örneklem büyüklüğü arttıkça, eşitlemenin standart hatası küçülür ve çok büyük örneklem için önemsiz hâle gelir. Random hata, büyük örneklem kullanılarak, uygun bir eşitleme deseni seçerek veya her ikisi de kullanılarak kontrol edilebilir (Kolen & Brennan, 1995). Bu hata türünde, cevaplayıcı örneklemini ortalama, standart sapma, yüzdelik sıra gibi parametreleri kestirmek için kullanıldığında fırsata dönüşür (Barnard, 1996).

Diğer bir eşitleme hatası ise, sistematik hatadır. Eşitleme yanlılığı olarak da ifade edilen sistematik hata, belirli eşitleme yöntemlerini kullanmak için gerekli koşulların, varsayımların ihlal edilmesinden ve yanlılıktan kaynaklanır (Zeng, 1991). Random hataya göre, sistematik hatayı kontrol etmek daha zordur (Kolen & Brennan, 1995). Sistematik hata özellikle denk olmayan gruplarda sorundur ve kullanılan yöntemin altında yatan varsayımlar karşılanmadığında oluşur (Zeng, 1991). Eğer gruplar büyük ölçüde farklıysa veya ortak maddeler içerik ve istatistiksel özellikler bakımından toplam test formunun bir temsilcisi değilse varsayımların karşılanması zordur. Bunun dışında eğer ortak maddeler bir uygulamadan diğerine farklılaşıyorsa sistematik hata oluşur (Felan, 2002). Sistematik hata, dikkatli bir test geliştirme, eşitleme desenlerinin yeterli şekilde uygulanması ve uygun istatistiksel tekniklerin kullanılmasıyla kontrol edilebilir. Toplam hata ise random hata ve sistematik hatanın toplamıyla tanımlanır (Kolen & Brennan, 2004).

Random hatanın miktarı eşitlemenin standart hatası kullanılarak belirlenebilmesine rağmen, sistematik hatanın miktarını belirlemek çok daha zordur. Eşitleme çalışmalarını tasarlama ve yürütmede, her iki hata türü mümkün olduğunca azaltılmalıdır. Eşitleme hatasının miktarı bazı uygulamalarda çok büyük olabilir. Random veya sistematik

hatayı kontrol etmenin yolu, yeterli örneklem büyüklüğü kullanmaktır. Buna rağmen eşitleme çalışmasının yürütüldüğü örneklemin büyüklüğünün artması random hatayı azaltırken sistematik hatayı azaltmamaktadır (Kolen & Brennan, 1995). Michaelides ve Haertel (2004)'e göre, ortak madde örnekleminde kaynaklanan hata örneklem büyüklüğüne değil, ortak maddelerin sayısına bağlıdır. Bu yüzden de ortak maddelerin sayısı, hataların en önemli kaynağını oluşturur (Monseur, Sibberns & Hastedt, 2006).

Sonuç olarak, eşitlemede hata istenmeyen etkilere ve gerçek performansından büyük ölçüde farklılık gösteren yetenek kestirimine yol açabilir. Bu yüzden eşitleme hatasını azaltmak için eşitleme desenleri ve eşitleme yöntemleri dikkatli bir biçimde seçilmelidir. Test puanları genellikle, büyük ölçekli sınavlar sonucunda verilen kararlarla ilişkilidir. Test puanlarının ve bu puanlar sonucunda verilen kararların doğruluğu, testlerden elde edilen puanların karşılaştırılmasından emin olmayı gerektirir. Bu şekilde hatasız ve doğru bir şekilde yapılan eşitleme, ölçme ve değerlendirmenin kalitesini de artırır. Bu yüzden hatanın olası kaynağı kontrol edilmeli, hatanın miktarı belirlenmeli ve her bir kestirim için rapor edilmelidir.

1.7. ARAŞTIRMANIN ÖNEMİ VE AMACI

Eğitim sisteminin amacına ulaşp ulaşmadığını görmek için öğrencilerin performanslarının bir göstergesi olarak sınavlardan aldıkları puanlar, her yıl sınıf düzeylerine ve gruplara göre karşılaştırılır. Örneğin, pek çok ülkede sınav sonuçları açıklandığında eğitimin kalitesi hakkında genellikle bir tartışma olur. Sınavı geçenlerin yüzdesi daha önceki yıl yapılan sınav sonuçlarıyla karşılaştırılır ve sonuçlar öğrencilerin önceki yıl yapılan sınav sonuçlarına göre daha iyi ya da daha kötü olup olmadığına ve puanların yüksekliğine göre değerlendirilir. Bazen sonuçlarda düşüş olduğu görüldüğünde standartlarda değişikliğe gidilir ve bu da sınav sisteminde değişikliğe yol açar. Sınav sonuçlarının bir önceki yıla göre yüksek ya da düşük olduğuna karar verebilmek farklı yıllarda alınan puanların karşılaştırılabilir olmasını gerektirir. Kişilerin geleceği hakkında verilen kararlar da, karşılaştırılabilirliği sağlanan ve geçerliği ciddi biçimde incelenen test puanlarına dayalı olarak yapılır. Herhangi bir karşılaştırma yapmadan önce, bir testin farklı yıllarındaki uygulamalardan elde edilen

puanlarının birbirleriyle ilişkisini belirlemek ve bu ilişkiye dayalı dönüşümler yapmak gereklidir. Bu yüzden de puanlar arasında bir ilişki tanımlanmadıkça farklı testlerden elde edilen puanları doğrudan karşılaştırmak anlamsızdır. Bu tür ilişkiyi oluşturan istatistiksel süreç ancak eşitleme ile yerine getirilir. Bu açıdan düşünüldüğünde, test eşitleme test sonuçlarının doğru bir şekilde karşılaştırılmasını sağlar. Birbirleriyle karşılaştırılabilen ve birbirinin yerine kullanılabilen puanlar elde etmek, aynı amaçla kullanılan farklı test formlarına dayalı olarak verilen kararların adil olmasını sağlamak için gereklidir.

Testlerden elde edilen puanlar kişilere ya da kurumlara önemli bilgiler sağlar. Örneğin, bireysel açıdan bakıldığında, test sonuçları bir öğrencinin hangi üniversiteye gireceğine karar verilmesine yardımcı olur. Kurumsal düzeyde ise bireyleri bir okula veya mesleğe yerleştirmek için test puanının ne olacağına karar vermeye gereksinim duyulabilir. Burada önemli olan, verilen kararın türü göz önüne alınmaksızın mümkün olan en doğru bilgiye ulaşmaktır. Bilgi ne kadar doğruysa verilen karar da o ölçüde geçerlidir. Pek çok durumda karar verme süreci birden fazla uygulamanın yapılmasını gerektirir. Bir testin farklı formlarının kullanılması ya da bir testin farklı zamanlarda uygulanması önemli bir probleme neden olabilir. Örneğin, ALES, KPSS, ÜDS, YGS gibi sınavlar bir yılda iki farklı dönemde ya da ardarda yıllarda öğrencilere uygulanmaktadır. Bu sınavlarda sınavın güvenliğini sağlamak amacıyla öğrencilere her dönem ya da her yıl farklı sorular sorulmaktadır. İki farklı dönemde uygulanan bu sınavlar benzer içerikte olmakla birlikte testlerin güçlüklerinin farklı olması öğrencileri değerlendirirken ve öğrenciler hakkında önemli kararlar alırken hatalara neden olabilir. Test geliştirme uzmanları istatistiksel özellikler ve içerik açısından mümkün olduğu kadar benzer test formları oluşturmaya çalışsa da formlar genellikle güçlük açısından bir miktar farklılık gösterir. Bu durum, öğrenciler hakkında önemli kararların alınmasına dayanak oluşturan sınavlardan elde edilen sonuçların geçerliğini olumsuz yönde etkilemektedir. Bu nedenle, ifade edilen bu problemten kaçınmak, aynı amaçla kullanılan farklı test formlarından alınan puanları karşılaştırabilmek ve bu puanların birbirinin yerine kullanılmasını sağlamak için test puanlarının eşitlenmesi gerekir. Bu nedenle eşitleme çalışmaları, test puanlarına dayalı olarak verilen kararların geçerliğini arttıran çok önemli çalışmalardır. Ancak eşitleme hatası, eşitlemenin doğruluğunu etkilemektedir.

Eşitleme hatasının kullanılan eşitleme deseni ve yöntemi, örneklem büyüklüğü, yetenek dağılımı, madde sayısı gibi değişkenlerden etkilendiği ve bunların etkileşimine bağlı olduğu pek çok çalışmada gösterilmiştir (Skaggs & Lissitz, 1986; Kim & Cohen, 2002; Hanson & Beguin, 2002; Keller III, 2007; Cao, 2008; Kim, Choi, Lee & Um, 2008; Nozawa, 2008; Dvorack, 2009; Puhan, 2010). Eşitlemenin doğruluğu, eşitleme yapılacak koşullara en uygun desenin ve yöntemin seçilmesine bağlıdır. Bu nedenle, bu çalışmada eşitleme hatasını etkileyen faktörlerden test uzunluğu, örneklem büyüklüğü, yetenek dağılımı ve model türü ele alınarak eşitleme hatasının hangi durumlarda daha az olacağı belirlenmeye çalışılmıştır. Çalışmada, değişkenler üzerinde kontrol sağlamak amacıyla simülasyon verileri kullanılmıştır.

MTK'ye dayalı eşitleme yöntemlerini araştırmak amacıyla pek çok simülasyon çalışması yürütülmüştür (Kim & Cohen, 1998; Wang, Hanson & Harris, 2000; Hanson & Beguin, 2002). Lehman ve Bailey (1968) ampirik bir çalışmanın mümkün olmadığı ya da çok masraflı olduğu durumlarda, simülasyon çalışmalarının rahatlıkla yapılabileceğini ifade etmişlerdir. Oh (2000), genellikle analiz için ihtiyaç duyulan veri miktarının ancak simülasyonla sağlanabildiğini, bu yüzden de simüle edilmiş verinin eşitleme çalışmalarında sıklıkla kullanıldığını ifade etmiştir. Türkiye'de yapılan ve sürekliliği olan seçme sınavlarında (YGS, KPSS, SBS gibi) yoklanan davranışlar değişmemekle birlikte, sınav güvenliği nedeniyle her yıl önceki yılların paraleli olduğu varsayılan yeni bir test uygulanır ve bu testlerin hiçbirinde bir önceki yıl ile veya farklı test formları arasında ortak maddeler yer almaz. Farklı yıllarda sınava giren ya da aynı sınavı tekrar alan bireylerin denkliliğini sağlamak güç olduğundan, denk grup deseninin uygulanması da problemlidir. Ayrıca ülkemizde uygulanan sınavlarda ortak maddeler yer almadığından ortak madde deseni kullanılarak eşitleme çalışması yapmak güçtür. Bu nedenle bu çalışmada bir simülasyon çalışması yoluyla MTK'ye dayalı kestirim yöntemlerinin performansı karşılaştırılmıştır.

Ülkemizde yapılan eşitleme çalışmalarında, genellikle gerçek veri ve tek grup deseni kullanıldığı görülmektedir (Çetin, 2009; Bozdağ & Kan, 2010, Öztürk, 2010; Kan, 2011). Bununla birlikte ortak madde deseni kullanılarak yatay eşitleme üzerine yapılan simülasyon çalışmalarının sayısı ise yok denecek kadar azdır. Bu çalışmada veriler türetilirken Türkiye'deki büyük ölçekli sınavların eşitleme sürecine katkı sağlamak

amacıyla, bu sınavların örneklem büyüklüğü, soru sayısı, yetenek dağılımı gibi özellikleri dikkate alınmıştır. Türkiye’de yapılan sınavların koşullarına göre, eşitleme yöntemlerinin performansının belirlenmesinin ülkemizde büyük ölçekli sınavlar üzerinde yapılacak eşitleme çalışmalarına da örnek olacağı düşünülmektedir.

Test eşitleme yöntemleri, aynı yapıyı ölçen farklı testlerden elde edilen puanları düzenlemede kullanılan psikometrik ve istatistiksel yöntemlerdir ve testler bu yöntemler kullanılarak karşılaştırılır (Dorans & Holland, 2000). Bu nedenle araştırmada farklı koşullara göre türetilen test formlarını MTK’ye dayalı kestirim yöntemlerini kullanarak eşitlemek ve bu yöntemlerden elde edilen sonuçları karşılaştırmak amaçlanmıştır. Ayrıca araştırmada hangi koşullarda hangi yöntemin daha iyi sonuç verdiği de incelenmiştir.

1.8. PROBLEM CÜMLESİ

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak iki ve üç parametrelili lojistik modele uyumlu veriler ile denk olmayan gruplarda ortak maddelere dayalı olarak yapılan yatay eşitlemelerin simülasyon koşullarına göre yanlılıkları ve hataları nasıl değişmektedir?

1.9. ALT PROBLEMLER

1. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak üç parametrelili lojistik modele uyumlu veriler ile kestirilen madde parametrelerinin yanlılık ve eşitleme hatası;
 - a) Test uzunluğuna
 - b) Örneklem büyüklüğüne
 - c) Yetenek dağılımına göre nasıl değişmektedir?
 - d) Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğu, örneklem büyüklüğü, yetenek dağılımına göre en iyi sonucu vermektedir?

2. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak iki parametrelili lojistik modele uyumlu veriler ile kestirilen madde parametrelerinin yanlılık ve eşitleme hatası;
 - a) Test uzunluğuna
 - b) Örneklem büyüklüğüne
 - c) Yetenek dağılımına göre nasıl değişmektedir?
 - d) Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğu, örneklem büyüklüğü, yetenek dağılımına göre en iyi sonucu vermektedir?
3. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak elde edilen madde parametrelerinin yanlılık ve eşitleme hatası model türüne göre nasıl değişmektedir?
4. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak üç parametrelili lojistik modele uyumlu veriler ile kestirilen yetenek parametrelerinin yanlılık ve eşitleme hatası;
 - a) Test uzunluğuna
 - b) Örneklem büyüklüğüne
 - c) Yetenek dağılımına göre nasıl değişmektedir?
 - d) Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğu, örneklem büyüklüğü, yetenek dağılımına göre en iyi sonucu vermektedir?
5. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak iki parametrelili lojistik modele uyumlu veriler ile kestirilen yetenek parametrelerinin yanlılık ve eşitleme hatası;
 - a) Test uzunluğuna
 - b) Örneklem büyüklüğüne
 - c) Yetenek dağılımına göre nasıl değişmektedir?
 - d) Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğu, örneklem büyüklüğü, yetenek dağılımına göre en iyi sonucu vermektedir?

6. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak elde edilen madde parametrelerinin yanlılık ve eşitleme hatası model türüne göre nasıl değişmektedir?
7. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak iki parametrelili lojistik modele uyumlu veriler ile gerçek puan eşitlemesine dayalı olarak yapılan eşitlemelerin yanlılık ve eşitleme hatası;
 - a) Test uzunluğuna
 - b) Örneklem büyüklüğüne
 - c) Yetenek dağılımına göre nasıl değişmektedir?
 - d) Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğu, örneklem büyüklüğü, yetenek dağılımına göre en iyi sonucu vermektedir?
8. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak iki parametrelili lojistik modele uyumlu veriler ile gerçek puan eşitlemesine dayalı olarak yapılan eşitlemelerin yanlılık ve eşitleme hatası;
 - a) Test uzunluğuna
 - b) Örneklem büyüklüğüne
 - c) Yetenek dağılımına göre nasıl değişmektedir?
 - d) Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğu, örneklem büyüklüğü, yetenek dağılımına göre en iyi sonucu vermektedir?
9. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak gerçek puan eşitlemesine dayalı olarak yapılan eşitlemelerin yanlılık ve eşitleme hatası model türüne nasıl değişmektedir?

1.10. SINIRLILIKLAR

1. Araştırma iki kategorili verilerle sınırlıdır.
2. MTK'ye dayalı iki ve üç parametrelili lojistik modellerle sınırlıdır.
3. Araştırmada ele alınan simülasyon koşulları; madde sayısı, kişi sayısı, yetenek dağılımı ve kullanılan modeller ile sınırlıdır.

1.11. KISALTMALAR

MTK: Madde tepki kuramı

KTT: Klasik test kuramı

3PLM: Üç parametrelili lojistik model

2PLM: İki parametrelili lojistik model

NEAT: Denk olmayan gruplarda ortak test/madde deseni

RMSE: Hata kareleri ortalamasının karekökü (Eşitleme hatası)

BIAS: Eşitleme yanlılığı

EAP: Expected a posteriori

OO: Ortalama-ortalama

OS: Ortalama-standart sapma

SL: Stocking-Lord

1.12. İLGİLİ ARAŞTIRMALAR

Stocking ve Lord (1983) yaptıkları araştırmada Stocking-Lord ve ortalama-sigma yöntemlerini karşılaştırmışlardır. Araştırma sonucunda Stocking-Lord yönteminin daha doğru kestirime yol açtığı bulunmuştur.

Wingersky, Cook ve Eignor (1987) farklı uzunluktaki ortak testler altında eşitlemenin doğruluğunu araştırmak için simülasyon çalışması yürütmüşlerdir. Çalışmada ortak test uzunlukları 10, 20 ve 40 madde olarak belirlenmiştir. Çalışmada ortak madde parametre kestirimlerinin standart hatalarının büyüklüğünün etkisi de araştırılmıştır. Tüm çalışma koşulları hem eş zamanlı kalibrasyon hem de Stocking-Lord kalibrasyon/kestirim yöntemleri kullanılarak incelenmiştir. Sonuçlar ortak maddelerin sayısının artmasının daha doğru eşitleme sonuçlarına neden olmakla birlikte tek biçimli cevaplayıcı dağılımlarının daha iyi parametre kestirimlerine yol açtığını ve böylece sivri dağılımların daha doğru eşitleme sonuçları verdiğini ortaya koymuştur. Bununla birlikte araştırmada, küçük eşitleme hatasına sahip maddelerin büyük eşitleme hatasına sahip maddelerden daha iyi performans gösterdiği de bulunmuştur. Yöntemler açısından bakıldığında ise, eşzamanlı kestirim yönteminin daha az eşitleme hatasına yol açtığı saptanmıştır.

Baker ve Al-karni (1991) yaptıkları çalışmada eşitleme katsayılarını hesaplamak için MTK'ye dayalı iki yöntemi (Loyd-Hoover, Stocking-Lord) karşılaştırmışlardır. Çalışmada iki tür eşitleme koşulu ele alınmıştır: dikey eşitleme ve yatay eşitleme. Araştırmadan elde edilen sonuçlar genel olarak her iki eşitleme yönteminin de tüm koşullarda benzer eşitleme katsayılarına yol açtığını göstermiştir. Araştırmada Stocking-Lord yönteminin test özelliklerine daha az duyarlı olduğu ortaya çıkmakla birlikte, bu yöntemin daha tercih edilebilir olduğu sonucuna da varılmıştır.

Kim ve Cohen (1998) yaptıkları çalışmada parametre kestirimlerini ortak bir ölçeğe yerleştirmek için kullanılan üç MTK yöntemini karşılaştırmışlardır: Karakteristik eğri yöntemlerinden elde edilen eşitleme katsayılarını kullanarak yapılan ayrı kalibrasyon yöntemi, marjinal maksimum likelihood yoluyla eş zamanlı kalibrasyon yöntemi ve marjinal maksimum posteriori yoluyla eş zamanlı kalibrasyon yöntemi. Araştırmada karakteristik eğri yöntemi kullanılarak yapılan kestirimin ortak maddelerin sayısı az olduğunda hem güçlük hem de ayırıcılık parametreleri için daha küçük RMSE değerlerine yol açtığı, ortak maddelerin sayısı arttığında ise her üç yöntemin de aynı sonuçları verdiği görülmüştür.

Hanson ve Beguin (1999, 2002) yaptıkları çalışmada ortalama-ortalama, ortalama-standart sapma ve karakteristik eğri yöntemlerinin performansını belirlemek için bir simülasyon çalışması yürütmüşlerdir. Araştırmada ortak maddelerin yer aldığı ve her bir test formunun 60 maddeden oluştuğu iki test formu kullanılmıştır. Bu çalışmada ele alınan üç faktör, örneklem büyüklüğü (1000, 3000), ortak maddelerin sayısı (10, 20) ve cevaplayıcı gruplarının denk olup olmamasıdır. Örneklem büyüklüğü ve ortak madde sayısı arttıkça, incelenen MTK eşitleme yöntemlerinin her biri için, hata kareleri ortalaması daha az bulunmuştur. Çalışmada denk grup koşulunda denk olmayan gruba göre daha az hatalı sonuçlar bulunmuştur. Bu durumun kaynağı olarak bir grubun ortalama yeteneği ile ilişkili yanlılık olabileceği ifade edilmiştir. Araştırmada eş zamanlı kestirimin farklı kestirimden daha az hatalı sonuçlara yol açtığı ortaya çıkmakla birlikte, araştırmanın hangi yöntemin daha tercih edilebilir olduğunu söylemek için yeterli olmadığı sonucuna da varılmıştır.

Beguın ve Hanson (2001) tarafından yapılan simülasyon çalışmasında eş zamanlı ve farklı kalibrasyon yöntemlerinin performansı karşılaştırılmıştır. Denk olan ve denk olmayan gruplar için veri türetilmiştir. Araştırmada parametrik model varsayımları karşılandığında (tek boyutlu MTK modeli), eş zamanlı kestirimin farklı kestirimden daha doğru sonuçlara yol açtığı görülmüştür. Çok boyutlu MTK modelinin tek boyutlu modelden daha az hatalı kestirime neden olduğu ve çok boyutlu modelin tek boyutlu modele göre daha yanlış sonuçlar ürettiği görülmüştür. Verinin çok boyutlu olmasının kestirim yöntemlerinin performansını etkilediği de bulunmuştur.

Kaskowitz ve De Ayala (2001) tarafından yapılan araştırmada, karakteristik eğri yöntemlerinin madde parametre kestirim hatasının varlığına duyarlılığı incelenmiştir. Bu simülasyon çalışmasında, iki ve üç parametrelı lojistik modeller kullanılarak ortak madde parametre kestirimlerinin etkisi incelenmiştir. 3PLM için hata düzeyleri 0.30, 0.91 ve 2.73'tür. 2PLM için ise 0.11, 0.25 ve 0.63'dir. Araştırmada 3PLM için karakteristik eğri yöntemlerinin daha yüksek hataya neden olduğu ortaya çıkmıştır. Bununla birlikte sonuçlar karakteristik eğri yöntemlerinin madde parametre kestirim hatasının varlığına daha duyarlı olduğunu göstermiştir. Aynı zamanda araştırmada karakteristik eğri yöntemlerinin diğer MTK eşitleme yöntemlerinden daha iyi performans gösterdiği ve diğer yöntemlere tercih edilebilir olduğu sonucuna da varılmıştır. Bu araştırma ve bu konuda yapılan diğer araştırmalar karakteristik eğri yöntemlerinin ortak maddelerin sayısı ve örneklem büyüklüğü gibi faktörlerden etkilendiğini göstermiştir (Cook & Petersen, 1987; Hanson & Beguin, 2002). Bununla birlikte araştırma, örneklem büyüklüğüne ek olarak ortak maddelerin sayısının artmasının daha doğru eşitleme sonuçlarına yol açtığını da ortaya koymuştur.

Kim ve Cohen (2002) yaptıkları simülasyon çalışmasında karakteristik eğri yöntemleri ile eş zamanlı kalibrasyonu karşılaştırmışlardır. Araştırmada koşul olarak 300 ve 1000 kişilik örneklem büyüklükleri; $N(0,1)$ (1, 1) olmak üzere iki yetenek dağılımı; 5, 10 ve 30 olmak üzere üç farklı ortak madde uzunluğu alınmıştır. Araştırmanın sonucunda 1000 kişilik örneklemelerin 300 kişilik örnekleme göre, yetenek dağılımının benzer olduğu grupların farklı olan gruplara göre ve ortak madde sayısının büyük olduğu grupların küçük gruplara göre daha düşük hataya sahip oldukları görülmüştür. Bununla

birlikte arařtırmada, eř zamanlı kalibrasyonun ayrı kalibrasyondan daha iyi sonuçlar verdiđi de bulunmuřtur.

Keller, Skorupski, Swaminathan ve Jodoin (2004) tarafından yapılan alıřmada simülasyon verisi kullanılarak FCIP (fixed common item parameters-sabit ortak madde parametreleri) eř zamanlı, ortalama-standart sapma ve Stocking-Lord kestirim yöntemleri karşılaştırılmıřtır. alıřmada genel olarak ortalama-standart sapma ve eř zamanlı kestirim yöntemlerinin en dođru sonuçları verdiđi, FCIP yönteminin ise en hatalı kestirimlere yol atıđı bulunmuřtur.

Kim ve Kolen (2006) yaptıkları arařtırmada karıřık formattaki testler (oktan semeli ve aık ulu soruların bir arada bulunduđu testler) için moment yöntemleri, karakteristik eđri yöntemleri ve eř zamanlı kestirim yöntemlerini karşılařtırmıřlardır. Arařtırmanın sonucunda eř zamanlı kestirim yönteminin diđer yöntemlere göre daha az hatalı sonuçlara yol atıđı bulunmuřtur. Bununla birlikte arařtırmada karakteristik eđri yöntemlerinin moment yöntemlerinden daha iyi sonuçlar verdiđi görülürken, en fazla hataya sahip yöntemin de ortalama-ortalama yöntemi olduđu bulunmuřtur.

Kim ve Lee (2006) tarafından yapılan simülasyon alıřmasında karıřık formattaki testler için ortalama-ortalama, ortalama-standart sapma, Stocking-Lord ve Haebara yöntemleri karşılaştırılmıřtır. Arařtırmada 3PLM ve ok kategorili puanlanan maddeler için geliştirilmiř yöntemler kullanılmıřtır. Arařtırma sonucunda, genel olarak Haebara ve Stocking-Lord yöntemlerinin ortalama-ortalama ve ortalama-standart sapma yöntemlerinden daha dođru sonuçlar verdiđi görülmüřtür.

Keller III (2007) yaptıđı arařtırmada, ortak maddelerin uzunluđu, veri-model uyumsuzluđu, ölekleme yöntemi, yetenek dađılımı gibi faktörlere göre türetilen veri üzerinde MTK gerek puan eřitileme ile kestirilen gerek eřitilemenin performansını karşılařtırmıřtır. Sonuçlar kararın dođruluđu, eřitileme yanlılıđı (BIAS) ve eřitileme hatası (RMSE) ölçütlerine göre deđerlendirilmiřtir. Arařtırmada ölekleme yöntemlerinin performansı da incelenmiřtir. Sonuçlar, kestirilen gerek eřitilemenin gerek puan eřitilemeden daha düşük yanlılık ve RMSE deđerlerine yol atıđını

göstermiştir. Bununla birlikte sonuçlar hangi ölçekleme yönteminin diğerine tercih edilebileceğini göstermede başarısız olmuştur.

Norman-Dvorak (2009) farklı koşullar altında kernel eşitleme ile karakteristik eğri eşitleme yöntemlerinin doğruluğunu araştırmıştır. Araştırmada denk olmayan gruplarda ortak madde deseni kullanılmıştır. Simülasyon verisi üzerinde test uzunluğu, ortalama faktör yükü, örneklem büyüklüğü, ortak maddelerin yüzdesi gibi faktörlerin test eşitleme yöntemlerinin doğruluğuna etkisi incelenmiştir. Araştırmada her iki eşitleme yönteminin de yukarıda ifade edilen faktörlere göre iyi bir performans gösterdiği bulunmuştur. Örneğin, uzun testlerde (75 maddelik) örneklem büyüklüğü veya ortak maddelerin yüzdesini göz önüne almaksızın ortalama faktör yükü 0.62 olduğunda kernel eşitleme her zaman daha doğru sonuçlar verirken, ortalama yük 0.50 olduğunda, büyük örneklemelerde ve ortak maddeler tüm testin %30'unu oluşturduğunda karakteristik eğri eşitleme yöntemleri çok daha doğru sonuçlar vermiştir. Sonuç olarak araştırmada tüm koşullar genel olarak incelendiğinde kernel eşitleme yönteminin karakteristik eğri eşitleme yöntemlerine göre daha iyi performans gösterdiği bulunmuştur.

Speron (2009) yaptığı araştırmada ortalama-ortalama, ortalama-standart sapma ve karakteristik eğri yöntemlerini (Stocking-Lord, Haebara) karşılaştırmayı amaçlamıştır. Bu çalışmada veriler 2PLM'ye göre türetilmiştir. Araştırmada simülasyon koşulu olarak örneklem büyüklüğü, ortak maddelerin sayısı, ortak maddelerin toplam testi temsil etme derecesi ve grupların yetenek dağılımları alınmıştır. Araştırmanın sonucunda, karakteristik eğri yöntemleri ve ortalama-ortalama yöntemlerinin benzer sonuçlar verdiği ve daha az hatalı kestirime neden olduğu görülmüştür. Bununla birlikte araştırmada ortalama-standart sapma yönteminin diğer yöntemlerden farklı sonuçlar verdiği ve en fazla hataya sahip yöntem olduğu da belirlenmiştir. Örneklem büyüklüğü arttıkça eşitleme hatalarının düştüğü gözlenmiştir.

Özetle, MTK'ye dayalı kestirim yöntemlerini karşılaştıran araştırmaların sonuçları gözden geçirildiğinde, genel olarak MTK'ye dayalı bir kestirim yöntemi olan karakteristik eğri yöntemlerinin, diğer yöntemlerden daha duyarlı ve daha doğru sonuçlar verdiği bulunmuştur (Baker & Al-Karni, 1991; Kolen & Brennan, 2004; Stocking & Lord, 1983) ve bu yöntemin son zamanlarda uygulamada yaygın olarak

kullanıldığı görülmüştür. Buna rağmen koşullar değiştikçe (genellikle uç durumlarda) yöntemlerin hatalarında da farklılıklar olduğu gözlenmiştir. Örneğin, bazı araştırmalarda en hatalı yöntemin ortalama-ortalama yöntemi olduğu bulunurken (Kim ve Kolen, 2006), bazı araştırmalarda ise ortalama-standart sapma yönteminin en fazla hataya sahip yöntem olduğu bulunmuştur (Speron, 2009). Araştırma sonuçları, test uzunluğu, örneklem büyüklüğü, ortak maddelerin sayısı ve özellikleri vb. koşulların MTK yöntemlerinin performansını etkilediğini ortaya çıkarmıştır. Bununla birlikte genel olarak, bazı istisnalar dışında MTK'ye dayalı kestirim yöntemlerinin ortak maddelerin sayısı, test uzunluğu ve örneklem büyüklüğü arttıkça hatalarının azaldığı görülmüştür.

BÖLÜM II

YÖNTEM

Bu bölümde araştırmanın türü, veri toplama deseni, simülasyon faktörleri ve koşulları, verilerin türetilmesi (simülasyon çalışması), araştırmada kullanılan eşitleme yöntemleri, eşitleme sürecinin uygulanması, değerlendirme ölçütü ve verilerin analizi üzerinde durulmuştur.

2.1. ARAŞTIRMANIN TÜRÜ

Bu araştırmada, belirlenen koşullara göre türetilen veriyi kullanarak madde tepki kuramına dayalı kestirim yöntemlerinin karşılaştırılması amaçlanmıştır. Böylece en az hatalı yöntemler ve koşullar belirlenerek kuramsal çalışmalara katkıda bulunulması planlanmıştır. Araştırma kestirim yöntemlerini karşılaştırma amacı taşıdığından temel araştırma niteliği taşımaktadır.

2.2. VERİ TOPLAMA DESENİ

Bu araştırmada iki test formunu eşitleyebilmek için “denk olmayan gruplarda ortak madde/test (NEAT) deseni” kullanılmıştır. NEAT deseni uygulamada yaygın olarak kullanılan desenlerden biri olmakla birlikte, en esnek ve en karmaşık desenlerden biridir (Sinharay & Holland, 2008). Bununla birlikte diğer desenlere göre daha az sınırlayıcıdır ve pratiklik açısından da tercih edilen bir yöntemdir (Zhu, 1998). Bu desende yeni test formu form X ile gösterilir, eski test formu ise form Y ile gösterilir. Form X ve form Y’de ortak maddeler vardır. İki cevaplayıcı grubu iki farklı popülasyondan elde edilmiştir. Form X’i alan cevaplayıcı grubu grup 1, form Y’yi alan cevaplayıcı grubu ise grup 2 ile ifade edilir. Araştırmada standart normal dağılım için kullanılan NEAT deseni şekil 2.1’de gösterilmiştir. Aynı durum çarpık dağılımlar için de geçerlidir.

Grup 1~ N(0,1) Form X

X formuna ait maddeler
Ortak maddeler

Grup 2~ N(0.5, 1) Form Y

Y formuna ait maddeler
Ortak maddeler

Şekil 2.1. Araştırmada kullanılan eşitleme deseni

Eşitleme çalışmalarında, genel olarak iki farklı veri seti (gerçek veri ve simüle edilmiş veri) kullanılır. Ülkemizde uygulanan büyük ölçekli sınavlar eşitleme desenlerine uygun olarak hazırlanmadığı için bu araştırma gerçek veri yerine simülasyon verisi üzerinde yürütülmüştür. Bununla birlikte, gerçek veride madde ve cevaplayıcı sayısı sınırlıdır. Bu yüzden Douglas (1997)'ın da ifade ettiği gibi madde ve cevaplayıcı sayısı yeterli olmadığında yöntemleri birbirleriyle karşılaştırmak çok doğru sonuçlar vermeyebilir. Bununla birlikte, simülasyon çalışması dikkatli bir şekilde tasarlanırsa, gerçek veriye uygun özellikte olması sağlanabilir ve simülasyonlar gerçek veri için ön deneme modeli olarak da görülebilir (Holland, von Davier, Sinharay & Han, 2006). Simülasyon verisinin kullanılması her bir eşitleme yönteminin doğruluğunun belirlenmesine imkân tanır ve simülasyon çalışmasıyla gerçek eşitleme ilişkisi bilinebilir. Bununla birlikte, kesin bir değerlendirme ölçütünün eksikliğinden dolayı gerçek veri kullanılarak hangi eşitleme yönteminin daha doğru sonuçlar verdiğini bilmek imkânsızdır. Yöntemlerden elde edilen eşitleme sonuçları gerçek puanlar ile karşılaştırılabilir. Böylece, eşitleme yöntemlerinin performansı ve hangi yöntemin hangi koşulda en iyi sonuçlar verdiği belirlenebilir.

2.3. SİMÜLASYON FAKTÖRLERİ VE KOŞULLARI

Bu simülasyon çalışmasının amacı, çeşitli koşullar altında eşitleme yöntemlerinin etkililiğini değerlendirmektir. Çalışmada kullanılan simülasyon faktörleri; testlerdeki madde sayısı, her bir gruptaki cevaplayıcı sayısı, yetenek dağılımı ve veri türetmede kullanılan modelin türüdür. Araştırmada ele alınan bu faktörlerin eşitleme hatasını etkilediği çeşitli araştırmalarda ortaya konmuştur (Hanson & Beguin, 1999, 2002; Kim & Cohen, 2002; Speron, 2009). Uygulamada, bu faktörlerin kestirim sürecini etkileyen önemli faktörler olduğu ifade edilmektedir (Kang & Petersen, 2009). Simülasyon çalışmalarının koşulları, gerçek uygulamaları yansıttığı ölçüde bu uygulamalara

genellenebilir. Bu nedenle, bu çalışmada ele alınan koşulların, ülkemizdeki büyük ölçekli sınavların özelliklerine uygun olmasına dikkat edilmiştir. Araştırmada eşitleme deseni olarak tüm faktörlerde ortak maddeye dayalı yatay eşitleme kullanılmıştır.

Araştırmada ele alınan ilk faktör test uzunluğudur. Test uzunluğu için her bir formda 30, 60 ve 80 maddenin yer aldığı üç koşul belirlenmiştir. Bu koşullar, ülkemizde uygulanan sınavlardaki ortalama madde sayıları temel alınarak belirlenmiştir (Bu sınavlar sırasıyla SBS, YGS ve ALES'tir). Bu çalışmada kullanılan test uzunlukları, genel olarak eşitleme üzerine yapılan pek çok simülasyon çalışmasında kısa, orta ve uzun olarak ifade edilir (Lee, 2007; Hou, 2007; Norman-Dvorak, 2009).

Her üç koşul için maddelerin en az %20'si ortak madde olarak alınmıştır. Bu nedenle n=30 koşulunda 10 madde, n=60 koşulunda 20 madde, n=80 koşulunda ise 30 madde ortak madde olarak düşünülmüştür. Genel olarak, ortak madde deseninde, maddelerin en az %20'si ortak madde olarak kabul edilmekle birlikte (Kolen & Brennan, 2004), yapılan çalışmalarda test uzunluğunun bir fonksiyonu olarak ortak maddelerin en fazla kaç olması gerektiği belirlenmemiştir (Sinharay & Holland, 2007). Ortak maddeler "iç ortak madde" olarak alınmıştır. Bu araştırmada ortak madde sayısı testlere göre değişmekle birlikte, bir faktör olarak ele alınmamıştır.

Araştırmada ele alınan ikinci faktör, her bir gruptaki cevaplayıcı sayısıdır. Her bir grup için 1000 ve 3000 cevaplayıcı olmak üzere iki koşul belirlenmiştir. Bu konuda yapılan diğer araştırma sonuçlarında da görüldüğü gibi, (Hanson & Beguin, 2002; Kim & Lee, 2004; Cao, 2008; Cui & Kolen, 2008; Zhao, 2008; Nozawa, 2009) seçilen bu örneklem büyüklüklerinin doğru eşitleme sonuçlarına yol açmak için yeterli olduğu söylenebilir. Başarılı bir eşitleme için en az 500 kişilik örneklem büyüklüğüne ihtiyaç duyulduğu ifade edilmekle birlikte (Spence, 1996), 1000 ve daha büyük örneklem için yöntemlerin daha iyi sonuçlar verdiği görülmüştür (Liou, Cheng & Johnson, 1997; Han, 2008). Skaggs ve Lissitz (1986) ise yaptıkları çalışmada yatay eşitleme için 3PLM kullanıldığında, örneklem büyüklüğünün en az 1000 olması gerektiğini ifade etmişlerdir. Ayrıca alanyazında Bilog ve Parscale gibi programlarla yapılan analizlerde en az 1000 cevaplayıcı için madde parametrelerinin doğru bir şekilde kestirildiği de belirtilmiştir (Yen, 1987; Mislevy & Stocking, 1989). Bununla birlikte 1000

cevaplayıcının bazı arařtırmalarda “orta” (Han, 2008), bazı arařtırmalarda ise “byk” rnekleme byklg olarak ifade edildiđi grlrken (Lee, 2007; Douglas & Cohen, 2001), bu arařtırmada ele alınan rnekleme byklkleri (1000 ve 3000), Hanson ve Beguin (2002)’in yaptıkları alıřmada sırasıyla “kk” ve “byk” rnekleme byklg olarak ifade edilmiřtir. Bunun yanı sıra, arařtırmada kullanılan rnekleme byklklerinin alanyazında gerek test etme uygulamalarında kullanılan rnekleme byklkleri olarak ifade edildiđi grlmřtr (Hou, 2007).

nc faktr, grup 1 ve grup 2 iin yetenek dađılımlarıdır. Bu faktr altında  kořul ele alınmıřtır. İlk yetenek dađılımı kořulunda grup 1 ortalaması 0 standart sapması 1 olan standart normal dađılıma sahiptir. Grup 2’nin yetenek dađılımı ise grup 1’den 0.5 daha yksek olmakla birlikte her iki grubun standart sapmaları aynıdır. İki grup arasında ortalamalar arasında 0.5’lik farkın seilmesinin nedeni, bu deđerin eřitleme zerine grup farklılıklarının etkisini yansıtmak iin yeterli byklkte olduđunun gsterilmiř olmasıdır (Li & Lissitz, 2000). Bu alıřmada pek ok simlasyon alıřmasında olduđu gibi (Kim & Kolen, 2006; Kim & Lee, 2006; Keller III, 2007; Cao, 2008; Zhao, 2008) sadece ortalamadaki farklılıklar gz nne alınarak standart sapmalar sabit tutulmuřtur. Eřitleme alıřmaları, genellikle simetrik veya normal dađılıma sahip byk lekli sınavlar iin yapılırken, yksek geme standartlarına sahip sınavlar ve yetenek testleri gibi arpık dađılıma sahip testleri eřitlemek iin de gereklidir. Gerek test etme uygulamalarında arpık dađılımlara sıklıkla rastlanır (Kolen, 1985). Bu yzden, ikinci yetenek dađılımı kořulunda, grup 1 pozitif arpık olurken, grup 2 ise standart normal dađılıma sahip olacak řekilde belirlenmiřtir. nc yetenek dađılımı kořulunda ise grup 1 negatif arpık olurken, grup 2 ise standart normal dađılıma sahiptir. Yapılan pek ok arařtırmada arpık yetenek dađılımlarının eřitleme yntemleri zerindeki etkisinin arařtırılmasının nemli olduđu ifade edilmiřtir (Kang & Petersen, 2009). Bununla birlikte, Hanson ve Beguin (2002) denk olmayan yetenek dađılımına sahip grupların bir kořul olarak alınmasının olduka nemli olduđunu ifade etmiřlerdir.

Pozitif ve negatif arpık dađılım kořullarında, yetenek parametreleri beta dađılımdan tretilmiřtir. Bylece iki dađılım arasındaki temel farklılık, arpıklıktan kaynaklanır. Dađılımlardan birinin pozitif arpık, diđerinin ise negatif arpık seilmesinin nedeni ise Trkiye’de yapılan byk lekli sınavların fen ve matematik gibi sayısal alt testlerinin

pek çoğunda dağılım pozitif çarpık olmakla birlikte, Türkçe gibi sözel alt testlerinde ise genel olarak negatif çarpık dağılım gözlenmesidir (Bekci, 2007; Bakan Kalaycıoğlu, 2008; Çetin, 2009; Öztürk, 2010). Çarpıklık katsayısı testlerde genellikle 0.30–0.40 arasında değiştiğinden, bu araştırmada da ortalama çarpıklık bu değerler arasındadır.

Dördüncü faktör, veriyi türetmede kullanılan modelin türüdür. Bu koşulda maddeler iki ve üç parametrelili lojistik modele göre türetilmiştir. Form X, form Y ve ortak maddeler için iki ve üç parametrelili lojistik modele göre türetilen maddelerin sayısı Tablo 2.1’de verilmiştir:

Tablo 2.1. İki ve üç parametrelili lojistik modele göre türetilen madde sayısı

Maddenin türetildiği model	Form X	Ortak maddeler	Form Y	Toplam
2PLM	20	10	20	30
3PLM	20	10	20	30
2PLM	40	20	40	60
3PLM	40	20	40	60
2PLM	50	30	50	80
3PLM	50	30	50	80

Araştırmada yukarıda ifade edilen bu dört simülasyon faktörü ve bu faktörlere ait koşullar çarpılarak koşulların tüm olası kombinasyonları incelenmiştir (3x2x3x2 olmak üzere toplam 36 kombinasyon). Bu koşullar Tablo 2.2’de verilmiştir:

Tablo 2.2. Araştırmada kullanılan simülasyon koşulları

Yetenek Dağılımı		Model Türü Madde sayısı						
		2 PLM			3PLM			
Grup 1	Grup 2	n=30	n=60	n=80	n=30	n=60	n=80	
N=1000	N (0,1)	N (0.5,1)	1 ^a	2	3	4	5	6
	PÇ ^b	N (0,1)	7	8	9	10	11	12
	NÇ ^c	N (0,1)	13	14	15	16	17	18
N=3000	N (0,1)	N (0.5,1)	19	20	21	22	23	24
	PÇ	N (0,1)	25	26	27	28	29	30
	NÇ	N (0,1)	31	32	33	34	35	36

^a Eğik sayılar 1 ile 36 arasındaki simülasyon koşullarını gösterir.

^b PÇ pozitif çarpık dağılım, beta (6,14) dağılımdan türetilmiştir.

^c NÇ negatif çarpık dağılım, beta (14, 6) dağılımdan türetilmiştir.

2.4. VERİLERİN TÜRETİLMESİ (SİMÜLASYON ÇALIŞMASI)

Araştırmada hangi eşitleme yönteminin hangi koşullar altında daha iyi sonuçlar verdiğini ortaya koymak amacıyla Monte-Carlo simülasyon çalışması yürütülmüştür. Bu tür çalışmalar, örneklem büyüklüğü, test uzunluğu, dağılım özellikleri gibi faktörlerin yöntemler üzerine etkilerini değerlendirmenin yollarından biridir. Bu çalışmaların önemli bir parçası, madde cevap verisini gerçek parametre değerlerine göre türetmektir (Han, 2007). Araştırmada madde cevap verisi iki kategorili olarak türetilmiştir. Form X'i alan grup 1'de yer alan cevaplayıcılar ile form Y'yi alan grup 2'de yer alan cevaplayıcıların madde cevapları ayrı ayrı türetilmiştir. Her bir koşul altında her bir grup için cevaplayıcıların cevapları iki ve üç parametrelili lojistik MTK modelleri kullanılarak türetilmiştir. Araştırmada kullanılan test formlarının dağılımları gerçek test formlarına benzer olacak şekilde türetilmiştir. Ortak maddelerin güçlük düzeyleri ise, form X ve form Y'ye benzerdir.

Araştırmada ele alınan iki kategorili madde cevapları Wingen3 simülasyon programı kullanılarak türetilmiştir (Han, 2007). Bu program, MTK'ye göre iki kategorili ve çok kategorili madde cevaplarını türetmek için geliştirilmiştir. Veri türetme süreci üç basamakta gerçekleştirilmiştir. Her bir basamak aşağıda açıklanmıştır:

a. Yetenek parametrelerinin türetilmesi: Yetenek dağılımları her bir grup için standart normal dağılımdan ($\theta \sim N(0,1)$) örneklenmekle birlikte, aynı zamanda pozitif ve negatif çarpık dağılım da kullanılmıştır.

b. Madde parametrelerinin türetilmesi: Araştırmada gerçeğe yakın parametre değerleri elde etmek için bu konuda yapılan diğer araştırmalarda kullanılan parametre değerlerinden yararlanılmıştır (Kim & Lee, 2006; Cao, 2008). Eşitleme için iki test formu türetilmiştir. Form X, form Y ile ortak ve form Y'den farklı maddelerden oluşmaktadır. Form Y ise, Form X ile ortak olan maddelerden ve Form Y'ye özgü olan farklı maddelerden oluşmaktadır. Bu üç madde seti ayrı ayrı türetilmiştir.

Form 1 için farklı maddelere ait madde parametreleri şu şekilde belirlenmiştir: Madde ayırıcılık parametreleri (a_i), ortalaması 0 standart sapması 0.5 olan log-normal dağılımdan türetilmiştir. Bu değerler BILOG-MG'deki madde ayırıcılık

parametrelerinin varsayılan dağılımıdır. Madde güçlük parametreleri (b_i) ortalaması 0 standart sapması 1 olan normal dağılımdan türetilmiştir. Şans parametresi (c_i) ise Kim ve Lee (2006) ile Cao (2008) tarafından da kullanıldığı gibi $\alpha=8$, $\beta=32$ olduğu beta dağılımından türetilmiştir.

Form Y için farklı maddelere ait madde parametreleri güçlük parametresi dışında form X'dekiyle aynı madde parametre dağılımlarından türetilmiştir. Eşitlemenin amacı, formlardaki güçlük farklılıklarını istatistiksel olarak düzeltmek olduğu için, madde güçlük parametrelerinin dağılımı form Y için form X'e göre 0.5 artırılmıştır. Bu durum form Y'nin form X'den daha zor olduğunu gösterir. Kim (2006) yaptığı araştırmada güçlük ve ayırıcılık parametreleri arasında bir ilişki olmadığını ortaya koyduğu için araştırmada bu parametreler birbirinden bağımsız olarak türetilmiştir.

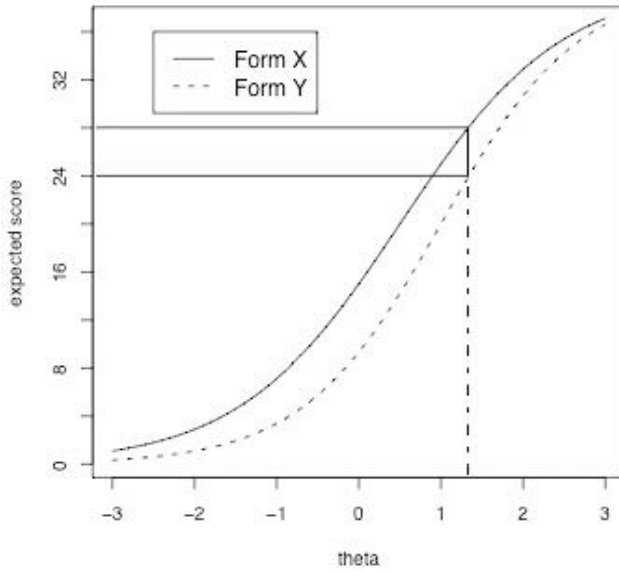
c. Madde cevaplarının türetilmesi: Her bir form için madde parametreleri ve her bir grup için yetenek parametreleri belirlendikten sonra, araştırmanın verilerini türetmede MTK modelleri kullanılmıştır. Wingen simülasyon programında maddeleri türetmek için iki ve üç parametrelili lojistik model kullanılmıştır. Madde cevapları random olarak türetilmiştir. Simülasyon çalışmalarının pek çoğunda, madde cevapları iki veya üç parametrelili lojistik modelden türetilmiştir (Nozawa, 2008). Ülkemizde ve yurt dışında yapılan büyük ölçekli sınavlarda genellikle cevaplayıcıların şansla doğru cevabı bulmaları olası olduğu için verilerin türetilmesinde 3PLM kullanılırken, şansla doğru cevabı bulma ihtimalinin olmadığı durumlarda yöntemlerin nasıl sonuçlar verdiğini görmek amacıyla 2PLM'de kullanılmış ve her iki modelden türetilen cevaplara göre modellerden elde edilen sonuçlar karşılaştırılmıştır. Bununla birlikte c parametresi 0'a sabitlendiğinde en iyi güçlük ve ayırıcılık parametreleri elde edildiğinden, madde parametrelerinin kalitesini incelemek için de 2PLM simülasyon koşulu olarak araştırmaya dâhil edilmiştir.

Form X, form Y ve ortak maddeler için türetilen madde parametreleri Ek 1'de sunulmuştur. Araştırmada kullanılan 36 koşulun her biri için 50 tekrar yapılmıştır.

2.5. ARAŞTIRMADA KULLANILAN EŞİTLEME YÖNTEMLERİ

Parametrik MTK'ye dayalı eşitleme yöntemleri her bir simülasyon koşulu için türetilen 50 veri setine uygulanmıştır. Araştırmada 2PLM ve 3PLM'ye dayalı ayrı kestirim yöntemleri ve gerçek puan eşitleme kullanılmıştır.

Genellikle puanları rapor etmede yorumlanabilir bir ölçek elde etmek ve negatif puanları elimine etmek için θ ölçeğinden başka bir ölçeğe ihtiyaç duyulur. Bu durumda, bir sınavın paralel formlarına ait gerçek veya gözlenen puanları eşitlemek için farklı yöntemler kullanılır. Bu yöntemlerden bazıları Levine gerçek puan eşitleme (Levine, 1959), IRT gerçek puan eşitleme (Lord, 1965) gibi gerçek puanlar arasındaki istatistiksel ilişkiye, bazıları ise gözlenen puanlar arasındaki istatistiksel ilişkiye dayalıdır. Bu araştırmada kullanılan modeller ve kestirim yöntemleri, aynı θ yeteneğine sahip cevaplayıcıların hangi test formunu aldıkları göz önüne alınmaksızın ortak bir ölçek üzerinde aynı beklenen puana sahip oldukları varsayımına dayalı “*gerçek puan eşitlemeye*” dayalı olarak analiz edilmiştir. Bu yöntem MTK'nin gerçek puan tanımını
$$T(\theta_i) = \sum_j p_{ij}(\theta_i; a_j, b_j, c_j)$$
 kullanarak her iki formdan elde edilen gerçek puanlar arasındaki ilişkiyi ifade eder. Bu süreç üç aşamada gerçekleşir: İlk olarak form X üzerinde bir gerçek puan seçilir, daha sonra Newton-Raphson gibi bir iteratif yöntem kullanılarak bu gerçek puanla ilişkili yetenek düzeyi (θ_i) kestirilir, son olarak da bu yetenek düzeyi ile ilgili form Y üzerinde gerçek puan hesaplanır (Kolen & Brennan, 2004). Bu durum grafiksel olarak şekil 2.2'de gösterilmiştir:



Şekil 2.2. MTK gerçek puan eşitleme örneği

Şekil 2.2 incelendiğinde, yetenek düzeyi (θ) 1.3 olan bir cevaplayıcının form Y'den 24 puan, form X'ten ise 28 puan aldığı görülmektedir. Bu nedenle form Y ve form X'ten alınan puanların birbirine eşit olduğu düşünülür. Bu sürecin doğru olması için test karakteristik eğrileri aynı ölçekte olmalıdır. Bu nedenle eşitleme yapmadan önce bir ölçekleme basamağına gereksinim duyulur.

Gerçek puan eşitlemenin avantajı, grup 1 ve grup 2 yetenek dağılımlarını kestirmeksizin yürütülebilmesidir. Bununla birlikte teoride, gerçek puan eşitleme ilişkisi cevaplayıcı alt grupları için yerine getirilir. Dezavantajı ise, gerçek test etme koşullarında gerçek puanın gözlenememesidir. Kolen (1981) yaptığı çalışmada gerçek puan eşitlemenin gözlenen puan eşitlemeden daha kararlı sonuçlar verdiğini bulmuştur.

Bu tür eşitlemede eşitleme yöntemlerinin doğruluğunu belirlemek için cevaplayıcıların verdikleri cevaplardan elde edilen gerçek puanlar, eşitlenmiş puanlarla karşılaştırılır. Eşitleme sürecinde puanların eşitlenmesi, X formundan elde edilen ham puanları Y'den elde edilen puanlara dönüştürmek için bir eşitleme fonksiyonunun kullanılmasıyla gerçekleşir (Holland & Dorans, 2006).

2.6. EŞİTLEME SÜRECİNİN UYGULANMASI

Bu bölümde “denk olmayan gruplarda ortak madde deseni” için eşitleme sürecinin nasıl gerçekleştiği aşama aşama açıklanmıştır.

NEAT deseniyle MTK eşitleme madde kalibrasyonu, ölçek dönüştürmesi ve eşitleme basamaklarını içerir. Denk olmayan gruplar için madde parametreleri ayrı ayrı kestirildiğinde, kestirilen madde parametreleri yetenek dağılımındaki farklılıkları yansıtmaz ve bu yüzden de parametreler aynı ölçek üzerinde olmaz. Kestirim ve dönüştürme yoluyla parametre kestirimlerini aynı ölçeğe yerleştirme süreci “kalibrasyon” olarak adlandırılır (Kolen & Brennan, 2004). Kalibrasyon süreci boyunca tüm madde parametreleri aynı ölçeğe yerleştirilir ve daha sonra cevaplayıcıların yetenekleri kestirilir. Araştırmada kullanılan eşitleme desenine uygun olarak madde parametrelerini kestirmek için ayrı kestirim yöntemleri kullanılmıştır.

Araştırmada kullanılan ayrı kalibrasyon yönteminde her iki test formu için madde cevapları ayrı ayrı kalibre edilmiştir. İkinci form kalibre edilirken ortak maddeler için madde parametreleri ilk formun kalibrasyonunda oluşturulan değerlere sabitlenmiştir. Böylece ilk formun ölçeğine tüm madde parametreleri yerleştirilmiş ve ortak bir ölçek oluşturulmuştur. Bunun sonucunda da, grup 2'nin yetenek dağılımı ortalaması 0 standart sapması 1 olacak şekilde ölçeklenmiştir. Madde parametreleri kestirildikten ve ortak bir ölçeğe yerleştirildikten sonra bu değerler, MTK gerçek puan eşitleme yöntemi için test formları üzerinde puanları eşitlemede kullanılmıştır. Bu amaçla bir yetenek ölçeği üzerindeki değerler, bu değerlere denk gelen gerçek puanlara dönüştürülmüştür. Daha sonra ise farklı formlardan elde edilen gerçek puanlar lineer dönüştürme yoluyla eşitlenmiştir. Bu anlamda MTK gerçek puan eşitleme gözlenen puanların değil, gerçek puanların eşitlenmesini ifade eder. Böylece yeni formdan elde edilen ham puanlar eski formdan elde edilen ham puanlara dönüştürülmüştür.

2.7. VERİLERİN ANALİZİ VE DEĞERLENDİRME ÖLÇÜTÜ

Bu araştırmada kalibrasyon yöntemlerinden ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılmıştır. Stocking-Lord ve Haebara yöntemleri genellikle benzer ve yeterli sonuçlar ürettiği ve Stocking-Lord, Haebara'ya nazaran test etme uygulamalarında daha yaygın olarak kullanılan yöntemlerden birisi olduğu için, bu çalışmada Stocking-Lord yöntemi tercih edilmiştir (Way & Tang, 1991; Lee & Fitzpatrick, 2008; He, 2011). Madde ve yetenek parametrelerinin kestirilmesi PARSCALE 4.1 (Muraki & Bock, 2003) ile, ayrı kalibrasyon için test eşitleme ve ölçekleme IRTEQ (Han, 2007) ile yapılmıştır. Araştırmada bilinen evren parametreleri ile çalışıldığından, madde ve yetenek parametreleri expected a posteriori (EAP) ile kestirilmiştir.

Araştırmada daha kararlı kestirimler elde etmek amacıyla her bir koşul için ve her bir yöntem için veri türetme ve eşitleme süreci 50 kez tekrarlanmıştır. Eşitleme üzerine yapılan araştırmalar incelendiğinde tekrar sayısının 50 olarak alındığı çok sayıda çalışmanın olduğu (Liou & Bond, 1985; Livingston, 1993a, 1993b; Cohen & Kim, 1998; Hanson & Beguin, 1999, 2002; Turhan, Tong & Um, 2007; Hu ve diğerleri, 2008) ve her bir koşul için farklı eşitleme yöntemlerinden elde edilen sonuçları karşılaştırmak amacıyla 50 tekrarın yeterli olduğu (Harwell, Stone, Hsu & Kirisci, 1996; Hanson & Beguin, 2002; Hu ve diğerleri, 2008) görülmüştür. Ayrıca Harwell ve diğerleri (1996) MTK ile yapılan simülasyon çalışmalarında 25 tekrarın yeterli olduğunu ifade etmişlerdir.

Eşitleme sonuçlarının doğruluğunu değerlendirmek için pek çok ölçüt bulunmaktadır ve değerlendirme için birden fazla ölçütün kullanılması önerilmektedir. (Harris & Crouse, 1993). Bu doğrultuda sıklıkla kullanılan değerlendirme ölçütleri RMSE, MSE, RMSD, BIAS ve SEE'dir Bu çalışmada BIAS (Eşitleme yanlılığı) ve RMSE (Root mean square error-Eşitleme hatası) kullanılmıştır (Maris, 1999; van der Linden, 2006). BIAS ve RMSE eşitlemenin niteliğini değerlendirmede kullanılan iki önemli indekstir. Bu ölçütler kestirilen parametre değerlerinin gerçek değere ne kadar yakın olduğunu gösterir. Bu ölçütlerin yorumlanması kolay olduğundan pek çok çalışmada da sıklıkla

kullanıldığı görülmektedir (Kim & Lee, 2006; Hou, 2007; Sinharay & Holland, 2007; Cao, 2008; Wang, Lee, Brennan & Kolen, 2008; Walker & Kim, 2009).

BIAS: Sistemik hata olarak da ifade edilen yanlılık, kestirilmiş parametre değerleri ile gerçek parametre değerleri arasındaki farkın toplamının tekrar sayısına bölünmesiyle elde edilir. Bu değerın matematiksel ifadesi şu şekildedir:

$$BIAS(\tau_j) = \frac{\sum_{r=1}^R (\tau_{jr}^* - \tau_j)}{R}$$

τ_j : j parametresinin gerçek değeri

τ_{jr}^* : tekrar edilen veri seti ($r=1, \dots, R$) için j parametresinin kestirilen değeri

R: tekrar sayısı

Yukarıdaki formül BIAS değerinin matematiksel ifadesini göstermektedir ve bu formül sırasıyla a ve b parametreleri (ayırıcılık ve güçlük parametreleri), θ (yetenek parametresi) için ayrı ayrı hesaplanmış ve daha sonra yapılan tekrarlardan (replication) elde edilen BIAS değerlerinin ortalaması alınmıştır.

$$\sum_{i=1}^I BIAS(\tau_i) / I$$

I: rapor edilen parametreye bağlı olarak kişi veya madde sayısı

Ortalama BIAS değeri eşitleme sonuçlarının doğruluk düzeyini gösterir. Bu değer küçüldükçe yanlılık azalır ve daha doğru kestirimler elde edilir. Yanlılık, sistemik eşitleme hatasını tanımlamada sıklıkla kullanılan bir indekstir (Harris & Crouse, 1993).

RMSE: Eşitleme hatası ise kestirilmiş parametre değerleri ile gerçek parametre değerleri arasındaki farkın kareleri toplamının tekrar sayısına oranının karekökünü gösterir ve bu değer toplam hata olarak da adlandırılır ve aşağıdaki şekilde hesaplanır:

$$RMSE(\tau_j) = \sqrt{\frac{\sum (\tau_{jr}^* - \tau_j)^2}{R}}$$

τ_j : j parametresinin gerçek değeri

τ_{jr}^* : tekrar edilen veri seti ($r=1, \dots, R$) için j parametresinin kestirilen değeri

R: tekrar sayısı

Yukarıdaki eşitlik kestirilmiş parametrelerin gerçek parametrelere uyma derecesini gösterir. Bu çalışmada, yapılan tekrarlardan (50 kez) elde edilen RMSE değerlerinin ortalaması alınmıştır. Elde edilen sonuca göre, ortalama RMSE değeri küçüldükçe, daha az hatalı ve dolayısıyla daha doğru eşitleme sonuçları elde edildiği söylenebilir. RMSE indeksi, toplam eşitleme hatasını (random ve sistematik hatanın toplamı) ifade eder. Araştırmada yapılan tekrarlar karşısında parametre veya yetenek kestirimlerindeki değişkenliği değerlendirmek için 50 kez yapılan tekrarlardan elde edilen RMSE değerlerinin ortalaması alınmıştır.

$$\sum_{i=1}^I RMSE(\tau_i) / I$$

I: rapor edilen parametreye bağlı olarak kişi veya madde sayısı

Gerçek puan açısından ele alındığında ise, BIAS (yanlılık) kestirilmiş beklenen toplam puanlar ile gerçek beklenen toplam puanlar arasındaki farkı gösterir. Bu değer şu şekilde hesaplanır:

$$BIAS_r = \frac{\sum_{k=1}^n (E(X_k | \hat{\theta}_k) - E(X_k | \theta_k))}{N}$$

$E(X_k | \hat{\theta}_k)$: kestirilmiş parametreler kullanılarak hesaplanan “kestirilen beklenen toplam puanlar (eşitlenmiş puanlar)”

$E(X_k | \theta_k)$: popülasyon parametreleri kullanılarak cevaplayıcıların Form 2’deki maddelere verdiği cevaplardan elde edilen “gerçek beklenen toplam puanlar”

N: Cevaplayıcı sayısı

Araştırmada kapsamında, popülasyon parametreleri $E(X_k | \theta_k)$ kullanılarak hesaplanan form 2’deki tüm maddeler için cevaplayıcıların beklenen toplam puanları, kestirilen parametreler ($E(X_k | \hat{\theta}_k)$) kullanılarak hesaplanan toplam puanlar ile karşılaştırılmıştır. Her bir tekrar için BIAS değerleri hesaplandıktan sonra 50 tekrardan elde edilen BIAS

değerlerinin ortalaması alınmıştır. Ortalama BIAS değerleri, hem eşitlemenin yönünü hem de doğruluğunu gösterir.

Eşitleme hatası (RMSE) ise kestirilen beklenen toplam puanların gerçek beklenen toplam puanlara uyma derecesini gösterir. Bu değer aşağıdaki şekilde hesaplanır:

$$RMSE_r = \sqrt{\frac{\sum_{k=1}^n (E(X_k | \hat{\theta}_k) - E(X_k | \theta_k))^2}{N}}$$

Araştırmada her bir tekrar için RMSE değeri hesaplandıktan sonra, 50 tekrardan elde edilen RMSE değerlerinin ortalaması alınmıştır. Ortalama RMSE ve BIAS değerleri küçüldükçe daha doğru eşitleme sonuçları elde edilir. İyi bir yöntem küçük yanlılığa ve küçük RMSE değerine sahip olmalıdır (Chu & Kamata, 2003). Yukarıda da ifade edildiği gibi, araştırmada RMSE ve BIAS değerleri hem parametreler hem de gerçek puanlar için hesaplanmıştır.

Başarılı bir MTK eşitleme, madde ve yetenek parametrelerinin doğru bir şekilde kestirilmesine bağlıdır. Livingston, Dorans ve Wright (1990) eşitleme çalışmalarında yanlılığın araştırılmasının RMS değerinin neden yüksek olduğunu ortaya çıkarmada yardımcı olduğunu ifade etmişlerdir. Bu nedenle araştırmada RMSE ve BIAS değerleri birlikte ele alınarak değerlendirilmiştir.

BÖLÜM III

BULGULAR VE YORUM

Bu bölümde araştırmanın bulguları, araştırma soruları çerçevesinde ele alınan eşitleme yanlılığı (BIAS) ve eşitleme hatası (RMSE) ölçütlerine göre incelenmiştir.

3.1. ALT PROBLEM 1'E İLİŞKİN BULGULAR VE YORUMLAR

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak üç parametrelili lojistik modele uyumlu veriler ile kestirilen madde parametrelerinin yanlılık ve eşitleme hatası;

- a) *Test uzunluğuna*
- b) *Örneklem büyüklüğüne*
- c) *Yetenek dağılımına göre nasıl değişmektedir?*
- d) *Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğu, örneklem büyüklüğü ve yetenek dağılımına göre en iyi sonucu vermektedir?*

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak üç parametrelili lojistik modele uyumlu veriler ile kestirilen madde parametrelerinin yanlılık ve eşitleme hatası Tablo 3.1'de verilmiştir:

Tablo 3.1. 3PLM ile kestirilen madde parametrelerinin simülasyon koşullarına göre yanlılık (BIAS) ve eşitleme hataları (RMSE)

Yetenek dağılımı	Örneklem büyüklüğü	Test uzunluğu	Değerlendirme ölçütü	3PLM Madde parametreleri						
				a			b			c
				OO	OS	SL	OO	OS	SL	OO-OS-SL
G1:N(0,1) G2:N(0.5,1)	1000	30	BIAS	-0.107	-0.048	-0.075	-0.116	-0.106	-0.082	0.016
			RMSE	0.254	0.273	0.249	0.267	0.270	0.254	0.072
		60	BIAS	0.025	-0.050	-0.029	-0.112	-0.105	-0.06	-0.005
			RMSE	0.216	0.250	0.224	0.253	0.266	0.224	0.069
		80	BIAS	0.013	0.028	0.012	-0.108	-0.093	-0.006	-0.006
			RMSE	0.202	0.226	0.205	0.251	0.254	0.210	0.057
	3000	30	BIAS	0.031	0.009	-0.005	-0.112	-0.109	-0.043	-0.005
			RMSE	0.147	0.161	0.139	0.210	0.223	0.147	0.049
		60	BIAS	0.018	0.025	0.009	-0.067	-0.068	0.001	0.008
			RMSE	0.139	0.162	0.138	0.209	0.225	0.146	0.047
		80	BIAS	-0.070	-0.091	-0.120	-0.056	-0.054	-0.03	-0.022
			RMSE	0.124	0.158	0.133	0.206	0.223	0.138	0.042
G1:PÇ G2:N(0,1)	1000	30	BIAS	-0.270	-0.213	-0.372	1.310	1.260	1.657	-0.001
			RMSE	0.326	0.302	0.400	1.372	1.322	1.714	0.058
		60	BIAS	-0.249	-0.207	-0.335	1.180	1.181	1.370	0.008
			RMSE	0.321	0.254	0.351	1.205	1.207	1.395	0.083
		80	BIAS	-0.112	-0.090	-0.129	1.023	1.018	1.296	0.003
			RMSE	0.281	0.223	0.269	1.055	1.055	1.334	0.063
	3000	30	BIAS	-0.252	-0.179	-0.383	1.569	1.529	1.943	-0.003
			RMSE	0.279	0.263	0.395	1.581	1.544	1.956	0.037
		60	BIAS	-0.29	-0.257	-0.269	1.362	1.376	1.512	0.001
			RMSE	0.321	0.362	0.306	1.410	1.482	1.567	0.047
		80	BIAS	-0.152	-0.154	-0.192	1.184	1.221	1.365	0.010
			RMSE	0.199	0.287	0.227	1.213	1.323	1.392	0.039
G1:NÇ G2:N(0,1)	1000	30	BIAS	-0.542	-0.563	-0.536	-1.292	-1.278	-1.557	-0.009
			RMSE	0.565	0.591	0.559	1.327	1.340	1.585	0.053
		60	BIAS	-0.497	-0.468	-0.513	-0.916	-0.939	-1.235	0.005
			RMSE	0.542	0.493	0.527	1.027	1.035	1.344	0.051
		80	BIAS	-0.480	-0.457	-0.450	-0.793	-0.909	-1.209	0.002
			RMSE	0.513	0.479	0.465	0.871	0.964	1.279	0.049
	3000	30	BIAS	-0.216	-0.204	-0.247	1.426	1.435	1.928	0.006
			RMSE	0.165	0.197	0.275	1.442	1.462	1.948	0.064
		60	BIAS	-0.210	-0.105	-0.224	1.336	1.291	1.479	-0.005
			RMSE	0.251	0.244	0.267	1.383	1.368	1.537	0.051
		80	BIAS	-0.134	-0.101	-0.158	1.124	1.125	1.241	0.001
			RMSE	0.191	0.273	0.225	1.159	1.185	1.277	0.038

OO: ortalama-ortalama, OS:ortalama-standart sapma, SL:Stocking-Lord

a. Test uzunluđuna gre yntemlerin hata ve yanlılıkları nasıl deđiřmektedir?

Arařtırmada ele alınan yntemler test uzunluđuna gre incelendiđinde, genel olarak tm yntemlerde standart normal dađılımda ve arpık dađılımlarda hem 1000 hem de 3000 kiřilik rneklemelerde test uzunluđu arttıka, hata ve yanlılık deđerlerinin dřtđ ve 0'a yaklařma eđiliminde olduđu grlmektedir. Bu aıdan, genel olarak hem standart normal dađılıma sahip hem de arpık dađılımlarda her iki rneklem byklđ iin test uzunluđu arttıka eřitlemenin dođruluđunun da arttıđı grlmřtr. Arařtırmada elde edilen bu sonu, Fitzpatrick ve Yen (2001) ile Vale (1986) tarafından yapılan arařtırma sonularıyla benzerlik gstermektedir. Yapılan bu arařtırmalarda daha uzun testlerin daha dřk hataya yol atıđı bulunmuřtur. Ayrıca madde sayısı arttıka da yntemlerin birbirine yakın deđerler verdiđi grlmřtr. Elde edilen bu sonu Han (2008) tarafından 3PLM'ye uyumlu veriler zerinde yapılan alıřmada elde edilen sonula benzerlik gstermektedir.

b. rneklem byklđne gre yntemlerin hata ve yanlılıkları nasıl deđiřmektedir?

Yntemler rneklem byklđ aısından karřılařtırıldıđında, gruplar arasındaki yetenek farklarının az olduđu standart normal dađılıma sahip rneklemelerde, rneklem byklđ arttıka RMSE ve BIAS deđerlerinin genel olarak dřtđ ve 0'a yaklařma eđiliminde olduđu gzlenmektedir. Aynı Őekilde gruplar arasındaki yetenek farklarının fazla olduđu pozitif ve negatif arpık dađılımlarda da genel olarak rneklem byklđ arttıka RMSE ve BIAS deđerlerinin dřtđ gzlenmektedir. Sonu olarak, byk rneklemeler ve daha uzun testler kullanıldıđında yntemlerin daha dođru sonular verdiđi grlmřtr. Wingersky (1983) ve Zhang (2010)'in 3PLM zerinde yaptıkları alıřmalarda benzer sonulara ulařmıřtır.

 parametrelili modelde madde kestirimlerinin dođruluđu hangi parametrenin kestirildiđine gre deđiřmektedir. Standart normal dađılımda a parametresi iin rneklem byklđ arttıka yanlılık deđerlerinin dřtđ gzlenirken, aynı durum b ve c parametreleri iin geerli deđildir. Bu durumun sonucu olarak rneklem byklđnn a parametresi zerinde diđer parametrelere gre daha byk ve olumlu ynde bir etkiye sahip olduđu grlrken, b ve c parametrelerinin kestiriminde daha az bir etkiye sahip olduđu sylenebilir. Arařtırmadan elde edilen bu sonu Swaminathan

ve Gifford (1983), Wingersky ve Lord (1984), Skaggs ve Stevenson (1989)'un 3PLM'ye göre türetilen veriler üzerinde yaptıkları araştırmada buldukları sonuç ile benzerlik göstermektedir. Bununla birlikte standart normal dağılımda a parametresi için RMSE değerlerinin tüm yöntemlerde düşük ve benzer performans gösterdiği görülmüştür. Araştırmadan elde edilen bu sonuç Keller, Kim, Nering ve Keller (2007) tarafından yapılan araştırmada bulunan sonuç ile benzerlik göstermektedir. Çarpık dağılımlarda a parametresi için RMSE ve BIAS değerlerinin 1000 kişilik örneklemlere nazaran 3000 kişilik örneklemlerde daha küçük olduğu görülürken, b parametresi için tam tersi bir durum söz konusudur. Yetenek dağılımı normal olduğunda çarpık olduğu durumlara göre madde parametreleri en iyi sonuçları vermiştir. Elde edilen bu sonuç Turhan ve diğerleri (2007) tarafından 3PLM'ye uygun veriler üzerinde yapılan çalışmada elde edilen sonuçla tutarlılık göstermektedir.

Araştırmanın bu bölümünden elde edilen sonuçlar hangi eşitleme yönteminin kullanıldığı göz önüne alınmaksızın incelendiğinde, test uzunluğu ve örneklem büyüklüğü arttıkça daha doğru sonuçlara ulaşıldığını göstermektedir. Araştırmadan elde edilen bu sonuç Yang ve Houang (1996), Bastari (2000), Hulin, Drasgow ve Parsons (1983), Skaggs ve Stevenson (1989), Zeng (1991), Kaskowitz ve De Ayala (2001) tarafından yapılan çalışmadan elde edilen sonuçla tutarlılık göstermektedir. Aynı şekilde Parshall, Houghton ve Kramley (1995) yaptıkları araştırmada örneklem büyüklüğü azaldıkça standart hatanın arttığını bulmuşlardır. Bu araştırmada sadece madde sayısı değişken olarak alınmış, ortak madde sayısı sabit tutulmuştur. Harris (1993) yaptığı çalışmada ortak maddelerin sayısının az ya da çok olmasının eşitleme sonuçlarını etkilemediğini bulmuştur.

c. Yetenek dağılımına göre yöntemlerin hata ve yanlılıkları nasıl değişmektedir?

Gruplar yetenek dağılımları açısından karşılaştırıldığında, diğer faktörler sabit tutulduğunda gruplar arasındaki yetenek farklarının az olduğu standart normal dağılıma sahip örneklemlerin, yetenek farklarının çok olduğu ve çarpık dağılıma sahip örneklemlerden daha az hatalı ve yanlı sonuçlar verdiği görülmektedir. Araştırmada elde edilen bu sonuç, Bastari (2000), Kim ve Lee (2006) tarafından yapılan araştırmada elde edilen sonuçla benzerlik göstermektedir. Bu konuda yapılan araştırmalar, gruplar

arasındaki farkların hata ve yanlılık üzerinde büyük etkiye sahip olduğunu göstermiştir (Ngudratoke, 2009; Wu, Huang, Nuh & Harris, 2009). Gruplar arasındaki yetenek farkları büyük olduğunda NEAT desenini kullanan tüm eşitleme yöntemleri büyük yanlılık ve hata üretir (Ngudratoke, 2009). Bu araştırmadan elde edilen sonuca benzer olarak, Turhan ve diğerleri (2007) tarafından 3PLM'ye uyumlu veriler üzerinde yapılan çalışmada tüm madde parametreleri için RMSE ve BIAS değerlerinin çarpık dağılımlarda standart normal dağılıma göre genel olarak daha büyük olduğuna dair bulunan sonuçla benzerlik göstermektedir. Bununla birlikte çarpık dağılımlarda madde parametrelerine ait hata ve yanlılık değerlerinin daha büyük olduğu bulunmuştur. Elde edilen bu sonuç Turhan ve diğerleri (2007)'nin yaptıkları araştırmada elde ettikleri sonuç ile tutarlılık göstermektedir. Kolen (1984) tarafından yapılan çalışmada da iki formun çarpıklığı farklı olduğunda eşitleme yöntemlerinin farklı sonuçlar verdiği bulunmuştur.

Negatif ve pozitif çarpık yetenek dağılımları kendi içerisinde karşılaştırıldığında, 3PLM için negatif çarpık dağılımın RMSE ve BIAS değerlerinin pozitif çarpık dağılımdan daha büyük olduğu görülmektedir. Araştırmadan elde edilen bu sonuç, Keller ve Keller (2008) tarafından yapılan çalışmada negatif çarpık yetenek dağılımında yanlılık değerlerinin diğer dağılımlardan daha büyük bulunmasına ilişkin olarak elde edilen sonuçla benzerlik göstermektedir. Dağılımın pozitif veya negatif çarpık olduğu durumlarda, madde parametrelerini kestirmek problemlidir ve negatif çarpık dağılım koşulunda yöntemlerin zayıf performans göstermesi kestirimin problemlidir yapısından kaynaklanabilir.

d. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğu, örneklem büyüklüğü, yetenek dağılımına göre en iyi sonucu vermektedir?

Yöntemler kendi içinde örneklem büyüklüğü açısından karşılaştırıldığında, 1000 ve 3000 kişilik örneklerde a ve b parametreleri için en küçük RMSE ve BIAS değerine sahip yöntemin SL yöntemi olduğu görülmektedir. Benzer şekilde, Hanson ve Beguin (2002) yaptıkları araştırmada 3PLM kullanıldığında Stocking-Lord ve Haebara

yöntemlerinin, ortalama-ortalama ve ortalama-standart sapma yöntemlerinden daha düşük hata ürettiğini bulmuşlardır.

Yöntemler test uzunluğu ve yetenek dağılımı açısından karşılaştırıldığında ise, standart normal dağılıma sahip gruplarda en iyi yöntemin SL olduğu görülmektedir. Bununla birlikte negatif çarpık dağılımlarda 1000 kişilik örneklemlerde 30 ve 60 maddelik testler için a parametresi göz önüne alındığında en iyi yöntemin SL yöntemi olduğu görülürken, 80 maddelik test için OS yöntemi olduğu görülmektedir. b parametresi içinse en iyi yöntemin OO yöntemi olduğu görülmektedir. 3000 kişilik örneklemlerde ise genel olarak her iki parametre içinde en az hataya ve yanlılığa sahip yöntemin OO yöntemi olduğu görülmektedir. Pozitif çarpık yetenek dağılımına sahip örneklemlerde ise genel olarak a ve b parametreleri için OO ve OS yöntemlerinin SL yönteminden daha az yanlı ve hatalı sonuçlar verdiği görülmektedir. Keller ve diğerleri (2004) yaptıkları araştırmada OS yönteminin SL ve OO yöntemlerinden daha doğru sonuçlar verdiğini bulmuşlardır. Keller ve Keller (2008) 3PLM'yi kullanarak yaptıkları araştırmada çarpık yetenek dağılımına sahip gruplarda OS yönteminin en çok hataya sahip yöntem olduğunu bulmuşlardır. Madde parametreleri genel olarak incelendiğinde ise a ve b parametreleri için en az hataya sahip yöntemin SL, en çok hatalı yöntemin ise OS yöntemi olduğu görülmektedir. Elde edilen bu sonuç Han (2008) tarafından yapılan araştırmada elde edilen sonuç ile tutarlılık göstermektedir.

Araştırmadan elde edilen sonuçlar yöntemler açısından genel olarak incelendiğinde ise üç parametrelili lojistik model kullanılarak veri türetildiğinde birkaç istisna dışında SL yönteminin OO ve OS yöntemlerinden daha doğru sonuçlar ürettiği bulunmuştur. Üç parametrelili lojistik modele uyumlu veriler için en az hata içeren yöntem genel olarak SL yöntemi olurken, en fazla hata içeren yöntem ise OS yöntemidir. Araştırmadan elde edilen bu sonuç, Kim ve Lee (2006), Kim ve Kolen (2006), Baker ve Al-Karni (1991), Speron (2009), Hanson ve Beguin (2002) tarafından yapılan araştırma sonuçlarıyla benzerlik göstermektedir. Bunun nedeni olarak ise ortalamaların standart sapmaya göre daha kararlı bir parametre olması gösterilmektedir. Karakteristik eğri kestirim yöntemleri (Stocking-Lord ve Haebara), madde parametrelerinin kestirilmesinde hataya daha dirençlidirler (Kaskowitz & De Ayala, 2001). Yapılan araştırmalarda SL yönteminin daha kararlı olduğu ve diğer MTK'ye dayalı yöntemlerden daha iyi

performans gösterdiği bulunmuştur (Kolen, 1981). Bu nedenle, SL yöntemi uygulamada en çok tercih edilen yöntemdir (Hills, Subhiyah & Hirsch, 1988; Kaskowitz & De Ayala, 2001; Kolen & Brennan, 2004). Ogasawara (2001) ise yaptığı çalışmada, hem iki hem de üç parametrelili modeller için SL yönteminin OO ve OS yöntemlerinden daha az hatalı ve yanlış sonuçlar verdiğini bulmuştur.

3.2. ALT PROBLEM 2'YE İLİŞKİN BULGULAR VE YORUMLAR

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak iki parametrelili lojistik modele uyumlu veriler ile kestirilen madde parametrelerinin yanlışlık ve eşitleme hatası;

a. Test uzunluğuna

b. Örneklem büyüklüğüne

c. Yetenek dağılımına göre nasıl değişmektedir?

d. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğu, örneklem büyüklüğü, yetenek dağılımına göre en iyi sonucu vermektedir?

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak iki parametrelili lojistik modele uyumlu veriler ile kestirilen madde parametrelerinin yanlışlık ve eşitleme hatası Tablo 3.2'de verilmiştir:

a. Test uzunluğuna göre yöntemlerin hata ve yanlışlıkları nasıl değişmektedir?

Tablo 3.2 incelendiğinde, test uzunluğu arttıkça, standart normal dağılım ve çarpık yetenek dağılımlarında hem 1000 hem de 3000 kişilik örneklemelerde (pozitif çarpık dağılım 1000 kişilik örneklemeler dışında), test uzunluğu 30'dan 60'a çıktıkça hata ve yanlışlık değerleri düşerken, 80 maddelik testlerde test uzunluğunun kestirim yöntemleri üzerinde herhangi bir etkisinin olmadığı görülmektedir. Elde edilen bu sonuca göre, test uzunluğunun 80 maddelik testlerde hata ve yanlışlık değerleri üzerinde olumlu bir etkiye sahip olduğu söylenemez.

Tablo 3.2. 2PLM ile kestirilen madde parametrelerinin simülasyon koşullarına göre yanlılık (BIAS) ve eşitleme hataları (RMSE)

Yetenek dağılımı	Örneklem büyüklüğü	Test uzunluğu	Değerlendirme ölçütü	2PLM					
				Madde parametreleri					
				a			b		
				OO	OS	SL	OO	OS	SL
G1:N(0,1) G2:N(0.5,1)	1000	30	BIAS	-0.017	-0.023	-0.024	-0.096	-0.095	-0.027
			RMSE	0.119	0.136	0.122	0.159	0.158	0.095
		60	BIAS	-0.006	0.013	-0.007	-0.091	-0.094	-0.016
			RMSE	0.108	0.152	0.106	0.147	0.166	0.085
		80	BIAS	-0.043	-0.155	-0.107	-0.082	-0.070	-0.048
			RMSE	0.16	0.228	0.182	0.177	0.179	0.127
	3000	30	BIAS	-0.033	-0.022	-0.032	-0.090	-0.096	-0.013
			RMSE	0.083	0.085	0.080	0.125	0.126	0.045
		60	BIAS	-0.018	0.005	-0.023	-0.078	-0.087	-0.008
			RMSE	0.061	0.064	0.062	0.121	0.121	0.045
		80	BIAS	-0.023	-0.008	-0.026	-0.103	-0.111	-0.029
			RMSE	0.079	0.12	0.11	0.136	0.156	0.075
G1:PÇ G2:N(0,1)	1000	30	BIAS	-0.359	-0.365	-0.362	1.604	1.574	1.89
			RMSE	0.368	0.404	0.373	1.662	1.64	1.958
		60	BIAS	-0.342	-0.288	-0.337	1.265	1.289	1.455
			RMSE	0.353	0.319	0.348	1.355	1.565	1.602
		80	BIAS	-0.254	-0.237	-0.235	0.829	0.82	0.96
			RMSE	0.273	0.262	0.254	0.932	0.92	1.091
	3000	30	BIAS	-0.473	-0.413	-0.398	-1.082	-1.088	-1.252
			RMSE	0.488	0.472	0.434	1.238	1.234	1.418
		60	BIAS	-0.342	-0.33	-0.306	-0.696	-0.702	-0.818
			RMSE	0.37	0.362	0.336	1.01	1.003	1.174
		80	BIAS	-0.453	-0.481	-0.422	1.207	1.315	1.566
			RMSE	0.459	0.487	0.43	1.266	1.376	1.634
G1:NÇ G2:N(0,1)	1000	30	BIAS	-0.319	-0.341	-0.325	-0.92	-0.891	-1.115
			RMSE	0.344	0.365	0.346	1.104	1.128	1.362
		60	BIAS	-0.361	-0.336	-0.318	-0.488	-0.487	-0.557
			RMSE	0.375	0.39	0.339	0.865	0.88	0.994
		80	BIAS	-0.329	-0.33	-0.302	-0.577	-0.574	-0.666
			RMSE	0.339	0.345	0.312	0.852	0.858	1.01
	3000	30	BIAS	-0.398	-0.428	-0.386	-1.437	-1.444	-1.625
			RMSE	0.400	0.43	0.388	1.45	1.468	1.638
		60	BIAS	-0.377	-0.427	-0.363	-1.074	-1.058	-1.275
			RMSE	0.385	0.448	0.372	1.251	1.247	1.49
		80	BIAS	-0.491	-0.63	-0.541	-1.245	-1.205	-1.556
			RMSE	0.516	0.649	0.566	1.275	1.240	1.589

OO: ortalama-ortalama, OS:ortalama-standart sapma, SL:Stocking-Lord

b. Örneklem büyüklüğüne göre yöntemlerin hata ve yanlılıkları nasıl değişmektedir?

İki parametrelili model için örneklem büyüklüğü açısından sonuçlar incelendiğinde, standart normal dağılımlarda hem a hem de b parametresi için örneklem büyüklüğünün yanlılık (BIAS) ve hata üzerinde olumlu bir etkiye sahip olduğu görülürken, pozitif ve negatif çarpık dağılımlarda ise herhangi bir etkisinin olmadığı görülmektedir. Standart normal dağılıma sahip gruplarda birkaç istisna dışında örneklem büyüklüğü arttıkça

yöntemlerin hata ve yanlılık değerlerinin de düştüğü görülmüştür. Araştırmadan elde edilen bu sonuç Arai ve Mayekawa (2011) ile Speron (2009) tarafından 2PLM'ye uyumlu veriler üzerinde yapılan araştırmada elde edilen sonuç ile benzerlik göstermektedir.

İki parametrelili model için örneklem büyüklüğü açısından sonuçlar genel olarak incelendiğinde, örneklem büyüklüğü arttıkça hataların (RMSE) düştüğü gözlenirken, örneklem büyüklüğünün yanlılık üzerinde herhangi bir etkisinin olmadığı gözlenmiştir. Araştırmada elde edilen bu sonuç Wu ve diğerleri (2009)'nin yaptığı araştırmada elde ettiği sonuç ile benzerlik göstermektedir. Kolen ve Brennan (2004) örneklem büyüklüğünün yanlılık üzerinde herhangi bir etkiye sahip olmadığını ve örneklem büyüklüğü arttıkça yanlılığın azalmadığını ifade etmişlerdir. Sistemik hata örneklem büyüklüğünden doğrudan etkilenmez. Aynı şekilde Wu ve diğerleri (2009) yaptıkları araştırmalarda örneklem büyüklüğünün yanlılığı etkilemediğini bulmuşlardır. Han (2007) ise yaptığı çalışmada, tüm parametreler için örneklem büyüklüğünün RMSE'nin tersine yanlılığı azaltmadığını bulmuştur. Yanlılık istatistiği, parametre kestirimlerinin sistemik hatasının bir ölçüsü olduğu için, örneklem büyüklüğünün sistemik hatayla ilişkisi yoktur ve sistemik hata örneklem büyüklüğünden doğrudan etkilenmez. Araştırmada 3000 kişilik örneklemelere nazaran 1000 kişilik örneklemelerde yanlılık değerlerinin daha düşük olduğu görülmüştür. Araştırmada örneklem büyüklüğünün yanlılık üzerinde herhangi bir etkiye sahip olmadığı görülürken, birkaç istisna dışında hata üzerinde olumlu bir etkiye sahip olduğu görülmüştür.

c. Yetenek dağılımına göre yöntemlerin hata ve yanlılıkları nasıl değişmektedir?

Araştırmada ele alınan yöntemler yetenek dağılımları açısından incelendiğinde ise yetenek dağılımlarının benzer olduğu standart normal dağılıma sahip grupların, yetenek dağılımları arasındaki farkların fazla olduğu pozitif ve negatif çarpık dağılımlara göre daha az hatalı ve yanlı sonuçlara yol açtığı bulunmuştur. Araştırmada elde edilen bu sonuç Kim ve Lee (2006)'nin araştırma sonucuyla benzerlik göstermektedir. Bununla birlikte çarpık dağılımlar kendi içinde karşılaştırıldığında, pozitif çarpık yetenek dağılımlarında hata ve yanlılık değerleri negatif çarpık dağılımlardan daha yüksektir.

d. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğu, örneklem büyüklüğü, yetenek dağılımına göre en iyi sonucu vermektedir?

Tüm koşullar için yöntemler birbirleriyle karşılaştırıldığında, normal yetenek dağılımına sahip gruplarda b parametresi için en düşük RMSE ve BIAS değerlerine sahip yöntemin SL yöntemi olduğu görülürken, a parametresi için ise genel olarak OO yöntemi olduğu görülmektedir. Gruplar arasındaki farkların fazla olduğu pozitif ve negatif çarpık dağılımlar için, moment yöntemlerinin (OO ve OS) SL yönteminden daha iyi sonuçlar verdiği görülmektedir. Bununla birlikte, bu tür dağılımlarda test uzunluğuna göre yöntemler arasındaki farkların oldukça küçük olduğu da görülmektedir. Bu sonuç Kim ve Kolen (2006)'nin yaptığı araştırmada yöntemler arasındaki farkların oldukça küçük olduğuna dair elde ettiği sonuç ile paralellik göstermektedir.

Yöntemler birbirleriyle karşılaştırıldığında, tüm koşullarda genel olarak 2PL model için yöntemler arasındaki farklılıkların oldukça küçük olduğu gözlenmiştir. Araştırmada elde edilen bu sonuç Wittington (1986) tarafından bulunan sonuç ile benzerlik göstermektedir. Bununla birlikte, iki parametrelili modele uyumlu veriler için yöntemler karşılaştırıldığında, en az hataya ve yanlılığa sahip yöntemin OO olduğu görülürken, en fazla hataya ve yanlılığa sahip yöntemin ise OS olduğu görülmektedir. Baker ve Al-Karni (1991) yaptıkları çalışmada OO yönteminin OS yöntemine göre daha iyi eşitleme sonuçlarına yol açtığını bulmuşlardır. Araştırmadan elde edilen bu sonuç Ogasawara (2001) tarafından yapılan iki parametrelili modeller için en iyi yöntemin SL olduğu bulgusuyla çelişmekte, en çok hataya sahip yöntemin ise OS olduğuna dair araştırma bulgusuyla benzerlik göstermektedir. Way ve Tang (1991) tarafından yapılan çalışmada ise OS yönteminin OO ve SL yöntemlerinden daha az kararlı kestirimler sağladığı bulunmuştur. Bununla birlikte OS yöntemi, eşitlemede en yaygın olarak kullanılan yöntemlerden biridir (Kolen & Brennan, 2004).

3.3. ALT PROBLEM 3'E İLİŞKİN BULGULAR VE YORUMLAR

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak elde edilen madde parametrelerinin yanlılık ve eşitleme hatası model türüne göre nasıl değişmektedir?

Araştırmadan elde edilen sonuçlar model türüne göre incelendiğinde, yapılan çalışmalarda çarpık dağılımlara sahip test formlarını eşitlerken şans parametresinin de olduğunu üç parametrelili modeli kullanmanın uygun olduğu bulunmuştur (Yang & Houang, 1996). Araştırmalarda elde edilen bu sonuç, bu araştırmadan elde edilen sonuçla benzerlik göstermektedir. Bu çalışma, çarpık dağılımlar için 3PLM'nin 2PLM'den daha iyi sonuçlar verdiğini göstermektedir. Bununla birlikte iki ve üç parametrelili model için yetenek dağılımları açısından hata ve yanlılık değerleri incelendiğinde, grupların yetenek dağılımları farklılaştıkça daha yanlı ve hatalı sonuçlar bulunmuştur. Araştırmada elde edilen bu sonuç, eşitleme yöntemlerinin etkililiğini incelemek için yapılan araştırmalarda grupların yetenek dağılımları farklılaştıkça daha yanlı sonuçlar elde edildiğine dair araştırma bulgusuyla paralellik göstermektedir (Livingston, 2004). Bununla birlikte Hambleton, Jones ve Rogers (1993) yetenek dağılımının madde parametrelerinin hatalı kestirilmesine yol açabileceğini ifade etmişlerdir.

Hem iki hem de üç parametrelili modeldeki madde parametreleri incelendiğinde ise a parametresi b parametresinden daha kararlı kestirilmeye yol açmıştır. Araştırmadan elde edilen bu sonuç Ironson (1983) ve Li, Tam ve Tompkins (2004)'in yaptıkları araştırmada elde ettikleri sonuçlarla tutarlılık gösterirken, Lord (1983) ile tutarlılık göstermemektedir. Lord (1983) tarafından yapılan çalışmada a parametresinin b parametresinden daha zayıf kestirildiği görülmüştür. Bu konuya ilişkin olarak alanyazında zayıf bir şekilde kestirilen madde güçlüklerinin MTK'ye dayalı eşitleme sonuçları üzerinde ciddi bir etkiye sahip olduğu ifade edilmektedir (Stocking & Lord, 1983).

Genel olarak iki ve üç parametrelili modelden elde edilen sonuçlar karşılaştırıldığında, hem a hem de b parametresi için iki parametrelili modelin üç parametrelili modelden daha

az hatalı ve yanlış sonuçlar ürettiği bulunmuştur. Araştırmada elde edilen bu sonuç, Kaskowitz ve De Ayala (2001)'nin yaptıkları çalışmada elde ettikleri sonuç ile paralellik gösterirken, Domaleski (2006) ile paralellik göstermemektedir. Kaskowitz ve De Ayala (2001) yaptıkları çalışmada a ve b parametreleri için 3PLM'nin 2PLM'den daha yüksek hataya sahip sonuçlar verdiğini bulurken, Domaleski (2006) madde parametreleri açısından model türüne göre yöntemler arasındaki farkın oldukça küçük olduğunu bulmuştur.

3.4. ALT PROBLEM 4'E İLİŞKİN BULGULAR VE YORUMLAR

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak üç parametrelili lojistik modele uyumlu veriler ile kestirilen yetenek parametrelerinin yanlışlık ve eşitleme hatası;

- a. Test uzunluğuna*
- b. Örneklem büyüklüğüne*
- c. Yetenek dağılımına göre nasıl değişmektedir?*
- d. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğu, örneklem büyüklüğü, yetenek dağılımına göre en iyi sonucu vermektedir?*

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak üç parametrelili lojistik modele uyumlu veriler ile kestirilen yetenek parametrelerinin yanlışlık ve eşitleme hatası Tablo 3.3'te verilmiştir:

- a. Test uzunluğuna göre yöntemlerin hata ve yanlışlıkları nasıl değişmektedir?*

Test uzunluğu bakımından yöntemler incelendiğinde, standart normal dağılım ve çarpık yetenek dağılımlarında test uzunluğunun yöntemler üzerinde olumlu bir etkiye sahip olduğu ve test uzunluğu arttıkça hata ve yanlışlık değerlerinin düştüğü gözlenmektedir. Araştırmadan elde edilen bu sonuç Kang ve Petersen (2009) tarafından yapılan ve test uzunluğu arttıkça (10'dan 40'a) yöntemlerin hata değerlerinin düştüğüne dair elde edilen araştırma sonucuyla benzerlik göstermektedir.

Tablo 3.3. 3PLM ile kestirilen yetenek parametrelerinin simülasyon koşullarına göre yanlılık (BIAS) ve eşitleme hataları (RMSE)

Yetenek dağılımı	Örneklem büyüklüğü	Test uzunluğu	Değerlendirme ölçütü	3PLM		
				Yetenek parametresi (θ)		
				OO	OS	SL
G1:N(0,1) G2:N(0.5,1)	1000	30	BIAS	0.498	0.463	0.478
			RMSE	0.518	0.481	0.488
		60	BIAS	0.410	0.414	0.402
			RMSE	0.457	0.447	0.421
		80	BIAS	0.369	0.408	0.355
			RMSE	0.409	0.423	0.408
	3000	30	BIAS	0.485	0.489	0.473
			RMSE	0.493	0.500	0.475
		60	BIAS	0.472	0.482	0.469
			RMSE	0.478	0.492	0.471
		80	BIAS	0.459	0.457	0.457
			RMSE	0.463	0.468	0.458
G1:PÇ G2:N(0,1)	1000	30	BIAS	1.577	1.526	1.710
			RMSE	1.597	1.559	1.720
		60	BIAS	1.434	1.465	1.388
			RMSE	1.460	1.495	1.427
		80	BIAS	1.354	1.345	1.333
			RMSE	1.425	1.420	1.424
	3000	30	BIAS	1.772	1.761	1.780
			RMSE	1.778	1.773	1.786
		60	BIAS	1.600	1.576	1.486
			RMSE	1.627	1.653	1.507
		80	BIAS	1.431	1.446	1.337
			RMSE	1.447	1.518	1.350
G1:NÇ G2:N(0,1)	1000	30	BIAS	-1.950	-1.972	-1.921
			RMSE	1.964	1.999	1.926
		60	BIAS	-1.456	-1.427	-1.582
			RMSE	1.482	1.453	1.611
		80	BIAS	-1.135	-1.208	-1.281
			RMSE	1.159	1.228	1.353
	3000	30	BIAS	1.604	1.567	1.648
			RMSE	1.656	1.586	1.619
		60	BIAS	1.496	1.530	1.391
			RMSE	1.526	1.578	1.417
		80	BIAS	1.336	1.332	1.190
			RMSE	1.363	1.388	1.211

OO: ortalama-ortalama, OS: ortalama-standart sapma, SL: Stocking-Lord

b. Örneklem büyüklüğüne göre yöntemlerin hata ve yanlılıkları nasıl değişmektedir?

Örneklem büyüklüğü açısından yöntemlerin performansı incelendiğinde ise, yetenek parametrelerine göre, negatif çarpık dağılımlar dışında örneklem büyüklüğünün yöntemler üzerinde herhangi bir etkisinin olmadığı görülmektedir. Araştırmadan elde edilen bu sonuç Kang ve Petersen (2009) tarafından yapılan araştırma sonucu ile tam olarak benzerlik göstermemektedir. Kang ve Petersen (2009) 3PLM'ye uyumlu veriler üzerinde yaptıkları yöntemlerin test ve madde karakteristik eğrilerini karşılaştırmayı

amaçlayan arařtırmalarında, tüm kořullarda örneklem büyüklüğü arttıkça (500'den 2000'e) yöntemlerin hata ve yanlılık deęerlerinin düřtüđünü bulmuřlardır. Bu arařtırmada ise sadece negatif çarpık daęılımlarda örneklem büyüklüğü arttıkça yöntemleri doęruluęu artmıř ve daha kararlı kestirimler elde edilmiřtir. Buradan hareketle, benzer kořullar altında bile eřitleme yöntemlerinin farklı sonuçlar verdięi sonucuna ulařılabilir. Bu durum, özellikle toplam hata dikkate alındıęında (RMSE), toplam hatayı oluřturan bileřenlerin farklı katkılarından dolayı sonuçların farklılařmasından kaynaklanabilir.

c. Yetenek daęılımına göre yöntemlerin hata ve yanlılıkları nasıl deęiřmektedir?

Gruplar arasındaki farklar arttıkça yöntemlerin hata ve yanlılık miktarlarının arttıęı da görölmektedir. Arařtırmada elde edilen bu sonuç, Kang ve Petersen (2009) tarafından yapılan arařtırmada gruplar arasındaki farkların arttıkça yöntemlerin daha hatalı kestirimlere yol açtıęına dair elde edilen sonuçla benzerlik göstermektedir. Aynı řekilde Keller ve Keller (2008) tarafından 3PLM'ye uygun yetenek parametreleri üzerinde yapılan çalıřmada negatif çarpık daęılımlarda en büyük hata ve yanlılık deęerlerinin elde edildięi görölmüřtür. Pozitif ve negatif çarpık yetenek daęılımları birbirleriyle karřılařtırıldıęında ise üç parametrelili modellerde 1000 kiřilik örneklemelerde pozitif çarpık, 3000 kiřilik örneklemelerde ise negatif çarpık yetenek daęılımlarının daha az hatalı sonuçlar verdięi görölmektedir.

d. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluęuna, örneklem büyüklüęüne ve yetenek daęılımına göre en iyi sonucu vermektedir?

Yetenek parametresi açısından yöntemler birbirleriyle karřılařtırıldıęında, hata ve yanlılık deęerlerinin tüm yöntemler için oldukça benzer olduęu görölmüřtür. Arařtırmada elde edilen bu bulgu, Keller ve diđerleri (2007), Keller ve Keller (2008) ile Baldwin, Nering ve Baldwin (2007) tarafından yapılan arařtırma ile benzerlik göstermektedir. Keller ve diđerleri (2007) yaptıkları çalıřmada yöntemler arasında küçük farklar olduęunu ve genel olarak tüm yöntemlerin mantıklı sonuçlar verdięini bulmuřlardır. Yöntemler tüm kořullar için genel olarak incelendięinde, pek çok durumda en iyi yöntemin Stocking-Lord olduęu görölmürken, bazı durumlarda ise

ortalama-standart sapma yöntemi daha az hatalı kestirimlerde bulunmuştur. Bununla birlikte en hatalı sonuçları ise ortalama-ortalama yönteminin verdiği görülmüştür. Araştırmadan elde edilen bu sonuç, Turhan ve diğerleri (2007) ile Baldwin ve diğerleri (2007) tarafından 3PLM üzerinde yapılan çalışmada yetenek parametreleri açısından ortalama-standart sapma yönteminin en az hatalı kestirime yol açtığına dair elde edilen sonuçla kısmen de olsa benzerlik göstermektedir. Keller III (2007) tarafından 3PLM kullanılarak yetenek parametresi üzerinde yapılan araştırmada çarpık dağılıma sahip gruplarda ve sadece ortalamanın değiştiği yetenek dağılımlarında tüm yöntemler için yanlılık değerlerinin benzer olduğu bulunmuştur. Bununla birlikte tüm yetenek dağılımlarında en iyi yöntemin SL yöntemi olduğu görülürken, en fazla hataya ve yanlılığa sahip yöntemin ise OO ve OS yöntemi olduğu görülmektedir. Aynı şekilde, Keller ve Keller (2008) yaptıkları çalışmada moment yöntemlerinin (OO ve OS) çarpıklıktan daha fazla etkilendiğini ve SL yönteminin en az hata ve yanlılığa sahip yöntem olduğunu bulmuşlardır. Bu durumun nedeninin, SL yöntemi ölçekleme sabitlerini belirlemede madde parametrelerinin tümünü kullanırken, OO ve OS yöntemlerinin kullanmaması olabileceği ifade edilebilir.

3.5. ALT PROBLEM 5'E İLİŞKİN BULGULAR VE YORUMLAR

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak iki parametrelili lojistik modele uyumlu veriler ile kestirilen yetenek parametrelerinin yanlılık ve eşitleme hatası;

- a. Test uzunluğuna,*
- b. Örneklem büyüklüğüne*
- c. Yetenek dağılımına göre nasıl değişmektedir?*
- d. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğu, örneklem büyüklüğü, yetenek dağılımına göre en iyi sonucu vermektedir?*

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak iki parametrelili lojistik modele uyumlu veriler ile kestirilen yetenek parametrelerinin yanlılık ve eşitleme hatası Tablo 3.4'te verilmiştir.

Tablo 3.4. 2PLM ile kestirilen yetenek parametrelerinin simülasyon koşullarına göre yanlılık (BIAS) ve eşitleme hataları (RMSE)

Yetenek dağılımı	Örneklem büyüklüğü	Test uzunluğu	Değerlendirme ölçütü	2PLM		
				Yetenek parametreleri (θ)		
				OO	OS	SL
G1:N(0,1) G2:N(0.5,1)	1000	30	BIAS	0.446	0.447	0.439
			RMSE	0.455	0.468	0.457
		60	BIAS	0.438	0.430	0.438
			RMSE	0.441	0.448	0.439
		80	BIAS	0.399	0.402	0.400
			RMSE	0.401	0.408	0.399
	3000	30	BIAS	0.517	0.507	0.505
			RMSE	0.518	0.508	0.506
		60	BIAS	0.512	0.489	0.497
			RMSE	0.513	0.490	0.498
		80	BIAS	0.493	0.479	0.473
			RMSE	0.495	0.489	0.479
G1:PÇ G2:N(0,1)	1000	30	BIAS	1.622	1.636	1.629
			RMSE	1.679	1.700	1.685
		60	BIAS	1.355	1.334	1.306
			RMSE	1.436	1.446	1.523
		80	BIAS	0.920	0.954	0.893
			RMSE	1.060	1.082	1.031
	3000	30	BIAS	-1.454	-1.471	-1.524
			RMSE	1.547	1.559	1.619
		60	BIAS	-1.376	-1.361	-1.305
			RMSE	1.544	1.546	1.472
		80	BIAS	-0.994	-0.994	-0.928
			RMSE	1.318	1.315	1.257
G1:NÇ G2:N(0,1)	1000	30	BIAS	-1.314	-1.303	-1.276
			RMSE	1.517	1.531	1.543
		60	BIAS	-0.789	-0.787	-0.716
			RMSE	1.109	1.152	1.064
		80	BIAS	-0.703	-0.691	-0.622
			RMSE	1.076	1.088	1.028
	3000	30	BIAS	-1.906	-1.940	-1.808
			RMSE	1.908	1.943	1.809
		60	BIAS	-1.514	-1.545	-1.419
			RMSE	1.721	1.786	1.699
		80	BIAS	-1.397	-1.352	-1.419
			RMSE	1.450	1.446	1.490

OO: ortalama-ortalama, OS: ortalama-standart sapma, SL: Stocking-Lord

a. Test uzunluđuna gore yontemlerin hata ve yanlılıkları nasıl deđiřmektedir?

Tablo 3.5 incelendiđinde 2PLM'ye uyumlu veriler uzerinde test uzunluđu aısından hem standart normal dađılım hem de arpık yetenek dađılımlarında test uzunluđu arttııa yontemlerin yanlılık ve hata deđerlerinde bir duřuř olmaktadır. Bu nedenle test uzunluđunun yetenek parametrelerine etkisinin olumlu yonde olduđu soylenebilir.

b. Orneklem buyukluđune gore yontemlerin hata ve yanlılıkları nasıl deđiřmektedir?

Orneklem buyukluđu aısından yontemler incelendiđinde ise genel olarak orneklem buyukluđunun birka istisna dıřında yetenek parametrelerinin hata ve yanlılık deđerleri uzerinde herhangi bir etkisinin olmadıđı gorulmektedir. Orneklem buyukluđu arttııa yontemlerin hata ve yanlılık deđerlerinde bir azalma olmamaktadır. Ozellikle sistematik hata aısından duřunulduđunde (yanlılık), sistematik hatanın orneklem buyukluđunden dođrudan etkilenmediđi ve buyuk orneklemelerin sistematik hatayı azaltmadıđı ifade edilmektedir (Kolen & Brennan, 2004). Buna gore 2PLM ile yapılan kestirimde orneklem buyukluđunun herhangi bir etkisinin olmadıđı soylenebilir. Orneklem buyukluđu aısından 2PLM'den elde edilen sonular 3PLM'den elde edilen sonular ile tutarlılık gostermektedir. Suanthong (1998) tarafından 1 PLM'ye uyumlu veriler uzerinde yapılan alıřmada orneklem buyukluđu arttııa hata ve yanlılıđın azalmadıđı 300 ve 500 kiřilik orneklemelere gore 100 kiřilik orneklemlerde daha az hatalı sonular elde edilmiřtir. Aynı řekilde bu arařtırmada da hem 3 PLM hem de 2 PLM'ye uyumlu verilerde 1000 kiřilik orneklemlerde daha az hatalı ve yanlı sonular elde edilmiřtir. Bu dođrultuda yetenek parametrelerinin kullanıldıđı eřitleme alıřmalarında 1000 kiřilik orneklemelerin yeterli olduđu soylenebilir.

c. Yetenek dađılımına gore yontemlerin hata ve yanlılıkları nasıl deđiřmektedir?

Yetenek dađılımları aısından, benzer yetenek dađılımına sahip gruplarda, farklı ve arpık yetenek dađılımlarına nazaran daha kuuk hata ve yanlılık deđerleri elde edildiđi gorulmektedir. Buna gore, gruplar arasındaki farklar arttııa hata ve yanlılık deđerlerinin de arttıđı soylenebilir. Bununla birlikte, gruplar kendi ierisinde karřılařtırıldıđında, pozitif arpık dađılımlarda negatif arpık dađılımlara gore daha buyuk hata ve yanlılık deđerleri elde edildiđi gorulmektedir.

d. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğuna, örneklem büyüklüğüne, yetenek dağılımına göre en iyi sonucu vermektedir?

Yöntemler test uzunluğu açısından karşılaştırıldığında, 30 maddelik testlerde SL ve OO yöntemleri en düşük hata ve yanlılığa sahipken, 60 ve 80 maddelik testlerde ise SL yönteminin en düşük hata ve yanlılık değerine sahip yöntem olduğu görülmektedir. Aynı şekilde, yöntemler örneklem büyüklüğü ve yetenek dağılımı açısından karşılaştırıldığında, standart normal dağılımda 1000 kişilik örneklemelerde en iyi yöntemin OO yöntemi olduğu görülürken, 3000 kişilik örneklemelerde ise SL olduğu görülmektedir. Pozitif çarpık dağılımda en iyi yöntemler, 1000 kişilik örneklemelerde OO ve SL iken, 3000 kişilik örneklemelerde ise SL yöntemidir. Negatif çarpık dağılımlarda, 1000 ve 3000 kişilik örneklemelerde SL en iyi yöntemdir. Sonuç olarak, çarpık yetenek dağılımlarında SL yönteminin daha başarılı olduğu; bir başka deyişle OO ve OS yöntemlerinin çarpıklıktan daha çok etkilendiği söylenebilir.

Koşullar göz önüne alınmaksızın yöntemler genel olarak karşılaştırıldığında ise en kararlı yöntemin SL, en hatalı kestirime yol açan yöntemin ise OO ve pek çok durumda OS olduğu görülmektedir. Bununla birlikte, genel olarak yöntemlere ilişkin RMSE ve BIAS değerlerinin de benzer olduğu görülmektedir. Bu alanda yapılan çalışmalar, yetenek parametreleri açısından OO, OS, SL ve Haebara yöntemleri arasındaki farkların oldukça az olduğunu göstermiştir (Keller ve diğerleri, 2007; Sukin & Keller, 2008; Kang & Petersen, 2009). Yetenek parametreleri üzerinde yapılan çalışmaların çok az bir kısmı belirli koşullar altında MTK eşitleme yöntemleri arasında önemli farklılıklar olduğunu göstermiştir.

Gruplar yetenek açısından farklı ya da benzer olduğunda, hangi eşitleme yönteminin kullanılması gerektiğine karar verilirken yanlılığın yanında random ve sistematik hatanın bileşimi olan eşitleme hatası da göz önüne alınmalıdır. Örneğin, bir eşitleme yöntemi başka bir yöntemle göre daha yanlıysa fakat daha küçük random eşitleme hatasına neden oluyorsa, daha küçük random eşitleme hatasına sahip yöntemi kullanmak daha istenilen bir durum olabilir. Yöntemlerin, grupların yetenekleri farklı olduğunda daha büyük eşitleme yanlılığı üretirken, daha az hataya neden olduğu da

görülmüştür. Bunun nedeni, ortak testin yanlılığı ortadan kaldırmada başarısız olması olabilir. Özellikle pozitif ve negatif çarpık dağılımlarda bazı durumlarda bir yöntemin daha büyük BIAS değerine sahipken aynı zamanda da daha az hataya yol açtığı görülmektedir. Araştırmada elde edilen bu bulgu, Holland, Sinharay, von Davier ve Han (2007) ile Wang, Lee, Brennan ve Kolen (2008) tarafından yapılan eşitleme yöntemlerinin karşılaştırıldığı çalışmada, yöntemlerin daha büyük yanlılık ve daha küçük standart hataya yol açtığına dair elde edilen sonuç ile tutarlılık göstermektedir.

3.6. ALT PROBLEM 6'YA İLİŞKİN BULGULAR VE YORUMLAR

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak elde edilen yetenek parametrelerinin yanlılık ve eşitleme hatası model türüne göre nasıl değişmektedir?

Yöntemler model türü açısından karşılaştırıldığında, iki parametrelili modelin üç parametrelili modelden daha düşük hata ve yanlılığa sahip olduğu görülmüştür. Araştırmadan elde edilen bu sonuç, Domaleski (2006) tarafından yapılan araştırmada elde edilen sonuç ile tutarlılık göstermektedir. Domaleski (2006) yaptığı araştırmada yetenek parametresi açısından 3PLM'nin 2PLM'den daha fazla hataya sahip olduğunu bulmuştur.

3.7. ALT PROBLEM 7'YE İLİŞKİN BULGULAR VE YORUMLAR

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerine göre üç parametrelili lojistik modele uyumlu veriler kullanılarak gerçek puan eşitlemesine dayalı yapılan eşitlemelerin yanlılık ve eşitleme hatası;

- a. Test uzunluğuna
- b. Örneklem büyüklüğüne
- c. Yetenek dağılımına göre nasıl değişmektedir?
- d. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğuna, örneklem büyüklüğüne, yetenek dağılımına göre en iyi sonucu vermektedir?

Araştırmada ele alınan koşullar doğrultusunda, yöntemler “madde ve yetenek parametreleri” açısından karşılaştırıldıktan sonra “puan” düzeyinde de karşılaştırmalar yapılmış ve sonuçlar Tablo 3.5’te verilmiştir. Puan düzeyinde karşılaştırmalar, gerçek puan eşitlemeye dayalı olarak yapılmıştır. Gerçek puan eşitlemede belirli bir θ ile ilişkili olarak bir formdan elde edilen gerçek ham puanların diğer bir formdaki gerçek ham puanlara denk olduğu düşünülür. Araştırmada her bir eşitleme yöntemi için ham puan ile eşitlenmiş puanlar arasındaki farklar RMSE ve BIAS değerleri ile değerlendirilmiştir.

Tablo 3.5. 3PLM ile yapılan gerçek puan eşitlemenin simülasyon koşullarına göre yanlışlık (BIAS) ve eşitleme hataları (RMSE)

Yetenek dağılımı	Örneklem büyüklüğü	Test uzunluğu	Değerlendirme ölçütü	3PLM		
				Gerçek puan eşitlemesi		
				OO	OS	SL
G1:N(0,1) G2:N(0.5,1)	1000	30	BIAS	-1.296	-1.299	-1.611
			RMSE	1.043	1.053	1.272
		60	BIAS	-2.135	-2.145	-2.324
			RMSE	2.946	2.993	3.486
		80	BIAS	-2.347	-2.338	-2.821
			RMSE	3.859	3.727	4.680
	3000	30	BIAS	-1.493	-1.489	-1.695
			RMSE	1.173	1.180	1.320
		60	BIAS	-2.579	-2.574	-3.013
			RMSE	3.022	3.045	3.497
		80	BIAS	-2.451	-2.489	-2.759
			RMSE	4.565	4.445	6.225
G1:PÇ G2:N(0,1)	1000	30	BIAS	-3.132	-3.116	-3.647
			RMSE	2.466	2.460	2.845
		60	BIAS	-6.900	-7.016	-7.780
			RMSE	8.646	8.593	9.518
		80	BIAS	-6.595	-6.507	-7.759
			RMSE	9.146	9.010	10.670
	3000	30	BIAS	-4.694	-4.734	-5.399
			RMSE	3.644	3.680	4.188
		60	BIAS	-7.219	-7.125	-8.117
			RMSE	8.616	8.506	9.460
		80	BIAS	-7.192	-7.112	-8.264
			RMSE	10.526	10.659	12.113
G1:NÇ G2:N(0,1)	1000	30	BIAS	3.878	3.781	4.595
			RMSE	3.113	3.143	3.738
		60	BIAS	4.002	4.436	4.972
			RMSE	4.627	4.873	5.615
		80	BIAS	4.401	5.565	6.143
			RMSE	6.138	7.697	8.607
	3000	30	BIAS	-3.895	-3.818	-4.579
			RMSE	3.033	2.984	3.559
		60	BIAS	-7.111	-7.084	-7.704
			RMSE	8.716	8.668	9.218
		80	BIAS	-7.169	-7.026	-7.701
			RMSE	11.387	11.014	11.717

OO: ortalama-ortalama, OS: ortalama-standart sapma, SL: Stocking-Lord

a. Test uzunluđuna gre yntemlerin hata ve yanlılıkları nasıl deđiřmektedir?

Yntemlerin hata ve yanlılık deđerleri test uzunluđu aısından incelendiđinde, test uzunluđunun yntemler zerinde herhangi bir etkisinin olmadığı; pek ok durumda test uzunluđu arttıa hata ve yanlılık deđerlerinin de arttıđı grlmektedir. 30 maddelik testlerde 60 ve 80 maddelik testlere gre hata ve yanlılık deđerlerinin daha dřk olduđu grlmektedir. Arařtırmada elde edilen bu sonu, Whittington (1986) tarafından yapılan arařtırmada elde edilen sonula tutarlılık gstermektedir. Lee ve Ban (2010) tarafından yapılan alıřmada da test uzunluđunun eřitleme yntemleri zerinde etkisinin olmadığı grlmř ve daha kısa testlerde (25 maddelik) daha dřk hata ve yanlılık deđerleri elde edilmiřtir.

b. rneklem byklđne gre yntemlerin hata ve yanlılıkları nasıl deđiřmektedir?

Yntemlerin hata ve yanlılık deđerleri rneklem byklđ aısından incelendiđinde, rneklem byklđnn yntemler zerinde herhangi bir etkisinin olmadığı; rneklem byklđ arttıa hata ve yanlılık deđerlerinin arttıđı grlmektedir. 1000 kiřilik rneklemlerde 3000 kiřilik rneklemelere gre hata ve yanlılık deđerlerinin daha dřk olduđu grlmektedir. Arařtırmada elde edilen bu sonu, Whittington (1986) tarafından yapılan arařtırmada 1000 kiřilik rneklem ve 30 maddelik testlerde daha dřk hata ve yanlılık deđerlerinin elde edilmesine dair bulunan sonula tutarlılık gstermektedir.

c. Yetenek dađılımına gre yntemlerin hata ve yanlılıkları nasıl deđiřmektedir?

Yetenek dađılımı aısından yntemlerin hata ve yanlılık deđerleri incelendiđinde ise, gruplar birbirine benzer olduđunda yntemlerin benzer sonular verdiđi grlrken, gruplar farklı olduđunda ise yntemler arasındaki farkların da arttıđı grlmektedir. Arařtırmada elde edilen bu sonu Kim, Brennan ve Kolen (2005) ile Tong ve Kolen (2005) tarafından yapılan alıřmalarda elde edilen sonularla tutarlılık gstermektedir. Gruplar arasındaki farklar arttıa hata ve yanlılık deđerlerinin de arttıđı, benzer gruplarda hata ve yanlılık deđerlerinin daha dřk olduđu grlmektedir. Bu dođrultuda, arařtırmada grup farkının sonular zerinde olduka nemli bir etkiye sahip olduđu bulunmuřtur. Arařtırmanın bu bulguları, Cao (2008) tarafından yapılan ve eřitleme yntemlerinin puan dzeyinde karřılařtırıldıđı alıřmada elde edilen bulgularla benzerlik gstermektedir. Aynı řekilde Puhan (2010) tarafından farklı eřitleme

yöntemlerinin karşılaştırıldığı çalışmada da farklı yetenek dağılıma sahip gruplarda hata ve yanlılık değerlerinin benzer yetenek dağılımına sahip gruplara göre daha büyük olduğu bulunmuştur.

d. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğuna, örneklem büyüklüğüne, yetenek dağılımına göre en iyi sonucu vermektedir?

Tablo 3.6’da pozitif fark eşitlenmiş puanın buna denk olan ham puandan daha büyük olduğunu gösterirken, negatif fark ise tam tersi bir durumu ifade etmektedir. Eşitleme sonucunda yöntemlerden elde edilen ham puan ve eşitlenmiş puanların RMSE ve BIAS değerlerinin benzer olması beklenir. Bu kapsamda araştırmadan elde edilen sonuçlar incelendiğinde de, tüm koşullarda özellikle OO ve OS yöntemlerinin genel olarak RMSE ve BIAS değerlerinin benzer olduğu görülmektedir. Bu sonuç, Whittington (1986) ve Wells (2004) tarafından yapılan çalışmada ham puanlar ve eşitlenmiş puanlar arasındaki farklara dayalı olarak hesaplanan hata ve yanlılık değerlerinin genel olarak benzer değerlere sahip olduğuna dair elde edilen araştırma sonucuyla benzerlik göstermektedir. Benzer şekilde Hagge (2010) tarafından yapılan çalışmada da birbirine benzer yetenek dağılımına sahip gruplarda tüm eşitleme yöntemlerinin benzer eşitlenmiş puanlar verdiği görülmüştür. Bu nedenle, gerçek test etme uygulamaları açısından düşünüldüğünde, 3PLM için eşitleme yönteminin seçiminin oldukça az bir öneme sahip olduğu söylenebilir.

Araştırmada kapsamında ele alınan yöntemler karşılaştırıldığında, genel olarak en az hataya sahip yöntemin OS yöntemi olduğu görülmektedir. Kolen ve Brennan (2004) OS yönteminin OO yönteminden bazen daha tercih edilebilir olduğunu çünkü b parametresi a parametresinden daha kararlı kestirime neden olduğunu ifade etmişlerdir. Bu alanda yapılan araştırmaların çok az bir kısmı belirli koşullar altında MTK eşitleme yöntemleri arasında önemli farklılıklar olduğunu göstermiştir. Keller ve diğerleri (2007) yaptıkları çalışmada yöntemler arasında küçük farklar olduğunu ve genel olarak tüm yöntemlerin mantıklı sonuçlar verdiğini bulmuşlardır. Bu açıdan araştırmadan elde edilen sonuç ile bu alanda yapılan araştırmaların sonuçları tutarlılık göstermektedir. Bununla birlikte tüm koşullarda en fazla hataya ve yanlılığa sahip yöntemin SL yöntemi olduğu

görülmektedir. Bu doğrultuda gerçek veri uygulamalarında 3PLM için tüm koşullarda OS yönteminin kullanılması tavsiye edilebilir.

Ayrı kestirim yöntemleri ortak maddeler için madde parametrelerinin karşılaştırılmasına imkân verir. Eş zamanlı kestirim yönteminde ise her bir ortak madde için tek bir parametre seti ortaya çıkar. Bu nedenle ayrı kestirim yöntemlerinde farklı yanlılık kaynakları yöntemleri etkilemektedir. Alanyazın incelendiğinde, ayrı kestirim yöntemlerinde iki test formunda yer alan maddelerin birebir eşitlenmesinden dolayı potansiyel yanlılık oluşabileceği ifade edilmektedir (Muraki ve diğerleri, 2000).

3.8. ALT PROBLEM 8'E İLİŞKİN BULGULAR VE YORUMLAR

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak iki parametrelili lojistik modele uyumlu veriler ile gerçek puan eşitlemesine dayalı yapılan eşitlemelerin yanlılık ve eşitleme hatası;

- a. Test uzunluğuna*
- b. Örneklem büyüklüğüne*
- c. Yetenek dağılımına göre nasıl değişmektedir?*
- d. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğu, örneklem büyüklüğü, yetenek dağılımına göre en iyi sonucu vermektedir?*

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak iki parametrelili lojistik modele uyumlu veriler ile gerçek puan eşitlemesine dayalı yapılan eşitlemelerin yanlılık ve eşitleme hatası Tablo 3.6'da verilmiştir:

a. Test uzunluğuna göre yöntemlerin hata ve yanlılıkları nasıl değişmektedir?

Araştırmadan elde edilen sonuçlar, test uzunluğu açısından incelendiğinde, test uzunluğunun yöntemler üzerinde olumlu düzeyde herhangi bir etkisinin olmadığı görülmekle birlikte, test uzunluğu arttıkça hata ve yanlılık değerlerinin de arttığı görülmektedir. Bununla birlikte kısa testlerde (30 maddelik) yöntemlerin hata ve yanlılıkları arasındaki farkların uzun testlere göre daha az olduğu görülmektedir.

Araştırmadan elde edilen bu sonuç Lee ve Ban (2010) tarafından yapılan araştırmada elde edilen sonuçla benzerlik göstermektedir.

Tablo 3.6. 2PLM ile gerçek puan eşitlemesinin simülasyon koşullarına göre yanlılık (BIAS) ve eşitleme hataları (RMSE)

Yetenek dağılımı	Örneklem büyüklüğü	Test uzunluğu	Değerlendirme ölçütü	2PLM		
				Gerçek puan ve eşitlenmiş puan		
				OO	OS	SL
G1:N(0,1) G2:N(0.5,1)	1000	30	BIAS	-2.016	-2.022	-2.372
			RMSE	1.602	1.629	1.882
		60	BIAS	-2.985	-2.984	-3.587
			RMSE	3.378	3.998	4.682
		80	BIAS	-2.694	-2.689	-3.320
			RMSE	4.139	4.186	4.099
	3000	30	BIAS	-2.497	-2.477	-2.998
			RMSE	1.938	1.922	2.324
		60	BIAS	-3.605	-3.565	-4.235
			RMSE	4.669	4.626	5.490
		80	BIAS	-3.269	-3.265	-4.041
			RMSE	4.814	4.826	5.949
G1:PÇ G2:N(0,1)	1000	30	BIAS	-5.838	-5.946	-7.062
			RMSE	4.655	4.768	5.634
		60	BIAS	-7.466	-7.670	-9.318
			RMSE	8.944	9.179	11.098
		80	BIAS	-6.601	-6.554	-7.604
			RMSE	11.397	11.353	13.234
	3000	30	BIAS	4.467	4.491	5.219
			RMSE	3.893	3.886	4.517
		60	BIAS	4.435	4.454	5.085
			RMSE	7.913	7.811	9.035
		80	BIAS	7.162	7.685	8.671
			RMSE	11.548	11.484	13.878
G1:NÇ G2:N(0,1)	1000	30	BIAS	4.130	4.014	4.742
			RMSE	3.802	3.872	4.462
		60	BIAS	3.018	3.056	3.309
			RMSE	7.464	7.702	8.388
		80	BIAS	3.955	3.933	4.481
			RMSE	10.106	10.194	11.501
	3000	30	BIAS	5.824	5.689	6.715
			RMSE	4.524	4.437	5.207
		60	BIAS	6.867	6.312	8.075
			RMSE	10.133	9.656	11.908
		80	BIAS	7.162	7.685	8.671
			RMSE	11.548	11.484	13.878

OO: ortalama-ortalama, OS:ortalama-standart sapma, SL:Stocking-Lord

b. Örneklem büyüklüğüne göre yöntemlerin hata ve yanlılıkları nasıl değişmektedir?

Örneklem büyüklüğü açısından incelendiğinde ise, test uzunluğu gibi örneklem büyüklüğünün de yöntemler üzerinde olumlu yönde herhangi bir etkisinin olmadığı görülmekle birlikte örneklem büyüklüğü arttıkça hata ve yanlılık değerlerinin de arttığı

görülmektedir. Bu doğrultuda, puan düzeyinde yapılan incelemelerde test uzunluğu ve örneklem büyüklüğünün 2PLM'ye uyumlu veriler üzerinde olumlu yönde herhangi bir etkisinin olmadığı söylenebilir.

c. Yetenek dağılımına göre yöntemlerin hata ve yanlılıkları nasıl değişmektedir?

Yetenek dağılımı açısından bakıldığında ise, gruplar arasındaki farklar arttıkça hata ve yanlılık değerlerinin de arttığı ve yöntemlerin daha zayıf performans gösterdiği görülmektedir. Bu açıdan gruplar arasındaki farkların fazla olduğu negatif ve pozitif yetenek dağılımlarında farkların az olduğu standart normal dağılımlara göre daha yanlı ve hatalı sonuçlar elde edilmiştir. Zeng (1991) tarafından yapılan araştırmada negatif çarpık dağılımlarda hata ve yanlılık değerlerinin standart normal dağılıma göre daha yüksek olduğuna dair elde edilen sonuç ile benzerlik göstermektedir. Buradan hareketle, 2PLM'ye uyumlu veriler üzerinde yetenek dağılımının oldukça önemli bir etkisinin olduğu söylenebilir.

Gruplar arasında büyük farklar olduğunda yöntemler daha yanlı sonuçlar verdiği için sonuçlar birbirinden farklı olma eğilimindedir ve bu yüzden de yöntemler arasındaki uyum düzeyi düşük olabilir. İki grubun yetenek dağılımı birbirine yakın olduğunda yöntemlerin benzer sonuçlar verdiği; hata ve yanlılık düzeylerinin de yüksek olduğu görülmektedir. Lee ve Ban (2010) tarafından yapılan çalışmada gruplar birbirinden farklı olduğunda yöntemler arasındaki uyum düzeyinin düşük olduğu bulunmuştur. Yöntemler için grup farklılığı yanlılığın temel kaynağı olarak görülmektedir.

d. Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemlerinden hangisi test uzunluğuna, örneklem büyüklüğüne, yetenek dağılımına göre en iyi sonucu vermektedir?

Araştırmada yetenek dağılımına göre yöntemler birbirleriyle karşılaştırıldığında, OS yöntemi ortalama grup farkına daha duyarlıyken, SL yöntemi ise ortalama grup farkına daha az duyarlıdır. Gruplar arasında farklılıklar olduğunda en az hatalı yöntem ortalama-ortalama olurken, en fazla hataya sahip yöntemin ise Stocking-Lord olduğu görülmektedir. Grupların yetenek dağılımları birbirine yakın olduğunda, yöntemlerin hata ve yanlılık değerleri daha düşük olurken ve yöntemler birbirine yakın sonuçlar

üretirken, gruplar arasındaki farklar arttıkça yöntemlerin hata ve yanlışlık değerleri de artmıştır. Grupların yetenek dağılımları benzer olduğunda OO ve OS yöntemlerinin benzer sonuçlar verdiği görülürken, farklı olduğunda ise birkaç istisna dışında OO yönteminin OS yönteminden daha az hataya sahip olduğu da görülmektedir. Araştırmadan elde edilen bu sonuç, Keller III (2007) tarafından puan düzeyinde yapılan karşılaştırmada elde edilen en iyi yöntemin OO ve en kötü yöntemin ise SL/Haebara olmasına dair araştırma sonucuyla benzerlik göstermektedir. Bunun nedeni ise, OO yönteminin OS yönteminden daha düşük varyansa sahip olmasından kaynaklanabilir. Baker ve Al-Karni (1991) ortalama-ortalama yönteminin bazen daha tercih edilebilir olduğunu çünkü ortalamaların standart sapmalardan daha kararlı olduğunu ifade etmiştir. Elde edilen bu sonuç, Hanson ve Beguin (1999) tarafından yapılan araştırma sonucuyla benzerdir. Bununla birlikte Baker ve Al-Karni (1991) tarafından yapılan araştırmada OO yönteminin OS yönteminden daha iyi sonuçlar verdiği ve bazı durumlarda OO yönteminin daha tercih edilebilir bir yöntem olduğu ifade edilmiştir.

2PLM’de tüm yöntemlerin RMSE ve BIAS değerlerinin benzer olduğu görülmektedir. Bununla birlikte tüm koşullarda en büyük hataya ve yanlışlığa sahip yöntemin SL olduğu görülürken, en az hataya ve yanlışlığa sahip yöntemin ise pek çok durumda OS yöntemi olduğu görülürken, bazı durumlarda da OO yöntemi olduğu görülmektedir. Bu doğrultuda 2PLM için gerçek test etme uygulamalarında OO ve OS yöntemlerinin kullanılması tercih edilebilir. Hu ve diğerleri (2008) tarafından yapılan araştırmada da ortalama-standart sapma yönteminin en küçük sistematik hataya sahip olduğu bulunmuştur.

3.9. ALT PROBLEM 9’A İLİŞKİN BULGULAR VE YORUMLAR

Ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri kullanılarak gerçek puan eşitlemesine dayalı olarak yapılan eşitlemelerin yanlışlık ve eşitleme hatası model türüne nasıl değişmektedir?

Araştırmadan elde edilen bulgular model türüne göre karşılaştırıldığında, puan düzeyinde yapılan karşılaştırmalarda üç parametreliliğin iki parametreliliğe göre daha düşük hata ve yanlışlık değerlerine sahip olduğu görülmektedir. Hem iki hem de üç

parametrelili lojistik modelde küçük örneklemlerde (1000 kişilik) ve kısa testlerde (30 maddelik) yöntemlerin daha az hatalı ve yanlış sonuçlar verdiği gözlenmektedir. Küçük örneklemler ve daha kısa testlere gereksinim duyulduğunda moment yöntemlerinin (OO ve OS) SL yönteminden daha iyi sonuçlar verdiği ifade edilebilir.

Araştırmadan elde edilen sonuçları genel olarak özetlemek gerekirse, madde ve yetenek parametreleri ve gerçek puan eşitlemesi için en iyi sonuç veren yöntemler Tablo 3.7’de verilmiştir.

Tablo 3.7. Madde ve yetenek parametreleri ile gerçek puan eşitlemesi için en az hatalı sonuç veren yöntemler

Yetenek dağılımı	Örneklem büyüklüğü	Test uzunluğu	Model türü	Madde parametreleri	Yetenek parametreleri	Gerçek puan eşitlemesi
G1:N(0,1) G2:N(0,5,1)	1000	30	3PLM	SL	OS	OO
			2PLM	OO	SL	OO
		60	3PLM	SL	SL	OO
			2PLM	SL	SL	OO
		80	3PLM	SL	SL	OS
			2PLM	a için OO, b için SL	SL	OO
	3000	30	3PLM	SL	SL	OO
			2PLM	SL	SL	OS
		60	3PLM	a için SL, b için OO	SL	OO
			2PLM	a için OS, b için SL	OS	OS
		80	3PLM	OO	SL	OS
			2PLM	a için OO, b için SL	SL	OO
G1:PÇ G2:N(0,1)	1000	30	3PLM	a için SL, b için OO	OS	OS
			2PLM	OS	OO	OO
		60	3PLM	a için OS, b için OO	SL	OO
			2PLM	OO	SL	OO
		80	3PLM	OS	SL	OS
			2PLM	a için SL, b için OS	SL	OS
	3000	30	3PLM	OS	OS	OO
			2PLM	a için SL, b için OO	OO	OO
		60	3PLM	OS	SL	OS
			2PLM	a için SL, b için OO	SL	OO
		80	3PLM	OO	SL	OO
			2PLM	a için SL, b için OO	SL	OO
G1:NÇ G2:N(0,1)	1000	30	3PLM	a için SL, b için OO	SL	OS
			2PLM	OO	OO	OO
		60	3PLM	a için OS, b için OO	OS	OO
			2PLM	a için SL, b için OO	SL	OO
		80	3PLM	a için SL, b için OO	OO	OO
			2PLM	a için SL, b için OO	SL	OO
	3000	30	3PLM	OO	OS	OS
			2PLM	a için SL, b için OO	SL	OS
		60	3PLM	OS	SL	OS
			2PLM	a için SL, b için OS	SL	OS
		80	3PLM	OO	SL	OS
			2PLM	a için OO, b için OS	OS	OO

OO: ortalama-ortalama, OS: ortalama-standart sapma, SL: Stocking-Lord

Tablo 3.7’de de görüldüğü üzere, madde parametreleri için pek çok koşulda en iyi yöntemin SL yöntemi olduğu görülmekle birlikte bazı koşullarda OO ve OS yöntemlerinin daha iyi sonuçlar verdiği görülmüştür. Yetenek parametreleri için en iyi yöntemin SL yöntemi olduğu görülürken, gerçek puan eşitlemesi durumunda ise birkaç istisna dışında her koşulda en iyi yöntemin OO yöntemi olduğu tespit edilmiştir.

Diğer koşullar göz önüne alınmaksızın sadece model türüne göre yöntemlerin en kararlı (1) sonuçlar verenden en hatalı (3) sonuçlar verene doğru sıralanışı Tablo 3.8’de verilmiştir.

Tablo 3.8. Model türüne göre yöntemlerin madde ve yetenek parametreleri ile gerçek puan eşitlemesi durumlarında en kararlıdan en hatalı sonuçlar verene doğru sıralanışı

Model Türü	Madde parametreleri			Yetenek parametreleri			Gerçek puan eşitleme		
	OO	OS	SL	OO	OS	SL	OO	OS	SL
3PLM	2	3	1	3	2	1	2	1	3
2PLM	1	3	2	2	2	1	2	1	3

OO: ortalama-ortalama, OS:ortalama-standart sapma, SL:Stocking-Lord

Tablo 3.8 incelendiğinde üç parametrelili modele uyumlu verilerde madde parametreleri kullanıldığında en az hataya sahip yöntemin SL, iki parametrelili modele uyumlu verilerde ise OO yöntemi olduğu görülürken, yetenek parametreleri için hem üç hem de iki parametrelili modele uyumlu verilerde SL yöntemi olduğu görülmektedir. Bununla birlikte gerçek puan eşitleme için ise hem iki hem de üç parametrelili modele uyumlu verilerde en az hatalı yöntemin OS olduğu belirlenmiştir.

BÖLÜM IV

SONUÇ VE ÖNERİLER

Bu bölümde araştırmadan elde edilen sonuçlara ve bu sonuçlara dayalı olarak sunulabilecek önerilere yer verilmiştir.

4.1. SONUÇLAR

Bu araştırmada madde tepki kuramına dayalı eşitleme yöntemlerinden OO, OS ve SL yöntemlerinin test uzunluğu, örneklem büyüklüğü, yetenek dağılımı ve model türüne göre karşılaştırılması amaçlanmıştır. Bu amaçla Türkiye'deki gerçek veri uygulamalarına benzer koşullarda veriler türetilerek ele alınan koşullarda hangi yöntemin en az hatalı sonuçlar verdiği incelenmiştir.

Araştırmada elde edilen sonuçlar, eşitleme yöntemlerinin performansı hakkında önemli ve yararlı bilgiler sağlar. Bu kapsamda, araştırma farklı eşitleme yöntemlerinden elde edilen sonuçlar doğrultusunda, gerçek test etme uygulamalarındaki koşullar göz önüne alınarak hangi yöntemi kullanmanın daha uygun olacağını göstergesi olabilir. Buna göre, araştırmada eşitlemenin doğruluğunu değerlendirmede kullanılan iki ölçüt olan RMSE ve BIAS arasında eşitlemenin doğruluğunu kestirmede küçük farklılıklar olduğu görülmekle birlikte, sonuçlar genel olarak incelendiğinde sonuçların birbiriyle uyumlu ve tutarlı olduğu da belirlenmiştir. Bununla birlikte hangi yöntemin hangi koşullarda en iyi sonuçlar verdiğini belirlemek için yöntemlerin sınırlılıklarını da göz önüne almak gerekir.

Araştırmadan elde edilen sonuçlar doğrultusunda;

1. Üç parametrelili lojistik modele uyumlu veriler için örneklem büyüklüğü ve test uzunluğu arttıkça daha kararlı madde parametre kestirimleri elde edilmiştir. Bu doğrultuda, 3PLM kullanıldığında en kararlı parametre kestirimlerine örneklem büyüklüğü 1000'in üzerinde olduğunda ve test uzunluğu da 60'in üzerinde olduğu durumlarda ulaşılabileceği söylenebilir. Benzer yetenek dağılımına sahip gruplarda ise

farklı yetenek dağılımına sahip gruplara göre yöntemlerin hata ve yanlışlık değerlerinin daha düşük olduğu bulunmuştur.

2. İki parametrelili lojistik modele uyumlu veriler üzerinden elde edilen madde parametreleri incelendiğinde ise, örneklem büyüklüğünün yanlışlık üzerinde herhangi bir etkiye sahip olmadığı görülürken, birkaç istisna dışında hata üzerinde olumlu bir etkiye sahip olduğu görülmüştür. Test uzunluğu 30'dan 60'a çıktıkça hata ve yanlışlık değerleri düşerken, 80 maddelik testlerde test uzunluğunun kestirim yöntemleri üzerinde herhangi bir etkisinin olmadığı görülmektedir. Test uzunluğu açısından ise 80 maddelik testlerde test uzunluğunun hata ve yanlışlık üzerinde herhangi bir etkisinin olmadığı görülmüştür. Yetenek dağılımı benzer gruplarda hata ve yanlışlık değerinin farklı gruplara göre daha düşük olduğu saptanmıştır.

3. Madde parametreleri açısından üç ve iki parametrelili modele göre yöntemler incelendiğinde, 3PLM'de en doğru kestirime yol açan yöntemin SL en hatalı kestirime yol açan yöntemin ise OS olduğu görülürken, 2PLM'de ise en doğru kestirime yol açan yöntemin OO olduğu, en hatalı kestirime yol açan yöntemin ise OS olduğu görülmüştür.

4. Model türüne göre yanlışlık ve hata değerleri incelendiğinde ise, 2PLM'nin 3PLM'den daha az yanlış ve hatalı sonuçlar verdiği görülmüştür.

5. Araştırmadan elde edilen sonuçlar yetenek parametresi açısından incelendiğinde, 3PLM için örneklem büyüklüğünün yöntemler üzerinde herhangi bir olumlu etkisinin olmadığı görülürken, test uzunluğu ve yetenek dağılımının yöntemler üzerinde olumlu bir etkiye sahip olduğu görülmüştür. Diğer bir deyişle, test uzunluğu arttıkça yöntemlerin hata ve yanlışlık değerleri düşerken, benzer yetenek dağılımına sahip gruplarda da daha düşük hata ve yanlışlık değerleri gözlenmiştir.

6. İki parametrelili modele uyumlu veriler üzerinde yetenek parametreleri incelendiğinde ise, yöntemlerin test uzunluğu ve yetenek dağılımında etkilendiği ve test uzunluğu arttıkça yöntemlerin hata ve yanlışlık değerlerinin azaldığı görülmüştür. Farklı yetenek dağılımlarına sahip gruplara nazaran benzer yetenek dağılımına sahip gruplarda daha

düşük hata ve yanlılık değerleri belirlenmiştir. Örneklem büyüklüğünün ise yöntemler üzerinde etkisinin olmadığı gözlenmiştir.

7. Yöntemler birbirleriyle karşılaştırıldığında, hem 3PLM hem de 2PLM için yöntemlerin hata ve yanlılık değerlerinin benzer olduğu görülmekle birlikte 3PLM için genel olarak tüm koşullarda SL ve OS yöntemlerinin OO yöntemine nazaran daha düşük hata ve yanlılığa sahip olduğu görülürken, 2PLM için ise en iyi yöntemin SL en kötü yöntemin ise OS olduğu saptanmıştır.

8. Yöntemler model türüne göre ele alındığında madde parametrelerinde olduğu gibi yetenek parametreleri için de 2PLM'nin 3PLM'den daha düşük hata ve yanlılığa sahip olduğu görülmüştür.

9. Araştırmada ele alınan koşullara göre yöntemlerin hata ve yanlılıkları puan düzeyinde incelendiğinde, 3PLM için yöntemlerin test uzunluğu ve örneklem büyüklüğünden etkilenmediği görülmüştür. Yetenek dağılımı açısından yöntemlerin hata ve yanlılık değerleri incelendiğinde ise, gruplar birbirine benzer olduğunda yöntemlerin benzer sonuçlar verdiği görülürken, gruplar farklı olduğunda ise yöntemler arasındaki farkların da arttığı saptanmıştır. Yöntemlerin hata ve yanlılık değerleri arasındaki farklar büyük olduğunda yöntemler arasında uyumun düşük düzeyde olduğu görülmüştür.

10. İki parametrelili modele uyumlu veriler üzerinde yöntemlerin hata ve yanlılıkları puan düzeyinde ele alındığında, test uzunluğu ve örneklem büyüklüğünden etkilenmediği görülmüştür. Yetenek dağılımı açısından bakıldığında ise, çarpık yetenek dağılımlarında (gruplar birbirinden farklı olduğunda) yöntemlerin hata ve yanlılık değerlerinin daha büyük olduğu görülürken, yetenek dağılımı benzer olduğu hata ve yanlılık değerlerinin daha küçük olduğu bulunmuştur.

11. Yöntemler birbirleriyle karşılaştırıldığında, 2PLM'de yöntemlerin birbirine benzer sonuçlar verdiği görülmekle birlikte, en kararlı kestirime sahip yöntemin OS yöntemi olduğu görülürken, en hatalı kestirime sahip yöntemin ise SL yöntemi olduğu bulunmuştur.

12. Puan düzeyinde yapılan karşılaştırmalarda iki ve üç parametrelî modelden elde edilen veriler birbirleriyle karşılaştırıldığında 3PLM'nin 2PLM'den daha az hatalı ve yanlı sonuçlar verdiği görülmüştür.

13. Model türü değişkenine göre yöntemler incelendiğinde, puan düzeyinde 3PLM 2PLM'den daha düşük hata ve yanlılık içermektedir.

Araştırmadan elde edilen sonuçlar genel olarak değerlendirildiğinde, en iyi eşitlemelerin 3000 kişilik örneklem, 80 maddelik testler, benzer yetenek dağılımına sahip gruplar, 2PLM ve SL yöntemi kullanılarak elde edilebileceği ifade edilebilir. Ayrıca araştırmada ele alınan koşullar doğrultusunda, büyük örneklem ile daha uzun testler kullanıldığında ve benzer yetenek dağılımına sahip gruplarda yöntemlerin daha az hatalı ve yanlı olduğu sonucuna da ulaşılmıştır.

Bu araştırmadan elde edilen sonuçların başka çalışmalara genellenmesi sadece bu çalışmada ele alınan koşullarla (örneklem büyüklüğü, test uzunluğu, yetenek dağılımı ve model türü) sınırlıdır. Bu kapsamda değerlendirildiğinde, araştırmada kullanılan koşulların birbirleriyle etkileşimi de araştırmının sonuçlarını etkileyen unsurlardan bir diğeridir. Araştırmada hem madde ve yetenek parametreleri hem de puan düzeyinde yapılan değerlendirmelerde, yetenek dağılımının eşitleme sonuçları üzerinde pek çok araştırmada olduğu gibi (Cao, 2008; Wu ve diğerleri, 2009; Ngudratok, 2009) en önemli etkiye sahip olduğu tespit edilmiştir. Bu doğrultuda, araştırmadan elde edilen sonuçlar belirli bir eşitleme yöntemini seçerken önemle üzerinde durulması gereken unsurun yetenek dağılımı olduğunu göstermiştir. Grupların denk olduğu ya da benzer olduğu durumlarda yöntemler denk ya da benzer olmayan gruplara göre daha iyi performans göstermiştir.

Bu çalışmanın sonucunda yöntemlerin farklı koşullarda farklı sonuçlar verdiği görülmüştür. Farklılığın miktarı ve hangi yöntemin tercih edileceği iki grubun yetenek dağılımının benzerliği, örneklem büyüklüğü, test uzunluğu ve model türüne bağlıdır. Bu doğrultuda, araştırmada ele alınan bu koşullar dikkate alınarak kullanılacak eşitleme yönteminin belirlenmesinin oldukça önemli olduğu sonucuna varılmıştır.

Yöntemlerdeki farklılıklar sonucunda, elde edilen sonuçlar bir yöntemi diğerine tercih etmek için yeterince kararlı değildir ve her koşulda etkili tek bir yöntem yoktur.

4.2. ÖNERİLER

Bu çalışmanın, ülkemizde test eşitleme çalışmaları yapacak kişilere uygulamanın yapıldığı örneklem büyüklüğü, test uzunluğu ve yetenek dağılımı için uygun olan eşitleme yönteminin kesin olarak bilinmemesi nedeniyle, söz konusu koşullar altında kullanılacak yöntemlerin belirlenmesinde yardımcı olacağı düşünülmektedir.

Ülkemizde uygulanan ALES, YGS, ÜDS, KPDS gibi büyük ölçekli sınavlar için test geliştirme uzmanları tarafından en doğru sonuçları verecek eşitleme yöntemini seçmek oldukça önemlidir. Öğrencileri belirli bir kuruma yerleştirmeyi amaçlayan bu sınavların sonuçları oldukça önemlidir. Güvenilir bir yolla öğrencilerin karşılaştırılmasını amaçlayan eşitleme yöntemini seçmek, bu tür sınavlardan elde edilen sonuçların geçerliğini etkileyecektir. Araştırmadan elde edilen sonuçlar, test geliştirme uzmanlarına test geçerliği için yararlı bilgiler sağlamakla birlikte, araştırmacılar ve test geliştirme uzmanlarına fikir verebilir.

Çalışmanın sonuçları farklı koşul ve durumlarda diğer yöntemlerden üstün olan tek bir yöntem olmadığı gibi, hangi yöntemi seçmenin en iyi sonuçları vereceği konusunda da açık bir kanıt yoktur; fakat eşitleme çalışmalarında elde edilen sonuçların daha önceki çalışmaların sonuçlarıyla tutarlılığı bir yöntemi seçmek için en bilgilendirici yoldur. Hanson ve Beguin (2002) pek çok eşitleme yöntemini birlikte kullanmanın ve sonuçları karşılaştırmanın uygulamada oldukça etkili olduğunu ifade etmişlerdir. Bu durum farklı koşullarda en uygun yöntemin seçilmesine yardımcı olacaktır.

Bu çalışma, tüm eşitleme sürecinin sadece bir parçasını içermektedir. Test etme programlarında ham-ölçek puanı dönüştürmeleri eşitlemenin son basamağı olarak yürütülür ve ölçek puanları beklenen toplam puanlar yerine rapor edilir. Bu yüzden test geliştirme uzmanları ölçek puanını etkileyen faktörler üzerinde çalışabilir. Bununla birlikte bu çalışmada formlar benzer güçlüktedir. Formların güçlükleri farklı olduğunda yöntemlerdeki farklılıklar incelenebilir.

Eşitleme konusunda yapılan arařtırmalar (Budesco, 1985; Wingsky ve diđerleri, 1987) ortak maddelerin uzunluęunun da yöntemlerin performansını etkiledięini ortaya koymuřtur. Bu kapsamda bu konuda yapılacak arařtırmalarda ortak madde uzunlukları da bir simülasyon kořulu olarak ele alınıp yöntemler üzerindeki etkisi incelenebilir. Bu çalışmada sadece sınırlı sayıda kořul ele alınmıř ve eşitleme sonuçlarını etkiledięi düşünölen pek çok faktör sabit tutulmuřtur. Örneęin, bu arařtırma 1000 ve 3000 kiřilik örneklemler üzerinde yapılmıřtır. Daha büyük örneklemler (örneęin gerçek veriye uygun olarak 10.000 kiři gibi) üzerinde de eşitleme çalışmaları yapılıp yöntemler bu örneklemler büyüklüklerine göre karşılaştırılabilir. Bununla birlikte arařtırma denk olmayan gruplar temel alınarak yürütölmüřtür. Denk olmayan gruplara ek olarak denk gruplar da simülasyon kořulu olarak alınıp iki farklı gruptan elde edilen sonuçlar karşılaştırılabilir. Sonuç olarak, arařtırmada ele alınan kořullar dıřında farklı kořullar ele alınıp yöntemlerin performansı bu kořullara göre karşılaştırılabilir.

Kolen (1981)'in de ifade ettięi gibi, farklı eşitleme yöntemlerinin doęruluęunu deęerlendirmek için tek ve kesin bir ölçüt yoktur. Bu yüzden eşitleme yöntemlerini karşılaştırırken pek çok ölçüt kullanmak gereklidir. Eşitleme konusunda yapılacak arařtırmalarda, eşitleme yöntemleri birbirleriyle karşılaştırılırken ve deęerlendirilirken RMSE ve BIAS (toplam hata ve sistematik hata) dıřında random eşitleme hatası ya da eşitlemenin standart hatası gibi farklı deęerlendirme indeksleri kullanılabilir. Bu indeksler kullanıldığında, bu çalışmada ele alınan yöntemlerin performansı deęiřebilir. Bu nedenle bundan sonra yapılacak arařtırmalarda random hata ya da standart hata kullanılarak yöntemlerdeki farklılıklar arařtırılabilir.

Arařtırmada MTK'ye dayalı eşitleme yöntemlerinden OO, OS ve SL yöntemleri karşılaştırılmıřtır. Farklı yöntemlerin farklı sonuçlar verdięini daha iyi anlayabilmek için diđer yöntemlerinde incelenmesi büyük önem taşımaktadır. Bu nedenle gelecekte bu konuda arařtırma yapacak kiřilere farklı eşitleme desenleri ve farklı eşitleme yöntemlerini kullanmaları önerilebilir.

Bu çalışmada eşitleme yöntemlerinin karşılaştırılması simülasyon verisi kullanılarak yapılmıřtır. Eđer çalışmada gerçek veri kullanılırsa yöntemlerin doęruluęunu belirlemek ve karşılařtırmak güçtür. Gerçek veri kullanıldığında sadece yöntemler arasında fark

olup olmadığını bilmek mümkündür. Ancak, simülasyon çalışması ile birlikte gerçek veri kullanılarak da benzer çalışmalar yapılabilir ve farklı türde iki veri setinden elde edilen sonuçlar karşılaştırılabilir.

KAYNAKÇA

- Albano, A. A. (2010). *Equate: Statistical methods for test equating [Computer software manual]*. Available from <http://CRAN.R-project.org/package=equate> (R package).
- Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.) *Educational measurement* (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.
- Ansley, N. A., & Forsyth, R. A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement*, 9, 37-48.
- Arai, S. & Mayekawa, S. (2011). A comparison of equating methods and linking designs for developing an item pool under item response theory. *Behaviormetrica*, 38(1), 1-16.
- Bakan Kalaycıoğlu, D. (2008). *Öğrenci Seçme Sınavı'nın madde yanlılığı açısından incelenmesi*. Doktora tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Barnard, J. J. (1996). *In search for equity in educational measurement: traditional versus modern equating methods*. Paper presented at ASEESA's national conference at the HSRC Conference Centre, Pretoria, South Africa.
- Baker, F. B. & Al-Karni, A. (1991). A comparison of two procedures for computing IRT equating coefficients. *Journal of Educational Measuremet*, 28(2), 147–162.
- Baldwin, S., Nering, M. & Baldwin, P. (April, 2007). *A comparison of IRT equating methods on recovering parameters and capturing growth in mixed-format tests*.

Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Bastari, B. (2000). *Linking multiple-choice and constructed-response items to a common proficiency scale*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Beguin, A. A. (2000). *Robustness of equating high-stakes tests*. Unpublished doctoral dissertation, University of Twente, The Netherlands.

Beguin, A. A. & Hanson, B. A. (2001). *Effect of noncompensatory multidimensionality on separate and concurrent estimation in IRT observed score equating*. Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle, WA.

Bekci, B. (2007). *Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı'nın değişen madde fonksiyonlarının cinsiyete ve okul türüne göre incelenmesi*. Yüksek Lisans tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.

Bozdağ, S. & Kan, A. (2010). Şans başarısının eşitlemeye etkisi. *H.Ü. Eğitim Fakültesi Dergisi*, 39, 91-108.

Cao, L. (2008). *Mixed format test equating: Effects of test dimensionality and common-item sets*. Unpublished doctorate thesis, University of Maryland.

Chu, K. & Kamata, A. (2000, April). *Nonequivalent group equating via I-P HGLLM*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Chu, K. L. & Kamata, A. (2003). *Test equating with the presence of DIF*. Paper presented at the annual meeting of American Educational Research Association, Chicago.

- Chulu, B. W. (2006). *Equating high stakes educational measurements: A study of design and consequences*. Unpublished doctorate thesis, University of Massachusetts, Amherst.
- Cohen, A. S. & Kim, S. H. (1998). An investigation of linking methods under the graded response model. *Applied Psychological Measurement*, 22(2), 116–130.
- 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.
- 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., & Petersen, N. S. (1987). Problems related to the use of conventional and item response theory equating methods in less than optimal circumstances. *Applied Psychological Measurement*, 11, 225-244.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart & Winston.
- Chui, Z. & Kolen, M. (2008). Comparison of parametric and nonparametric bootstrap methods for estimating random error in equipercentile equating. *Applied Psychological Measurement*, 32(4), 334-347.
- Çetin, E. (2009). *Dikey ölçeklemede klasik test ve madde tepki kuramına dayalı yöntemlerin karşılaştırılması*. Yayımlanmamış doktora tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- de Koning, E., Sijtsma, K. & Hamers, J. H. M. (2002). Comparison of four IRT models when analyzing two tests for inductive reasoning. *Applied Psychological Measurement*, 26, 302-320.

- Dismer- Davis, M. R. (1988). *Using the rasch model for vertical equating: A practical application study*. Unpublished doctorate thesis, Georgia State University, Atlanta, Georgia.
- Dorans, N. J. (1990). Equating methods and sampling designs. *Applied Measurement in Education*, 3(1), 3-17.
- 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.
- Douglas, J. (1997). Joint consistency nonparametric item characteristic curve and ability estimation. *Psychometrika*, 62(19), 7-28.
- Douglas, J. & Cohen, A. (2001). Nonparametric item response function estimation for assessing parametric model fit. *Applied Psychological Measurement*, 25, 234-243.
- Domaleski, C. S. (2006). *Exploring the efficacy of pre-equating a large scale criterion-referenced assessment with respect to measurement equivalence*. Unpublished doctorate thesis, Georgia State University, The College of Education, Atlanta, GA.
- Dongyang, L. (2009). *Developing a common scale for testlet model parameter estimates under the common-item nonequivalent groups design*. Unpublished doctorate thesis, University of Maryland.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum Associates Publishers.
- Felan, G. D. (2002). *Test equating: mean, linear, equapercentile and item response theory*. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin.

- Fitzpatrick, A. R. & Yen, W. M. (2001). The effects of test length and sample size on the reliability and equating of tests composed of constructed-response items. *Applied Measurement in Education, 14*(1), 31-57.
- Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research, 22*, 144-149.
- 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, The University of Iowa.
- Hambleton, R. K., & Murray, L. (1983). Some goodness of fit investigations for item response models. In R. K. Hambleton (Ed.) *Applications of item response theory* (pp. 74-94). Vancouver: Educational Research Institute of British Columbia.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hambleton, R. K. & Jones, R. W. (1993). An NCME instructional module on comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*, 38-47.
- Hambleton, R. K. & Jones, R. W. & Rogers, R. W. (1993). Influence of item parameter estimation errors in test development. *Journal of Educational Measurement, 30*(2), 143-155.
- Han, K. T. (2007). Wingen: windows software that generates item response theory parameters and item responses. *Applied Psychological Measurement, 31*, 457-459.

- Han, K. T. (2007). *WinGen2: Windows software that generates IRT parameters and item responses [computer program]*. Amherst, MA: Center for Educational Assessment, University of Massachusetts Amherst.
- Han, K. T. (2008). *Impact of item parameter drift on test equating and proficiency estimates*. Unpublished doctorate thesis, University of Massachusetts, Amherst.
- Han, T., Kolen, M. J., & Pohlmann, J. (1997). A comparison among IRT true- and observed score equating and traditional equipercentile equating. *Applied Measurement in Education, 10*, 105-121.
- Hanson, B. A. & Beguin, A. A. (1999). *Separate versus concurrent estimation of IRT item parameters in the common item equating design*. ACT research report series, Iowa City, IA. Eric document ED 438 310.
- 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.
- Harris, D. J. (1993). *Practical issues in equating*. Paper presented at the annual meeting of American Educational Research Association, Atlanta.
- Harris, D. J. (2003). Equating the multistate bar examination. *The Bar Examiner, 72*(3), 12-18.
- Harris, D. J. & Crouse, J. D. (1993). A study of criteria used in equating. *Applied Measurement in Education, 6*, 195-240.
- Harris, D. J. & Kolen, M. J. (1986). Effect of examinee group on equating relationships. *Applied Psychological Measurement, 10*, 35-43.
- Harwell, M., Stone, C. A., Hsu, T.-C. & Kirişci, L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement, 20*(2), 101-125.

- He, Y. (2011). *Evaluating equating properties for mixed-format tests*. Unpublished doctoral dissertation, University of Iowa, Iowa City.
- Hills, J. R., Subhiyah, R. G. & Hirsch, T. M. (1988). Equating minimum-competency tests: comparisons of methods. *Journal of Educational Measurement*, 25(3), 221-231.
- Holland, P. W. & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational measurement* (pp. 187–220). Westport, CT: Praeger Publishers.
- Holland, P. W., Dorans, N. J. & Petersen, N. S. (2007). Equating test scores. In Rao, C. R., Sinharay, S. (Eds.), *Handbook of statistics: Psychometrics* (Vol. 26, pp. 169-197). Amsterdam: Elsevier B. V.
- Holland, P. W., von Davier, A. A., Sinharay, S. & Han, N. (2006). *Testing the untestable assumptions of the chain and poststratification equating methods for the neat design*. Research Report, Educational Testing Service, Princeton, NJ.
- Holland, P. W., Sinharay, S., von Davier, A. A., & Han, N. (2008). An approach to evaluating the missing data assumptions of the chain and post-stratification equating methods for the neat design. *Journal of Educational Measurement*, 45, 17-43.
- Hou, J. (2007). *Effectiveness of the hybrid Levine equipercentile and modified frequency estimation equating methods under the common-item nonequivalent groups design*. Unpublished doctorate thesis, University of Iowa, Iowa City.
- Hu, H., Rogers, T. W. & Vukmirovic, Z. (2008). Investigation of IRT-based equating methods in the presence of outlier common items. *Applied Psychological Measurement*, 32(4), 311-333.
- Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory: Applications to psychological measurement*. Homewood IL: Dow Jones-Irwin.

- Ironson, G.H. (1983). Using item response theory to measure bias. In R.K. Hambleton (Ed.), *Applications of item response theory* (pp 155–174). Vancouver: Educational Research Institute of British Columbia.
- Kan, A. (2011). Test Eşitleme: OKS testlerinin istatistiksel eşitliğinin sınanması. *Eğitim ve Bilim*, 36(160), 38-51.
- Kang, T., & Petersen, N. S. (2009). *Linking item parameters to a base scale*. Paper presented at the National Council on measurement in education, San Diego, CA.
- Kaskowitz, G. S., & De Ayala, R. J. (2001). The effect of error in item parameter estimates on the test response function method of linking. *Applied Psychological Measurement*, 25, 39-52.
- 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.
- Keller, L. A., & Keller, R. R. (2008). *A comparison of transformation methods and calibration methods on the classification of students over time*. Paper presented at the meeting of the Psychometric Society, Dover, NH.
- Keller, R. R., Kim, W., Nering, M. & Keller, L. A. (2007). *What breaks the equating? A preliminary investigation into threats to a five-year equating chain*. Paper presented at the 2007 AERA Annual Meeting, Chicago, IL.
- Keller, L. A., Skorupski, W. P., Swaminathan, H., & Jodoin, M. G. (2004, April). *An evaluation of the effect of the linking method on capturing changes in examinee distributions with mixed-format tests in an item response theory context*. Paper presented at the meeting of the National Council on Measurement in Education, San Diego, CA.

- Kim, S. (2006). A comparative study of IRT fixed parameter calibration methods. *Journal of Educational Measurement, 43*, 355-381.
- Kim, D.-I., Brennan, R. L., & Kolen, M. J. (2005). A comparison of IRT equating and beta 4 equating. *Journal of Educational Measurement, 42*, 77-99.
- Kim, D., Choi, S. W., Lee, G. & Um, K. R. (2008). A Comparison of the common-item and random-groups equating designs using empirical data. *International Journal of Selection and Assessment, 16*(2), 83-92.
- 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.-H., & Cohen, A. S. (2002). A comparison of linking and concurrent calibration under the graded response model. *Applied Psychological Measurement, 26*, 25-41.
- 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., & Lee, W. (2004). *IRT scale linking methods for mixed-format tests* (ACT Research Report). Iowa City, IA: ACT, Inc.
- Kim, S., & Lee, W. (2006). An extension of four IRT linking methods for mixed-format tests. *Journal of Educational Measurement, 43*(1), 53-76.
- Kim, S., Walker, M. E. & Mchale, F. (2010). Comparisons among designs for equating mixed-format tests in large-scale assessment. *Journal of Educational Measurement, 47*(1), 36-53.
- Kolen, M. J. (1981). Comparison of traditional and item response theory methods for equating tests. *Journal of Educational Measurement, 18*, 1-11.
- Kolen, M. J. (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational Statistics, 9*, 25-44.

- Kolen, M. J. (1985). Standard errors of Tucker equating. *Applied Psychological Measurement, 2*, 209-223.
- Kolen, M. J. (1988). An NCME instructional module on traditional equating methodology. *Educational Measurement: Issues and Practice, 7*, 29-36.
- Kolen, M. J. & Brennan, R. L. (1995). *Test Equating: methods and practices*. New York: Springer-Verlag.
- Kolen, M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd. ed.). New York: Springer.
- Lee, Y. S. (2007). A comparison of methods for nonparametric estimation of item characteristic curves for binary IRTs. *Applied Psychological Measurement, 31*, 121-134.
- Lee, G. & Fitzpatrick, A. R. (2008). A new approach to test score equating using item response theory with fixed c-parameters. *Asia Pacific Education Review, 3*, 248-261.
- Lee, W. C. & Ban, J. C. (2010). Comparison of IRT Linking Procedures. *Applied Measurement in Education, 23*(1), 23-48.
- Lehman, R. S., & Bailey, D. E. (1968). *Digital computing: Fortran IV and its applications in behavioural science*. New York: John Wiley.
- Lei, P.-W., Dunbar, S. B., & Kolen, M. J. (2004). A comparison of parametric and nonparametric approaches to item analysis for multiple-choice tests. *Educational and Psychological Measurement, 64*, 565-587.
- Li, Y. H. & Lissitz, R. W. (2000). An evaluation of multidimensional IRT linking. *Applied Psychological Measurement, 24*, 115-138.

- Li, D. (2009). *Developing a common scale for testlet model parameter estimates under the common-item nonequivalent groups design*. Unpublished doctorate thesis, University of Maryland.
- Li, Y. H., Tam, H. P. & Tompkins, L. J. (2004). A comparison of using the fixed common-precalibrated parameter method and the matched characteristic curve method for linking multiple-test items. *International Journal of Testing*, 4(3), 267-293.
- Liou, M. & Bond, L. (1985). *A Monte Carlo simulation for evaluating item response theory equating*. Paper presented at the annual meeting of the American of the American Educational Research Association, Chicago.
- 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), 349-69.
- Livingston, S. A. (1993a). *An empirical tryout of kernel equating* (ETS Research Rep. No. RR-93-33). Princeton, NJ: ETS.
- Livingston, S. (1993b). Small-sample equatings with log-linear smoothing. *Journal of Educational Measurement*, 30, 23-39.
- Livingston, S. A. (2004). *Equating test scores (without IRT)*. Princeton, NJ: ETS
- Livingston, S. A., Dorans, N. J., & Wright, N. K. (1990). What combination of sampling and equating methods works best? *Applied Measurement in Education*, 3(1), 73-95.
- Lord, F. M. (1977). Practical applications of item characteristic curve theory. *Journal of Educational Measurement*, 14(2), 117-138

- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, N. J.: Lawrence Erlbaum.
- Lord, F.M. (1983). Statistical bias in maximum likelihood estimators of item parameters. *Psychometrika*, 48, 477-482
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Loyd, B. H. & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, 17(3), 179-193.
- Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, 14, 139-160.
- Mao, X., von Davier, A. A., & Rupp, S. (2006). *Comparisons of the kernel equating method with the traditional equating methods on PRAXIS data* (ETS RR-06-30). New Jersey: Educational Testing Service.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64, 187-212.
- McDonald, R. P. (1997). Normal-ogive multidimensional model. In W. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 258-269). New York: Springer-Verlag.
- McKinley, R. L. & Reckase, M. D. (1981). *A comparison of procedures for constructing large item pools* (Research Report 81-3). Missouri: University of Missouri, Department of Educational Psychology.
- Michaelides, M. P. & Haertel, E. H. (2004). *Sampling of common items: An unrecognized source of error in test equating* (CSE Report 636). Los Angeles: The Regents of the University of California.

- Mislevy, R. J. & Stocking, M. L. (1989). A consumer's guide to LOJIST and BILOG. *Applied Psychological Measurement*, 13, 57-75.
- Mohandas, R. (1998). *Test equating, problems and solutions: Equating English test forms for Indonesian junior secondary school final examinations administered in 1994*. Unpublished masters thesis, The Flinders University of South Australia.
- Mokken, R. J. (1971). *A theory and procedure of scale analysis*. The Hague: Mouton.
- Monseur, C., Sibberns, H. & Hastedt, D. (2006). *Equating errors in international surveys in education*. The Second IEA International Research Conference: Proceedings of the IRC-2006, pp. 61-65.
- Muraki, E., & Bock, R. (1991). *PARSCALE: Parametric scaling of rating data*. Chicago: Scientific Software International, Inc.
- Muraki, E., Hombo, C. M., & Lee, Y.-W. (2000). Equating and linking of performance assessments. *Applied Psychological Measurement*, 24, 325-337.
- Nandakumar, R. (1994). Assessing dimensionality of a set of item responses: Comparison of different approaches. *Journal of Educational Measurement*, 31(1), 17-35.
- Norman-Dvorak, R. L. (2009). *A comparison of kernel equating to the test characteristic curve method*. Unpublished doctorate thesis, University of Nebraska, Lincoln, Nebraska.
- Nozawa, Y. (2008). *Comparison of parametric and nonparametric IRT equating methods under the common-item nonequivalent groups design*. Unpublished doctorate thesis, The University of Iowa, Iowa City.

- Ngudratoke, S. (2009). *An investigation of using collateral information to reduce equating biases of the post-stratification equating method*. Unpublished doctorate thesis, Michigan State University.
- Ogasawara, H. (2000). Asymptotic standard errors of IRT equating coefficients using moments. *Economic Review (Otaru University of Commerce)*, 51(1), 1-23.
- Ogasawara, H. (2001). Standard errors of item response theory equating/linking by response function methods. *Applied Psychological Measurement*, 25, 53-67.
- Oh, S. (2000). *Comparison of traditional and item response theory equating using arm and shoulder girdle muscular strength and endurance tests*. Doctorate thesis, University of Georgia, Athens, Georgia.
- Öztürk, N. (2010). *Akademik personel ve lisansüstü eğitimi giriş sınavı puanlarının eşitlenmesi üzerine bir çalışma*. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Petersen, N. S., Kolen, M. J. & Hoover, H. D. (1989). Scaling, norming and equating. In: Linn, R. L. (Ed.) *Educational Measurement* (pp.221-262). New York: Macmillan.
- Petersen, N. S., Marco, G. L., & Stewart, E. E. (1982). A test of the adequacy of linear scores equating models. In P. W. Holland & D. B. Rubin (Eds.) *Test Equating*, 71-136. New York: Academic Press.
- Puhan, G. (2010). A comparison of chained linear and poststratification linear equating under different testing conditions. *Journal of Educational Measurement*, 47(1), 54-75.
- RaOSay, J. O. (1991). Kernel smoothing approach to nonparametric item characteristic curve estimation. *Psychometrika*, 56, 611-630.

- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*, 207-230.
- Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement, 21*(1), 25-36.
- Reckase, M. D., Ackerman, T. A., & Carlson, J. (1988). Building a unidimensional test using multidimensional item response theory. *Journal of Educational Measurement, 25*, 193–203.
- Sinharay, S. & Holland, P. W. (2007). Is it necessary to make anchor tests mini versions of the tests being equated or can some restrictions be relaxed? *Journal of Educational Measurement, 44*(3), 249-275.
- Sinharay, S. & Holland, P. W. (2008). *The missing data assumptions of the nonequivalent groups with anchor test (neat) design and their implications for test equating* (ETS RR-09-16 Research report). Princeton NJ: Educational Testing Service.
- Skaggs, G. & Lissitz, R. W. (1982). *Test equating, relevant issues and a review of recent research*. Paper presented at the annual meeting of the American Educational Research Association. Los Angeles, CA. Eric document ED 241 540.
- Skaggs, G. & Lissitz, R. W. (1986). IRT Test equating: relevant issues and a review of recent research. *Review of Educational Research, 56*(4), 495-529.
- Skaggs, G. & Stevenson, J. (1989). A comparison of pseudo-bayesian and joint maximum likelihood procedures for estimating item parameters in the three-parameter IRT model. *Applied Psychological Measurement, 13*(4), 391-402.
- Slinde, J. A. & Linn, R. L. (1979). A note on vertical equating via the rasch model for groups of quite different ability and tests of quite different difficulty. *Journal of Educational Measurement, 16*(3), 159–165.

- Spence, P. (1996). *The effect of multidimensionality on unidimensional equating with item response theory*. Doctorate thesis, University of Florida, FL.
- Speron, E. (2009). *A comparison of metric linking procedures in Item Response Theory*. Unpublished doctorate thesis, University of Illinois, Chicago, Illinois.
- Stocking, M. L. & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201-210.
- Suanthong, S. (1988). *An investigation of factors affecting test equating in latent trait theory*. Unpublished doctoral dissertation. University of North Texas, Denton.
- Sukin, T. & Keller, L. (2008). *The effect of deleting anchor on the classification of examinees*. Northeastern educational research association (NERA) Annual Conference. NERA Conference Proceedings.
- Swaminathan, H. & Gifford, J. A. (1983). Estimation of parameters in the three-parameter latent trait model. In D. Weiss (Ed.), *New horizons in testing*. New York: Academic Press.
- Thorndike, R. L. (1971). *Educational measurement*. Washington, D.C.: American Council on Education.
- 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.
- Tsai, T. H., Hanson, A. B., Kolen, J. M. & Forsyth, A. R. (2001). A comparison of bootstrap standard errors of IRT equating methods for the common-item nonequivalent groups design. *Applied Measurement in Education*, 14, 17-30.

- Tong, Y. & Kolen, M. (2005). Assessing equating results on different equating criteria. *Applied Psychological Measurement, 29*, 418-432.
- Turhan, A. (2006). *Multilevel 2PL item response model vertical equating with the presence of differential item functioning*. Unpublished doctorate dissertation, Florida State University.
- Turhan, Tong & Um (April, 2007). *Effects of anchor item properties and dimensionality of test on vertical scaling*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.
- Vale, C. D. (1986). Linking item parameters onto a common scale. *Applied Psychological Measurement, 10*(4), 333-344.
- Van der Linden, W. J. (2006). Equating error in observed-score equating. *Applied Psychological Measurement, 30*, 355-378.
- von Davier, A. A., Holland, P. W. & Thayer, D. T. (2004). *The kernel method of test equating*. New York: Springer.
- von Davier, A. A. & Wilson, C. (2008). Investigating the population sensitivity assumption of item response theory true-score equating across two subgroups of examinees and two test formats. *Applied Psychological Measurement, 32*(1), 11-26.
- Yang, W. L. & Houang, R. T. (1996). *The effect of anchor length and equating method on the accuracy of test equating: comparisons of linear and IRT-based equating using an anchor-item design*. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Yen, W. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika, 52*, 275-291.

- Walker, M. E. & Kim, S. (2009, April). *Linking mixed-format tests using multiple choice anchors*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Wang, T., Lee, W.-C., Brennan, R. J. & Kolen, M. J. (2008). A comparison of the frequency estimation and chained equipercentile methods under the common-item non-equivalent groups design. *Applied Psychological Measurement, 32*, 632-651.
- Wang, T.-Y., Hanson, B. A. & Harris, D. J. (2000). The effectiveness of circular equating as a criterion for evaluating equating. *Applied Psychological Measurement, 24*(3), 195-210.
- Way, W. D., & Tang, K. L. (1991, April). *A comparison of four logistic model equating methods*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Wells, C. S. (2004). *Investigation of model misfit in IRT and a new approach based on simultaneous parametric and nonparametric IRT estimation*. Unpublished doctoral dissertation. University of Wisconsin-Madison.
- Wingersky, M. S. (1983). LOGIST: A program for computing maximum likelihood procedures for logistic test models. In R. K. Hambleton (Ed.), *Applications of item response theory* (pp. 45-56). Vancouver, BC: Educational Research Institute of British Columbia.
- Wingersky, M. S., Cook, L. L. & Eignor, D. R. (1987). *Specifying the characteristics of linking item used for item response* (ECT Research Report 87-24). Princeton NJ: Educational Testing Service.

- Wingersky, M. S. & Lord, F. M. (1984). An investigation of methods for reducing sampling error in certain IRT procedures. *Applied Psychological Measurement*, 8, 347–364.
- Whittington, C. D. (1986). *Guessing, sample size, and equating with item response theory*. Unpublished doctorate thesis, Columbia University, Ohio, Ontario.
- Woldbeck, T. (1998). *Basic concepts in modern methods of test equating*. Paper presented at the annual meeting of the Southwest Psychological Association, New Orleans.
- Wolkowitz, A. A. (2008). *A comparison of classical test theory and item response theory methods for equating number-right scored to formula scored assessments*. Unpublished doctorate thesis, University of Kansas.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA Pres.
- Wu, N., Huang, C-Y., Huh, N., & Harris, D. (2009). *Robustness in using multiple choice items as an external anchor for constructed-response test equating*. Paper presented at the annual conference of the National Council on Measurement in Education, San Diego, CA.
- Zeng, L. (1991). *Standard errors of linear equating for the single-group design*. ACT Research Report Series. 91–4.
- Zhao, Y. (2008). *Approaches for addressing the fit of item response theory models to educational test data*. Unpublished doctorate thesis, University of Massachusetts.
- Zhang, Z. (2010). *Comparison of different equating methods and an application to link testlet-based tests*. Doctorate thesis, Villanova University, Philadelphia.
- Zhu, W. (1998). Test equating: What, why, who?. *Research Quarterly for Exercise and Sport*, 69(1), 11–23.

EKLER

Ek 1:

Tablo 1. 2PLM 1000 kiři normal dađılım 30 madde

Maddeler	Form1		Form2		Maddeler	Form1		Form2	
	a	b	a	b		a	b	a	b
1	0.925	-1.049	0.925	-1.049	16	0.781	0.613	1.703	-0.358
2	1.745	1.141	1.745	1.141	17	0.516	-1.038	1.313	1.433
3	0.777	-0.177	0.777	-0.177	18	0.698	0.283	1.422	-1.474
4	0.956	0.459	0.956	0.459	19	0.720	-0.213	0.495	1.104
5	0.534	1.253	0.534	1.253	20	0.455	-1.180	0.541	-0.289
6	0.660	1.036	0.66	1.036	21	1.092	-0.021	0.764	0.295
7	1.435	-1.806	1.435	-1.806	22	1.098	-0.223	0.618	-0.627
8	0.969	-0.843	0.969	-0.843	23	2.591	-0.077	2.46	0.695
9	0.714	-0.660	0.714	-0.66	24	1.590	-1.024	0.821	0.272
10	0.634	-1.326	0.634	-1.326	25	0.536	2.877	2.361	-0.454
11	1.479	0.760	0.793	0.301	26	1.111	0.651	0.621	-0.23
12	1.065	-1.407	0.798	1.194	27	2.711	-0.298	1.57	-0.211
13	1.097	1.099	1.124	-0.237	28	0.952	1.417	1.288	-1.813
14	0.797	-0.365	0.96	0.887	29	1.066	0.954	0.375	-0.259
15	0.623	-0.713	1.286	1.276	30	0.753	-0.388	1.701	1.938

1-10 arasındaki maddeler ortak maddelerdir.

Tablo 2. 2PLM 1000 kişi normal dağılım 60 madde

m	Form1		Form2		m	Form1		Form2	
	a	b	a	b		a	b	a	b
1	2.378	0.983	2.378	0.983	31	1.898	2.155	2.055	-0.1
2	0.468	0.209	0.468	0.209	32	0.995	1.437	0.676	0.501
3	2.077	0.019	2.077	0.019	33	0.373	-1.09	0.926	1.105
4	0.58	-0.153	0.58	-0.153	34	0.411	0.951	1.104	0.243
5	0.84	0.457	0.84	0.457	35	0.622	1.119	1.19	-0.136
6	0.76	0.564	0.76	0.564	36	1.148	-0.911	0.844	2.385
7	1.324	0.336	1.324	0.336	37	0.918	1.186	1.947	2.81
8	0.702	-0.65	0.702	-0.65	38	0.717	-0.002	1.368	-0.64
9	0.603	0.101	0.603	0.101	39	1.088	0.063	0.838	-1.692
10	1.986	-0.15	1.986	-0.15	40	0.894	0.509	1.775	0.164
11	0.804	-0.048	0.804	-0.048	41	0.773	1.003	0.588	2.841
12	1.059	-1.174	1.059	-1.174	42	0.447	-0.101	0.472	1.516
13	1.165	0.146	1.165	0.146	43	1.674	0.381	0.841	0.502
14	0.725	3.174	0.725	3.174	44	2.906	-0.788	1.181	-0.455
15	0.619	1.105	0.619	1.105	45	1.782	-1.451	1.374	0.212
16	0.568	0.7	0.568	0.7	46	1.178	-0.272	0.963	0.799
17	0.339	0.474	0.339	0.474	47	1.348	1.108	1.828	0.796
18	1.415	-1.033	1.415	-1.033	48	0.518	-0.737	0.721	2.092
19	0.524	0.002	0.524	0.002	49	1.654	0.039	1.009	0.461
20	1.146	-0.205	1.146	-0.205	50	0.885	-0.894	0.685	0.491
21	1.833	-1.41	2.485	1.251	51	1.471	-0.262	0.84	-0.22
22	1.009	0.412	1.936	2.21	52	1.118	-0.773	1.243	0.822
23	0.742	-0.044	0.382	1.054	53	0.483	1.875	0.598	-0.33
24	1.903	1.463	1.726	0.848	54	0.869	0.025	1.485	0.589
25	1.13	-1.498	0.853	1.773	55	1.204	0.737	1.244	1.406
26	0.538	0.224	0.957	-0.748	56	1.293	-0.456	1.242	-0.95
27	1.017	0.185	1.032	0.025	57	2.997	-0.624	0.661	0.64
28	0.815	-0.938	2.066	-0.27	58	0.122	0.514	0.76	-0.419
29	0.701	0.177	1.153	-0.09	59	0.936	0.121	2.709	0.315
30	0.881	-1.258	1.351	2.481	60	0.449	1.422	0.854	1.051

1-20 arasındaki maddeler ortak maddelerdir.

Tablo 3. 2PLM 1000 kişi normal dağılım 80 madde

m	Form1		Form2		m	Form1		Form2	
	a	b	a	b		a	b	a	b
1	1.27	-0.32	1.27	-0.32	41	2.541	-0.069	1.391	1.262
2	1.153	0.308	1.153	0.308	42	1.466	0.915	1.242	1.775
3	1.125	-0.82	1.125	-0.82	43	0.311	-1.201	0.703	0.776
4	0.537	0.657	0.537	0.657	44	1.293	1.061	0.712	1.032
5	1.271	0.175	1.271	0.175	45	0.949	-0.741	0.876	-0.771
6	0.537	0.985	0.537	0.985	46	0.837	-0.085	0.741	0.292
7	0.776	0.989	0.776	0.989	47	1.345	1.062	0.935	-0.59
8	1.167	1.507	1.167	1.507	48	0.846	1.122	0.975	-1.41
9	1.104	-1.171	1.104	-1.171	49	1.405	0.658	0.886	1.052
10	1.139	-0.924	1.139	-0.924	50	0.423	-0.913	0.95	-0.78
11	1.173	1.201	1.173	1.201	51	0.568	0.724	0.914	0.01
12	0.523	-0.232	0.523	-0.232	52	0.793	-1.104	0.352	0.518
13	1.255	-0.581	1.255	-0.581	53	1.171	-0.893	1.445	0.8
14	0.831	-0.022	0.831	-0.022	54	1.293	-0.052	1.135	-0.468
15	0.858	0.572	0.858	0.572	55	1.331	-0.735	0.949	1.429
16	0.875	0.311	0.875	0.311	56	0.896	0.197	0.431	1.371
17	0.604	0.082	0.604	0.082	57	0.596	0.032	0.927	1.888
18	2.077	1.046	2.077	1.046	58	2.282	0.908	0.451	0.88
19	0.52	-0.508	0.52	-0.508	59	3.036	0.668	1.03	0.04
20	1.798	0.047	1.798	0.047	60	0.99	-0.63	2.144	-0.318
21	1.038	0.162	1.038	0.162	61	0.503	-0.447	0.979	0.171
22	0.682	0.084	0.682	0.084	62	0.999	-0.284	2.025	-0.351
23	0.461	-0.361	0.461	-0.361	63	0.81	0.294	1.581	-0.665
24	1.878	-0.272	1.878	-0.272	64	1.101	0.445	0.436	0.661
25	2.081	-0.233	2.081	-0.233	65	4.107	0.043	1.045	0.235
26	0.743	2.229	0.743	2.229	66	1.112	-1.042	0.608	0.498
27	1.575	-0.548	1.575	-0.548	67	1.431	-0.791	1.034	0.128
28	1.008	-0.614	1.008	-0.614	68	1.881	-0.203	0.818	-0.185
29	1.229	-0.121	1.229	-0.121	69	1.76	-0.654	0.928	1.503
30	0.449	-0.884	0.449	-0.884	70	1.145	-0.027	1.339	1.15
31	2.03	0.64	2.638	0.993	71	0.791	-0.483	1.673	-0.622
32	2.118	0.358	0.342	-1.565	72	0.605	0.234	0.986	1.457
33	1.015	0.261	1.447	0.7	73	1.048	-0.821	0.885	0.671
34	0.954	0.253	1.079	-0.854	74	1.128	-1.081	0.74	1.009
35	1.74	-0.025	0.963	-0.553	75	1.551	1.412	0.688	-0.61
36	0.595	0.289	0.875	0.725	76	0.898	-0.011	2.247	-0.587
37	1.227	-1.028	0.825	0.896	77	0.801	0.236	0.803	1.276
38	0.843	0.396	1.754	0.538	78	0.875	1.207	0.73	-0.85
39	1.261	-0.399	1.686	1.104	79	0.874	0.184	0.994	0.632
40	1.065	-1.729	1.197	-0.605	80	0.573	0.348	2.299	-1.201

1-30 arasındaki maddeler ortak maddelerdir.

Tablo 4. 2PLM 3000 kiři normal dađılım 30 madde

Maddeler	Form1		Form2		Maddeler	Form1		Form2	
	a	b	a	b		a	b	a	b
1	1.459	0.795	1.459	0.795	16	1.216	1.127	1.379	0.859
2	1.085	-0.983	1.085	-0.983	17	0.938	-0.302	1.501	-0.942
3	2.013	1.06	2.013	1.06	18	1.174	-1.277	0.468	1.798
4	1.026	0.195	1.026	0.195	19	1.067	1.03	0.598	0.64
5	1.441	-0.55	1.441	-0.55	20	0.398	-0.794	1.321	1.29
6	0.59	-0.121	0.59	-0.121	21	0.919	-0.453	1.131	-0.893
7	0.561	-0.21	0.561	-0.21	22	1.569	-1.039	2.34	1.728
8	1.177	-0.23	1.177	-0.23	23	0.517	-0.2	2.019	1.662
9	0.876	-0.718	0.876	-0.718	24	0.971	-0.418	0.771	1.491
10	1.673	0.118	1.673	0.118	25	0.647	-1.048	0.974	-0.191
11	1.74	-0.585	0.53	1.938	26	0.63	0.622	0.832	0.613
12	0.56	-0.464	2.862	-0.342	27	0.551	0.18	0.898	0.92
13	2.075	0.966	1.167	1.107	28	1.017	1.091	0.905	0.196
14	1.157	0.402	1.126	0.652	29	1.024	-0.545	1.585	1.539
15	0.882	0.032	0.988	1.615	30	0.362	-0.042	0.294	1.234

1-10 arasındaki maddeler ortak maddelerdir.

Tablo 5. 2PLM 3000 kişi normal dağılım 60 madde

Maddeler	Form1		Form2		Mad.	Form1		Form2	
	a	b	a	b		a	b	a	b
1	1.56	1.012	1.56	1.012	31	1.265	0.189	1.25	-0.271
2	0.963	0.841	0.963	0.841	32	0.983	-0.091	1.065	1.944
3	0.929	0.872	0.929	0.872	33	1.419	-2.463	0.527	0.132
4	1.038	-1.137	1.038	-1.137	34	3.247	-0.658	1.431	0.481
5	1.215	-0.085	1.215	-0.085	35	0.954	1.327	0.557	0.006
6	1.111	0.62	1.111	0.62	36	2.389	-1.18	0.795	1.253
7	1.735	0.671	1.735	0.671	37	0.544	0.261	0.464	0.613
8	1.826	-1.147	1.826	-1.147	38	0.861	-2.295	0.781	1.604
9	1.16	-0.808	1.16	-0.808	39	0.719	-0.296	2.193	0.31
10	1.067	0.095	1.067	0.095	40	2.455	1.024	1.353	0.322
11	0.828	-1.494	0.828	-1.494	41	1.203	0.574	0.714	0.041
12	0.901	0.299	0.901	0.299	42	2.625	0.544	1.424	0.767
13	0.763	-1.51	0.763	-1.51	43	1.152	0.117	1.539	1.112
14	2.285	-0.052	2.285	-0.052	44	1.855	0.176	1	-0.535
15	1.322	-0.602	1.322	-0.602	45	2.066	1.092	0.843	0.861
16	1.015	0.175	1.015	0.175	46	1.024	-1.819	0.354	0.26
17	1.115	0.758	1.115	0.758	47	1.479	-0.851	1.111	0.696
18	1.064	0.991	1.064	0.991	48	0.765	-0.401	0.983	0.9
19	0.806	-0.086	0.806	-0.086	49	0.689	-1.176	0.746	0.399
20	1.668	0.355	1.668	0.355	50	0.515	0.921	1.133	0.471
21	1.641	0.001	0.552	-0.792	51	0.676	-1.06	0.762	1.228
22	0.598	1.544	0.855	1.125	52	0.736	-0.74	0.752	0.318
23	0.659	0.277	1.13	0.651	53	1.557	0.1	0.547	2.251
24	0.724	-0.269	2.057	-0.363	54	2.125	-0.856	0.515	0.682
25	4.507	-0.443	1.12	1.382	55	0.859	-0.894	0.719	-0.399
26	1.088	-1.316	0.721	2.211	56	0.935	0.008	0.475	0.903
27	2.386	0.772	0.581	0.204	57	1.057	-1.459	1.306	0.352
28	1.11	-0.242	2.157	0.847	58	0.388	0.478	0.745	-0.163
29	0.534	0.604	0.272	0.129	59	0.72	-0.455	1.238	-1.254
30	1.325	-0.187	0.756	-0.261	60	0.979	0.797	1.73	1.27

1-20 arasındaki maddeler ortak maddelerdir.

Tablo 6. 2PLM 3000 kişi normal dağılım 80 madde

Maddeler	Form1		Form2		Mad.	Form1		Form2	
	a	b	a	b		a	b	a	b
1	0.625	1.284	0.625	1.284	41	0.965	0.103	1.641	-0.285
2	1.005	1.607	1.005	1.607	42	1.039	-0.031	1.414	-0.991
3	1.045	-0.306	1.045	-0.306	43	0.594	2.09	1.518	1.384
4	0.591	2.28	0.591	2.28	44	0.454	0.469	2.041	0.459
5	1.558	-1.101	1.558	-1.101	45	1.36	-0.899	1.721	0.772
6	0.868	-0.858	0.868	-0.858	46	1.129	-0.244	0.746	-0.639
7	1.601	-0.808	1.601	-0.808	47	1.2	0.203	0.882	0.014
8	1.164	-2.789	1.164	-2.789	48	0.435	0.231	0.603	1.429
9	1.246	0.789	1.246	0.789	49	1.113	-0.282	1.267	2.265
10	1.967	-0.578	1.967	-0.578	50	0.7	0.293	1.727	-0.951
11	1.344	0.527	1.344	0.527	51	0.349	0.666	1.911	0.279
12	0.897	-1.259	0.897	-1.259	52	0.508	-2.701	0.803	1.453
13	1.689	-1.489	1.689	-1.489	53	1.037	1.571	0.51	0.474
14	0.776	0.039	0.776	0.039	54	1.335	1.253	1.487	1.644
15	0.989	0.44	0.989	0.44	55	0.714	-0.06	0.628	0.157
16	0.877	-0.378	0.877	-0.378	56	0.771	0.765	1.109	-1.634
17	2.802	1.132	2.802	1.132	57	1.24	-0.664	2.39	1.145
18	0.952	0.349	0.952	0.349	58	1.274	-0.479	1.029	0.19
19	0.891	0.386	0.891	0.386	59	1.103	0.318	2.086	0.004
20	0.322	0.564	0.322	0.564	60	0.843	-1.855	2.409	1.109
21	0.807	-0.269	0.807	-0.269	61	1.714	0.781	1.232	0.174
22	1.913	0.191	1.913	0.191	62	1.153	0.318	1.142	0.551
23	1.853	-0.769	1.853	-0.769	63	0.985	0.616	0.528	0.287
24	0.824	-0.781	0.824	-0.781	64	0.869	-0.405	2.28	1.659
25	1.232	-0.263	1.232	-0.263	65	0.623	-0.241	0.729	0.883
26	1.297	0.563	1.297	0.563	66	1.574	-0.633	0.604	-1.24
27	0.869	-0.489	0.869	-0.489	67	0.573	-0.435	0.826	2.073
28	0.525	-0.193	0.525	-0.193	68	2.204	1.014	0.997	1.161
29	1.158	0.747	1.158	0.747	69	0.762	0.096	1.066	-0.293
30	1.996	-1.254	1.996	-1.254	70	1.317	-0.125	0.906	-0.303
31	1.343	1.066	0.707	1.472	71	2.056	0.356	1.087	1.272
32	0.572	-0.288	1.613	-1.307	72	0.87	0.619	2.014	1.211
33	0.816	-0.796	1.074	2.764	73	0.498	0.586	1.473	0.752
34	0.703	-2.362	0.371	0.206	74	1.171	-0.478	0.997	1.148
35	0.594	-0.802	1.463	0.317	75	2.524	0.638	0.979	0.65
36	1.149	0.188	1.058	0.106	76	0.742	-0.468	1.272	-0.482
37	0.774	-0.285	0.478	0.92	77	1.131	0.496	0.977	0.241
38	0.726	0.527	0.648	1.588	78	1.192	0.177	0.529	-0.551
39	0.691	0.135	0.235	0.924	79	0.944	0.372	0.624	0.829
40	1.047	1.178	0.55	1.761	80	0.289	-0.285	0.805	1.123

1-30 arasındaki maddeler ortak maddelerdir.

Tablo 7. 2PLM 1000 kiři pozitif arpık 30 madde

Maddeler	Form1		Form2		Mad.	Form1		Form2	
	a	b	a	b		a	b	a	b
1	1.482	0.429	1.482	0.429	16	0.526	0.095	0.525	1.383
2	1.258	0.088	1.258	0.088	17	0.369	0.345	1.546	0.295
3	0.493	1.45	0.493	1.45	18	0.526	0.162	0.985	-1.187
4	1.247	-0.737	1.247	-0.737	19	1.267	0.859	1.022	0.818
5	0.474	-0.856	0.474	-0.856	20	0.981	0.503	1.531	1.464
6	1.167	0.01	1.167	0.01	21	2.071	-2.637	0.812	1.723
7	1.078	0.229	1.078	0.229	22	2.134	0.838	1.517	1.411
8	1.396	1.557	1.396	1.557	23	0.686	0.627	0.811	0.153
9	2.025	-0.726	2.025	-0.726	24	1.181	0.196	1.484	-2.222
10	0.734	0.061	0.734	0.061	25	0.711	-0.18	0.913	0.483
11	0.58	0.421	0.836	1.574	26	1.383	-0.768	0.602	-0.781
12	0.828	0.139	1.031	1.223	27	0.893	0.019	0.851	0.09
13	2.054	-0.851	0.58	2.36	28	1.002	-1.408	0.803	2.444
14	0.973	-0.453	1.086	-0.547	29	0.913	1.276	0.62	-0.316
15	1.642	-1.011	0.803	1.711	30	1.701	0.793	0.803	0.497

1-10 arasındaki maddeler ortak maddelerdir.

Tablo 8. 2PLM 1000 kişi pozitif çarpık 60 madde

Maddeler	Form1		Form2		Mad.	Form1		Form2	
	a	b	a	b		a	b	a	b
1	0.893	-0.456	0.893	-0.456	31	0.448	-1.084	1.615	0.585
2	2.788	-0.775	2.788	-0.775	32	0.958	-1.729	0.709	0.324
3	1.206	0.421	1.206	0.421	33	2.4	0.68	1.87	-0.387
4	0.859	-0.483	0.859	-0.483	34	0.369	2.167	0.619	-0.114
5	2.243	0.014	2.243	0.014	35	0.645	0.522	1.655	0.074
6	1.85	-0.003	1.85	-0.003	36	2.755	-1.139	0.958	0.66
7	1.31	-1.051	1.31	-1.051	37	0.839	0.315	0.414	1.036
8	1.143	-2.325	1.143	-2.325	38	1.459	0.546	1.795	-0.76
9	1.142	0.455	1.142	0.455	39	1.276	-0.189	0.436	-0.127
10	0.848	-0.509	0.848	-0.509	40	0.875	-0.18	1.118	-1.231
11	0.391	-0.248	0.391	-0.248	41	1.681	1.288	0.925	0.508
12	1.361	0.157	1.361	0.157	42	0.809	-1.025	1.537	1.474
13	1.286	0.745	1.286	0.745	43	4.058	-0.492	0.591	-1.417
14	0.457	1.127	0.457	1.127	44	2.219	0.064	1.221	-0.036
15	1.404	-0.111	1.404	-0.111	45	1.102	-0.654	1.806	-0.634
16	0.579	-0.316	0.579	-0.316	46	1.832	0.184	1.65	1.169
17	0.703	0.232	0.703	0.232	47	0.915	0.349	0.939	-0.486
18	0.466	1.303	0.466	1.303	48	1.858	0.36	1.037	0.046
19	0.884	1.827	0.884	1.827	49	1.429	-0.615	0.794	-1.166
20	1.295	2.064	1.295	2.064	50	1.645	-0.305	0.959	1.309
21	0.992	1.543	0.71	0.959	51	0.662	1.476	0.75	2.214
22	1.421	-0.151	1.835	2.082	52	0.797	-0.263	0.287	0.547
23	1.391	0.215	1.184	1.507	53	0.684	0.925	1.324	-0.813
24	0.282	-0.781	0.563	0.573	54	0.421	-0.195	0.553	-0.718
25	0.597	-1.538	0.417	1.281	55	0.96	0.616	0.42	0.564
26	0.491	1.548	1.167	2.327	56	0.716	1.185	0.549	0.464
27	0.8	1.322	1.674	2.035	57	1.448	0.26	2.847	-0.011
28	0.752	-3.005	2.091	-1.025	58	0.77	-0.529	0.867	-0.022
29	2.821	-0.443	0.787	0.496	59	1.267	0.302	1.631	-1.174
30	1.487	0.944	1.223	-0.644	60	2.319	0.181	0.953	-1.278

1-20 arasındaki maddeler ortak maddelerdir.

Tablo 9. 2PLM 1000 kişi pozitif çarpık 80 madde

Maddeler	Form1		Form2		Mad.	Form1		Form2	
	a	b	a	b		a	b	a	b
1	1.24	-0.844	1.24	-0.844	41	0.254	1.183	1.2	1.011
2	0.743	0.415	0.743	0.415	42	1.276	-1.056	1.592	1.5
3	0.755	-2.302	0.755	-2.302	43	1.387	-0.032	0.629	-0.997
4	1.494	0.137	1.494	0.137	44	0.605	-0.27	0.885	-0.951
5	1.512	-0.725	1.512	-0.725	45	1.096	-0.456	0.761	-0.919
6	1.241	-0.771	1.241	-0.771	46	1.341	-0.163	1.749	-1.228
7	1.041	0.21	1.041	0.21	47	1.045	-0.352	0.527	-0.289
8	1.066	2.556	1.066	2.556	48	1.779	-0.982	1.464	-0.912
9	1.051	-1.168	1.051	-1.168	49	2.157	0.358	0.542	-0.688
10	1.113	0.114	1.113	0.114	50	0.506	-0.651	1.303	0.584
11	0.69	0.414	0.69	0.414	51	1.498	-0.17	0.496	0.166
12	0.436	-0.081	0.436	-0.081	52	1.029	0.691	1.646	-0.376
13	0.497	1.546	0.497	1.546	53	1.172	-1.894	0.985	0.13
14	1.53	1.142	1.53	1.142	54	1.126	-1.034	1.49	0.62
15	0.739	0.214	0.739	0.214	55	0.677	-0.12	0.966	0.701
16	3.276	-0.076	3.276	-0.076	56	0.864	0.824	0.528	1.349
17	0.769	-1.653	0.769	-1.653	57	0.815	0.585	1.238	0.975
18	1.056	1.873	1.056	1.873	58	1.508	1.34	0.67	0.607
19	0.77	-1.775	0.77	-1.775	59	1.182	0.026	0.452	0.253
20	0.633	0.198	0.633	0.198	60	1.783	0.043	0.731	0.095
21	1.256	-0.402	1.256	-0.402	61	0.857	0.334	0.494	1.555
22	1.515	0.938	1.515	0.938	62	1.507	-0.529	0.866	2.229
23	0.519	0.319	0.519	0.319	63	4.501	0.945	1.33	2.635
24	1.463	0.954	1.463	0.954	64	0.864	0.919	1.091	0.698
25	0.458	1.137	0.458	1.137	65	0.884	-0.196	0.859	0.083
26	0.823	-1.084	0.823	-1.084	66	0.929	0.908	0.847	2.352
27	0.946	1.166	0.946	1.166	67	1.212	-0.951	1.445	0.741
28	0.917	-0.517	0.917	-0.517	68	0.649	-0.728	0.876	-0.45
29	0.695	0.164	0.695	0.164	69	2.081	-1.886	1.212	2.032
30	0.698	0.994	0.698	0.994	70	1.646	-1.762	1.073	1.847
31	1.313	0.099	1.648	1.468	71	1.382	0.836	1.201	0.683
32	1.676	0.968	0.937	1.12	72	1.017	-0.946	0.994	0.047
33	0.773	-1.062	1.079	-0.975	73	0.666	1.405	0.925	2.117
34	1.243	-0.557	0.837	2.109	74	0.804	-1.45	1.107	0.429
35	0.84	-0.689	0.739	0.146	75	0.91	-1.973	0.336	2.018
36	1.16	-2.073	1.463	0.418	76	0.547	-0.644	1.087	0.329
37	1.152	1.66	0.626	1.059	77	0.533	-0.774	1.18	2.133
38	0.809	0.896	0.681	0.736	78	0.922	0.407	0.906	-0.809
39	0.917	-0.225	2.467	1.183	79	2.832	-0.243	1.136	0.521
40	0.842	0.209	0.473	1.664	80	1.081	-1.269	1.142	0.565

1-30 arasındaki maddeler ortak maddelerdir.

Tablo 10. 2PLM 3000 kiři pozitif arpık 30 madde

Maddeler	Form1		Form2		Mad.	Form1		Form2	
	a	b	a	b		a	b	a	b
1	1.785	1.511	1.785	1.511	16	1.637	-0.096	1.373	-0.718
2	0.803	-0.001	0.803	-0.001	17	1.636	-0.963	0.558	0.626
3	2.286	-1.288	2.286	-1.288	18	1.138	-0.338	1.447	2.457
4	0.853	1.366	0.853	1.366	19	1.504	-1.133	1.374	1.534
5	0.687	-0.036	0.687	-0.036	20	0.679	0.92	1.614	-0.295
6	2.299	0.709	2.299	0.709	21	1.699	-0.6	0.93	-0.296
7	1.168	-0.179	1.168	-0.179	22	0.662	-0.514	1.128	-0.414
8	1.349	-1.382	1.349	-1.382	23	0.812	-0.412	0.857	0.935
9	2.934	-0.132	2.934	-0.132	24	1.861	-1.576	0.84	0.736
10	2.074	0.122	2.074	0.122	25	0.731	1.162	2.045	-1.349
11	0.436	1.507	1.222	0.445	26	0.286	-0.057	0.534	0.731
12	0.476	1.912	2.52	0.237	27	1.082	1.941	0.993	1.69
13	1.767	2.022	0.564	1.734	28	0.493	0.219	2.189	0.91
14	1.086	-1.303	1.015	-2.159	29	1.408	-0.355	0.466	-0.865
15	0.636	1.12	0.961	-0.143	30	1.44	-1.426	1.315	-1.301

1-10 arasındaki maddeler ortak maddelerdir.

Tablo 11. 2PLM 3000 kişi pozitif çarpık 60 madde

Maddeler	Form1		Form2		Mad.	Form1		Form2	
	a	b	a	b		a	b	a	b
1	1.977	-0.18	1.977	-0.18	31	0.488	-0.164	1.198	0.494
2	0.911	0.236	0.911	0.236	32	0.862	-0.235	1.325	0.11
3	1.3	-0.803	1.3	-0.803	33	1.245	-0.424	1.03	-0.029
4	1.923	1.013	1.923	1.013	34	0.856	1.96	1.056	1.581
5	0.945	-0.01	0.945	-0.01	35	0.692	1.11	2.815	0.306
6	1.096	-0.653	1.096	-0.653	36	0.864	2.291	0.516	1.834
7	1.593	0.415	1.593	0.415	37	0.868	0.052	0.66	1.71
8	2.103	0.99	2.103	0.99	38	0.91	-0.097	1.618	-1.198
9	0.688	1.539	0.688	1.539	39	2.001	0.891	0.674	1.615
10	0.58	-1.552	0.58	-1.552	40	0.763	-1.33	0.924	1.521
11	0.817	1.354	0.817	1.354	41	0.323	1.818	0.504	0.534
12	1.195	-0.035	1.195	-0.035	42	0.639	0.844	1.053	0.696
13	0.798	-0.012	0.798	-0.012	43	2.033	-0.124	0.951	2.148
14	0.891	-0.21	0.891	-0.21	44	0.395	0.397	0.937	-2.382
15	0.624	-0.499	0.624	-0.499	45	1.071	-1.22	0.925	-0.359
16	0.519	1.125	0.519	1.125	46	0.435	0.295	1.083	-0.788
17	1.7	-1.024	1.7	-1.024	47	1.533	-0.851	0.453	-0.469
18	0.174	-1.391	0.174	-1.391	48	0.874	-0.23	0.791	0.606
19	0.309	-0.229	0.309	-0.229	49	0.4	-0.144	0.789	-0.316
20	1.649	0.526	1.649	0.526	50	0.365	0.368	0.717	0.924
21	1.669	0.764	1.46	0.587	51	0.66	2.529	1.029	1.963
22	1.711	0.615	1.21	0.769	52	0.364	0.54	3.192	-0.466
23	0.542	-1.06	1.405	1.37	53	1.418	0.155	1.653	1.146
24	3.197	0.312	1.918	-0.425	54	0.565	0.481	1.03	-0.058
25	1.458	-0.76	1.82	1.534	55	3.971	-0.736	2.038	0.072
26	0.798	1.725	1.076	-0.041	56	0.817	-0.579	0.453	0.197
27	1.202	-1.44	0.815	3.445	57	0.694	-0.452	1.563	-0.064
28	1.351	-0.269	0.643	0.066	58	1.259	-0.025	1.122	1.367
29	0.947	-1.174	1.1	-0.229	59	1.271	1.476	1.098	0.305
30	0.535	-0.234	0.488	0.934	60	1.127	-1.396	2.081	0.426

1-20 arasındaki maddeler ortak maddelerdir.

Tablo 12. 2PLM 3000 kişi pozitif çarpık 80 madde

Maddeler	Form1		Form2		Mad.	Form1		Form2	
	a	b	a	b		a	b	a	b
1	1.359	0.562	1.359	0.562	41	3.139	-0.467	0.961	-1.324
2	1.018	1.406	1.018	1.406	42	0.667	0.459	0.7	0.867
3	0.769	-0.426	0.769	-0.426	43	0.693	0.426	1.348	-1.155
4	0.976	0.996	0.976	0.996	44	1.083	1.762	4.935	0.428
5	0.522	-1.298	0.522	-1.298	45	0.948	0.165	1.616	0.896
6	0.683	-0.18	0.683	-0.18	46	0.82	-0.4	1.25	-0.188
7	0.629	-0.289	0.629	-0.289	47	0.759	1.125	1.841	0.937
8	1.86	-0.699	1.86	-0.699	48	2.395	-0.773	0.234	0.374
9	0.892	-1.327	0.892	-1.327	49	0.902	1.53	0.927	-0.29
10	0.501	1.721	0.501	1.721	50	0.523	-0.855	0.924	1.002
11	0.625	-1.435	0.625	-1.435	51	1.516	-0.312	0.358	1.479
12	0.706	0.444	0.706	0.444	52	1.144	0.742	1.343	0.809
13	1.742	0.387	1.742	0.387	53	2.265	-1.414	1.093	-0.327
14	0.956	-0.034	0.956	-0.034	54	2.163	-0.811	0.485	1.034
15	0.828	-0.328	0.828	-0.328	55	1.115	0.197	1.68	2.734
16	0.658	2.213	0.658	2.213	56	1.43	-0.053	1.552	1.754
17	1.271	0.849	1.271	0.849	57	0.629	-0.683	1.668	-0.816
18	0.543	0.397	0.543	0.397	58	1.611	0.838	0.623	0.408
19	0.948	1.869	0.948	1.869	59	1.125	1.032	0.664	0.924
20	0.78	0.305	0.78	0.305	60	1.052	-2.891	5.322	-0.705
21	1.808	0.487	1.808	0.487	61	0.498	-1.454	0.531	0.093
22	1.228	-0.816	1.228	-0.816	62	0.682	-0.797	0.525	-1.011
23	1.434	-0.006	1.434	-0.006	63	0.593	0.484	0.676	0.474
24	0.486	-1.458	0.486	-1.458	64	2.198	-1.913	0.694	0.748
25	0.556	-1.094	0.556	-1.094	65	1.438	1.41	0.766	0.609
26	0.854	-0.586	0.854	-0.586	66	2.282	1.019	0.401	0.519
27	2.146	-1.827	2.146	-1.827	67	0.638	1.069	0.923	-0.784
28	0.936	0.253	0.936	0.253	68	1.917	-0.402	1.424	-1.463
29	0.454	0.17	0.454	0.17	69	2.372	-0.173	1.394	1.136
30	0.265	-0.534	0.265	-0.534	70	0.845	-1.637	0.758	0.201
31	0.917	-1.448	1.39	0.479	71	1.811	-2.028	1.187	1.762
32	0.938	0.91	1.292	-0.22	72	1.188	0.639	1.098	1.367
33	0.774	-0.703	2.316	0.705	73	0.77	-1.42	0.497	0.378
34	0.923	1.545	1.18	0.174	74	1.09	-0.675	0.914	1.805
35	1.704	0.448	0.778	0.746	75	2.081	-0.205	1.79	0.079
36	2.447	0.184	0.875	2.156	76	1.163	-0.324	0.652	0.065
37	1.251	1.124	2.387	1.135	77	1.145	0.976	0.389	1.563
38	0.756	0.839	0.478	3.148	78	0.548	-1.705	1.149	1.438
39	0.608	-1.226	0.946	-1.061	79	0.671	-0.723	1.714	1.45
40	0.563	-1.627	0.572	-0.642	80	0.772	2.108	0.476	0.715

1-30 arasındaki maddeler ortak maddelerdir.

Tablo 13. 2PLM 1000 kişi negatif çarpık 30 madde

Maddeler	Form1		Form2		Mad.	Form1		Form2	
	a	b	a	b		a	b	a	b
1	1.005	-0.588	1.005	-0.588	16	0.366	1.446	0.583	-0.279
2	0.666	-0.118	0.666	-0.118	17	0.669	-2.284	1.143	1.577
3	0.837	-1.39	0.837	-1.39	18	0.587	-0.648	1.212	1.902
4	0.867	-0.381	0.867	-0.381	19	1.755	-0.864	0.782	-0.207
5	1.64	0.074	1.64	0.074	20	1.159	1.641	1.193	1.895
6	1.234	0.578	1.234	0.578	21	0.931	0.593	2.301	-0.579
7	2.118	-0.008	2.118	-0.008	22	1.431	0.204	0.954	1.215
8	0.845	-1.195	0.845	-1.195	23	1.474	1.353	0.854	2.145
9	0.698	0.354	0.698	0.354	24	0.913	0.423	1.35	-0.183
10	0.541	0.576	0.541	0.576	25	3.451	-0.598	1.01	0.134
11	0.641	0.4	0.706	0.653	26	1.171	0.58	0.816	1.514
12	1.115	-1.422	0.904	1.351	27	3.064	-0.295	1.102	-0.331
13	1.185	-0.981	1.286	-0.362	28	1.79	0.806	0.887	0.049
14	1.384	-1.065	1.158	0.95	29	1.489	0.757	2.15	0.165
15	1.247	-1.696	1.118	2.212	30	1.786	1.344	0.3	-0.773

1-10 arasındaki maddeler ortak maddelerdir.

Tablo 14. 2PLM 1000 kişi negatif çarpık 60 madde

Maddeler	Form1		Form2		Mad.	Form1		Form2	
	a	b	a	b		a	b	a	b
1	0.632	0.542	0.632	0.542	31	1.087	-0.644	1.252	2.404
2	1.685	1.511	1.685	1.511	32	0.628	0.042	0.26	0.955
3	0.79	0.848	0.79	0.848	33	0.825	0.821	2.308	-0.719
4	1.116	-0.484	1.116	-0.484	34	0.914	-2.07	0.796	0.718
5	1.335	0.052	1.335	0.052	35	1.485	0.764	1.449	-0.58
6	0.683	1.391	0.683	1.391	36	1.203	-1.466	1.293	0.568
7	0.68	-0.888	0.68	-0.888	37	0.586	0.823	0.745	1.106
8	0.818	0.093	0.818	0.093	38	1.464	-0.482	0.895	-0.095
9	1.079	-0.738	1.079	-0.738	39	1.911	-0.884	0.877	1.244
10	0.624	-2.125	0.624	-2.125	40	1.313	0.883	2.237	-0.09
11	1.428	-0.806	1.428	-0.806	41	1.523	0.53	1.558	1.741
12	1.008	-0.196	1.008	-0.196	42	2.143	0.232	0.812	0.956
13	0.943	0.738	0.943	0.738	43	0.55	-0.789	1	1.56
14	0.578	1.126	0.578	1.126	44	0.837	-0.781	0.632	0.311
15	1.873	-0.193	1.873	-0.193	45	1.731	-0.667	2.878	1.272
16	1.589	-0.417	1.589	-0.417	46	1.363	1.064	1.013	-0.223
17	1.035	0.471	1.035	0.471	47	1.295	-1.003	2.497	1.186
18	0.714	-0.148	0.714	-0.148	48	0.877	-0.117	0.845	0.56
19	3.137	0.372	3.137	0.372	49	1.205	-0.11	0.983	0.111
20	1.281	0.223	1.281	0.223	50	4.125	0.343	0.428	0.66
21	1.135	2.745	0.688	0.202	51	1.392	0.27	0.4	0.387
22	0.39	0.906	1.288	-0.35	52	0.505	-0.32	0.612	0.99
23	0.62	1.241	1.067	0.747	53	1.55	0.166	0.933	-0.297
24	0.762	-0.041	1.953	0.476	54	0.555	-0.178	1.146	0.808
25	1.944	0.183	1.773	0.398	55	0.978	1.119	1.351	-0.039
26	1.282	2.087	0.962	-0.207	56	1.334	-0.774	1.416	0.333
27	0.677	-0.571	1.095	2.263	57	0.648	1.807	1.236	-0.448
28	0.621	-1.943	0.666	-0.622	58	0.879	0.071	0.739	0.122
29	1.617	1.539	1.133	1.125	59	1.041	0.685	0.981	2.691
30	1.355	-0.391	0.737	1.663	60	0.578	-1.346	1.396	0.965

1-20 arasındaki maddeler ortak maddelerdir.

Tablo 15. 2PLM 1000 kişi negatif çarpık 80 madde

Maddeler	Form1		Form2		Mad.	Form1		Form2	
	a	b	a	b		a	b	a	b
1	0.748	-1.17	0.748	-1.17	41	1.594	0.68	0.729	1.335
2	0.728	0.358	0.728	0.358	42	0.718	-0.466	1.024	-0.885
3	0.702	-0.194	0.702	-0.194	43	0.735	-0.145	1.181	-0.5
4	2.594	-0.022	2.594	-0.022	44	0.669	-1.317	0.689	-1.994
5	1.567	0.645	1.567	0.645	45	0.647	0.775	0.865	0.594
6	0.547	-0.991	0.547	-0.991	46	0.78	-0.559	1.16	-0.882
7	0.463	0.645	0.463	0.645	47	0.278	0.281	0.808	-0.054
8	0.847	0.637	0.847	0.637	48	1.028	0.82	0.86	0.606
9	0.689	0.757	0.689	0.757	49	1.902	0.055	1.677	-1.213
10	1.431	0.252	1.431	0.252	50	1.29	-0.067	2.129	1.024
11	1.296	0.682	1.296	0.682	51	0.67	0.764	1.154	1.632
12	1.16	1.276	1.16	1.276	52	2.003	0.765	2.302	0.886
13	0.694	-0.576	0.694	-0.576	53	0.715	-0.135	0.435	-0.833
14	0.717	-1.735	0.717	-1.735	54	1.851	-0.963	0.98	0.111
15	0.349	1.69	0.349	1.69	55	1.673	-0.019	1.096	-0.53
16	0.374	-0.627	0.374	-0.627	56	0.882	0.413	0.987	-0.207
17	0.449	-1.745	0.449	-1.745	57	0.499	-0.955	1.093	0.684
18	1.009	-0.448	1.009	-0.448	58	0.86	-2.098	0.762	1.007
19	0.709	2.097	0.709	2.097	59	1.987	0.635	0.802	1.348
20	1.156	1.852	1.156	1.852	60	1.313	1.427	0.67	0.012
21	0.543	0.046	0.543	0.046	61	0.943	1.182	0.552	0.129
22	0.319	0.544	0.319	0.544	62	1.629	-1.197	0.609	1.485
23	0.744	-1.747	0.744	-1.747	63	0.643	-0.425	0.591	-0.042
24	0.809	0.53	0.809	0.53	64	1.511	0.434	1.388	-0.961
25	0.574	-0.389	0.574	-0.389	65	0.275	-1.684	1.335	0.818
26	0.747	-0.739	0.747	-0.739	66	2.864	0.181	1.11	1.147
27	0.706	0.215	0.706	0.215	67	1.669	0.03	1.135	1.28
28	0.609	0.085	0.609	0.085	68	1.356	1.812	1.071	1.627
29	2.237	-0.396	2.237	-0.396	69	0.984	-0.734	0.984	1.543
30	1.107	0.849	1.107	0.849	70	0.444	-0.135	1.157	0.014
31	1.261	0.031	0.569	3.158	71	0.578	-1.388	3.011	0.626
32	0.789	-1.884	0.374	1.512	72	1.603	0.631	0.51	0.494
33	0.984	0.36	0.724	0.058	73	0.94	-0.983	0.358	0.693
34	0.407	-0.98	1.908	0.319	74	1.263	-0.083	1.287	0.816
35	0.974	0.342	1.566	1.579	75	1.001	-1.394	0.889	0.536
36	0.881	-0.962	0.903	0.525	76	1.53	-1.128	1.28	0.983
37	2.313	-1.031	1.528	0.255	77	1.138	1.454	0.985	0.832
38	1.475	0.178	1.482	0.163	78	0.595	0.659	1.364	0.04
39	2.342	1.724	1.356	0.649	79	0.791	-1.362	1.287	0.051
40	0.938	0.592	1.475	0.429	80	1.551	1.624	1.895	-0.391

1-30 arasındaki maddeler ortak maddelerdir.

Tablo 16. 2PLM 3000 kişi negatif çarpık 30 madde

Maddeler	Form1		Form2		Mad.	Form1		Form2	
	a	b	a	b		a	b	a	b
1	1.741	0.499	1.741	0.499	16	1.186	0.209	1.101	0.026
2	0.76	1.945	0.76	1.945	17	2.314	1.327	0.659	0.633
3	0.579	-0.371	0.579	-0.371	18	1.219	-0.572	0.742	0.888
4	0.935	-0.811	0.935	-0.811	19	0.801	-1.414	1.527	-0.279
5	1.596	0.417	1.596	0.417	20	1.043	1.042	0.922	-0.635
6	1.216	0.306	1.216	0.306	21	0.795	-1.495	1.276	2.6
7	0.691	1.108	0.691	1.108	22	0.952	0.115	0.489	-1.265
8	1.472	1.149	1.472	1.149	23	1.461	2.307	0.634	0.929
9	1.216	-0.627	1.216	-0.627	24	0.852	1.784	1.475	0.586
10	1.663	0.957	1.663	0.957	25	0.718	-0.733	0.713	-0.951
11	1.085	-0.134	0.809	2.071	26	1.391	0.307	1.331	-0.507
12	1.099	1.017	0.757	0.108	27	0.632	-0.329	0.656	0.576
13	0.474	-0.832	0.909	-0.699	28	0.789	-1.222	0.899	0.4
14	0.901	-1.12	0.665	-1.576	29	0.708	2.003	1.027	1.698
15	0.832	-0.994	2.365	1.044	30	1.322	-0.411	0.87	0.451

1-10 arasındaki maddeler ortak maddelerdir.

Tablo 17. 2PLM 3000 kişi negatif çarpık 60 madde

	Form1		Form2			Form1		Form2	
	a	b	a	b		a	b	a	b
1	0.976	0.159	0.976	0.159	31	0.487	0.81	0.712	1.312
2	1.087	-0.754	1.087	-0.754	32	0.935	-0.454	1.043	-0.461
3	1.077	-0.848	1.077	-0.848	33	1.624	0.238	1.072	-0.533
4	0.701	-0.157	0.701	-0.157	34	1.179	-0.113	0.855	0.057
5	1.183	1.118	1.183	1.118	35	0.937	-1.226	0.808	-0.145
6	2.269	-0.345	2.269	-0.345	36	1.853	0.037	1.007	-0.042
7	1.491	0.335	1.491	0.335	37	0.893	1.03	0.956	0.767
8	0.584	0.253	0.584	0.253	38	0.914	-1.115	1.103	-0.174
9	1.19	0.236	1.19	0.236	39	1.843	-1.245	0.937	0.734
10	0.987	1.564	0.987	1.564	40	1.455	1.084	1.313	1.241
11	1.673	0.842	1.673	0.842	41	0.436	0.112	1.347	1.709
12	0.557	0.195	0.557	0.195	42	1.475	1.955	0.582	1.622
13	0.919	-0.016	0.919	-0.016	43	1.123	-1.205	0.991	0.816
14	1.398	0.735	1.398	0.735	44	1.837	2.355	1.21	0.184
15	1.291	-0.668	1.291	-0.668	45	3.599	0.608	0.608	0.583
16	1.23	0.515	1.23	0.515	46	1.605	0.382	0.688	0.838
17	1.129	-0.897	1.129	-0.897	47	0.714	-0.184	1.634	2.091
18	0.961	0.4	0.961	0.4	48	0.378	1.077	0.388	0.274
19	0.625	-0.096	0.625	-0.096	49	0.741	1.364	0.671	-0.192
20	0.407	0.433	0.407	0.433	50	1.557	1.957	1.224	-0.715
21	0.694	1.51	0.85	-0.163	51	1.263	-0.996	0.662	0.885
22	0.705	-0.636	0.591	0.468	52	0.793	1.042	1.512	-0.61
23	0.856	-1.096	1.157	-0.058	53	1.342	0.336	0.816	1.892
24	0.544	-1.821	1.335	0.456	54	0.663	0.236	1.591	-1.269
25	1.093	-0.392	0.857	0.661	55	0.733	0.096	0.563	-0.783
26	1.257	-0.205	0.766	0.914	56	0.922	0.402	1.049	1.005
27	1.114	0.152	0.691	0.787	57	0.74	-0.466	0.52	-0.551
28	1.215	0.775	0.928	0.278	58	1.239	0.293	1.925	1.85
29	0.842	0.051	1.214	0.917	59	0.566	0.594	2.598	0.445
30	0.78	-1.09	1.55	0.693	60	0.748	-0.85	0.831	-0.146

1-20 arasındaki maddeler ortak maddelerdir.

Tablo 18. 2PLM 3000 kişi negatif çarpık 80 madde

Maddeler	Form1		Form2		Mad.	Form1		Form2	
	a	b	a	b		a	b	a	b
1	0.875	0.365	0.875	0.365	41	0.979	0.697	0.629	2.239
2	1.284	1.803	1.284	1.803	42	0.956	-0.443	1.489	3.073
3	1.536	-1.238	1.536	-1.238	43	1.389	0.828	1.07	0.678
4	0.836	1.532	0.836	1.532	44	0.742	0.012	1.361	0.425
5	0.816	-0.705	0.816	-0.705	45	1.179	-1.181	0.773	1.697
6	0.345	0.216	0.345	0.216	46	1.076	-1.913	0.647	0.814
7	0.345	-1.384	0.345	-1.384	47	0.604	-1.276	1.028	1.641
8	1.753	0.709	1.753	0.709	48	0.6	0.485	1.658	-0.259
9	2.114	-0.834	2.114	-0.834	49	0.881	-0.453	2.158	0.335
10	0.642	0.055	0.642	0.055	50	1.513	0.991	1.216	1.613
11	0.501	0.017	0.501	0.017	51	3.61	3.118	0.855	-0.194
12	1.725	0.249	1.725	0.249	52	1.795	0.31	0.878	0.574
13	1.509	0.72	1.509	0.72	53	1.176	-0.418	3.871	0.374
14	0.726	-0.294	0.726	-0.294	54	1.907	-0.363	1.436	0.102
15	1.34	2.267	1.34	2.267	55	1.153	-0.33	0.884	-0.147
16	0.549	-0.847	0.549	-0.847	56	0.638	0.699	0.633	1.481
17	0.359	-1.084	0.359	-1.084	57	0.883	0.422	0.748	1.185
18	1.453	1.76	1.453	1.76	58	1.05	-0.638	0.492	0.574
19	0.804	0.138	0.804	0.138	59	1.003	0.787	0.391	0.728
20	0.867	-0.84	0.867	-0.84	60	0.627	-0.388	1.09	0.986
21	0.769	-1.179	0.769	-1.179	61	0.844	-0.64	1.037	1.306
22	1.628	0.247	1.628	0.247	62	1.402	-0.647	1.745	0.492
23	1.839	-0.992	1.839	-0.992	63	1.142	-1.128	0.333	-0.761
24	0.931	-0.082	0.931	-0.082	64	0.568	0.286	0.856	2.044
25	1.561	-1.602	1.561	-1.602	65	0.784	-2.294	4.995	1.888
26	0.895	0.835	0.895	0.835	66	0.742	0.604	1.139	0.295
27	2.166	-0.533	2.166	-0.533	67	0.545	0.892	0.872	-0.21
28	1.468	0.078	1.468	0.078	68	0.974	0.876	0.771	1.283
29	1.157	-0.414	1.157	-0.414	69	1.355	0.896	0.731	0.916
30	0.508	-1.726	0.508	-1.726	70	1.133	0.412	0.728	0.769
31	0.749	-0.19	2.226	0.654	71	2.031	1.404	0.777	-0.08
32	0.784	1.648	0.846	0.918	72	0.648	-0.281	2.503	-0.176
33	4.628	0.223	2.018	-0.301	73	1.404	-0.98	1.061	0.632
34	0.572	0.299	1.794	0.297	74	0.873	0.445	1.802	-0.529
35	1.798	-0.325	2.027	0.37	75	0.877	0.335	1.805	1.019
36	0.911	0.699	2.258	0.061	76	1.654	-0.315	0.608	1.037
37	1.587	0.522	1.315	-0.001	77	0.919	-0.146	1.032	1.407
38	0.623	-0.311	2.086	-0.279	78	0.472	-0.084	1.341	0.023
39	1.001	0.524	1.693	1.935	79	0.795	0.813	1.858	-0.106
40	1.109	-0.997	4.037	0.608	80	1.149	-0.94	1.654	0.196

1-30 arasındaki maddeler ortak maddelerdir.

Tablo 19. 3PLM 1000 kişi normal dağılım 30 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	0.734	-0.946	0.128	0.734	-0.946	0.128	16	0.889	0.771	0.211	1.133	1.434	0.147
2	2.508	0.721	0.198	2.508	0.721	0.198	17	0.564	0.583	0.217	0.723	0.866	0.414
3	1.198	0.192	0.128	1.198	0.192	0.128	18	1.989	0.151	0.255	0.739	-0.556	0.164
4	0.492	-1.118	0.16	0.492	-1.118	0.16	19	0.915	-0.081	0.142	2.344	3.396	0.174
5	0.885	2.07	0.142	0.885	2.07	0.142	20	1.093	-1.18	0.234	1.337	0.86	0.226
6	1.315	-0.248	0.236	1.315	-0.248	0.236	21	0.92	0.189	0.289	0.469	-0.923	0.174
7	0.536	0.164	0.102	0.536	0.164	0.102	22	0.912	0.195	0.417	2.848	0.237	0.218
8	1.009	0.363	0.14	1.009	0.363	0.14	23	0.45	-0.613	0.196	1.04	-1.159	0.059
9	3.362	0.752	0.208	3.362	0.752	0.208	24	0.867	0.926	0.207	0.754	-1.566	0.106
10	0.793	0.144	0.107	0.793	0.144	0.107	25	2.825	-0.376	0.185	1.069	0.183	0.108
11	1.115	0.044	0.196	0.572	1.076	0.259	26	1.895	0.027	0.211	0.655	2.187	0.086
12	1.658	-0.074	0.22	2.37	2.327	0.185	27	0.703	-0.307	0.173	1.048	-0.836	0.188
13	0.415	0.962	0.336	0.498	1.071	0.252	28	1.342	1.84	0.146	0.849	2.147	0.113
14	1.662	-0.929	0.198	1.345	-0.734	0.186	29	0.583	-0.448	0.157	0.951	1.4	0.268
15	0.545	0.045	0.273	0.352	-1.157	0.156	30	0.432	-2.46	0.193	3.573	0.845	0.207

1-10 arasındaki maddeler ortak maddelerdir.

Tablo 20. 3PLM 1000 kişi normal dağılım 60 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	0.906	0.814	0.21	0.906	0.814	0.21	31	1.16	-0.168	0.289	1.463	-0.142	0.211
2	0.975	0.086	0.133	0.975	0.086	0.133	32	2.812	1.328	0.129	0.463	1.369	0.168
3	1.24	-0.858	0.098	1.24	-0.858	0.098	33	1.197	-0.031	0.141	1.276	-1.588	0.149
4	1.605	1.735	0.272	1.605	1.735	0.272	34	0.395	0.335	0.221	0.807	-0.098	0.259
5	0.926	-0.594	0.22	0.926	-0.594	0.22	35	1.685	1.347	0.174	1.6	1.043	0.12
6	1.077	0.367	0.151	1.077	0.367	0.151	36	3.202	0.396	0.21	0.746	0.748	0.238
7	0.993	-0.897	0.158	0.993	-0.897	0.158	37	0.988	0.618	0.198	0.422	1.036	0.305
8	1.717	-0.66	0.214	1.717	-0.66	0.214	38	1.116	0.315	0.228	1.947	0.755	0.153
9	0.433	-0.017	0.178	0.433	-0.017	0.178	39	3.468	-0.019	0.253	1.483	0.363	0.105
10	1.446	-0.271	0.153	1.446	-0.271	0.153	40	0.533	-0.724	0.245	1.29	0.118	0.238
11	0.305	-0.62	0.159	0.305	-0.62	0.159	41	1.232	0.589	0.166	2.439	-0.259	0.32
12	0.905	0.803	0.275	0.905	0.803	0.275	42	0.735	1.87	0.288	1.306	-0.295	0.217
13	0.649	0.009	0.206	0.649	0.009	0.206	43	0.536	0.433	0.188	1.563	0.334	0.386
14	1.724	0.034	0.232	1.724	0.034	0.232	44	2.062	-0.389	0.153	1.494	-0.385	0.127
15	0.986	-0.85	0.115	0.986	-0.85	0.115	45	0.734	1.536	0.251	1.67	0.291	0.087
16	1.034	0.726	0.157	1.034	0.726	0.157	46	1.208	-0.136	0.134	1.219	-0.965	0.157
17	0.369	1.379	0.174	0.369	1.379	0.174	47	1.602	-0.065	0.193	2.507	0.998	0.173
18	2.095	-0.967	0.128	2.095	-0.967	0.128	48	1.268	0.438	0.102	1.584	0.695	0.172
19	1.079	1.216	0.281	1.079	1.216	0.281	49	0.733	-0.85	0.159	0.609	0.363	0.241
20	1.918	1.022	0.187	1.918	1.022	0.187	50	2.219	-0.409	0.376	1.623	2.509	0.14
21	0.879	0.783	0.298	0.462	0.611	0.299	51	1.302	0.79	0.262	2.065	1.766	0.162
22	0.95	0.75	0.098	0.851	-1.235	0.239	52	0.616	0.205	0.261	0.973	1.202	0.158
23	0.387	1.195	0.178	2.005	0.844	0.34	53	0.478	1.217	0.206	1.067	0.682	0.216
24	0.433	-2.014	0.184	0.542	-0.909	0.241	54	0.744	0.408	0.159	2.292	0.907	0.153
25	1.062	1.324	0.213	0.877	0.489	0.296	55	0.668	1.469	0.155	1.832	1.154	0.154
26	1.022	-1.469	0.195	0.424	0.746	0.125	56	1.238	2.123	0.206	1.243	-0.056	0.144
27	1.152	0.273	0.137	0.855	-0.404	0.175	57	1.292	3.144	0.22	0.993	1.729	0.223
28	0.489	-0.276	0.351	1.298	-2.185	0.215	58	1.355	0.43	0.223	1.612	0.314	0.182
29	1.018	0.489	0.27	0.302	-0.051	0.172	59	0.543	0.427	0.239	1.361	1.523	0.107
30	1.472	0.057	0.29	0.796	0.238	0.163	60	1.014	-0.896	0.19	0.634	0.73	0.18

1-20 arasındaki maddeler ortak maddelerdir.

Tablo 21. 3PLM 1000 kişi normal dağılım 80 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	1.134	-0.037	0.183	1.134	-0.037	0.183	41	0.931	-1.453	0.226	0.813	0.123	0.257
2	0.759	1.28	0.136	0.759	1.28	0.136	42	0.506	0.379	0.259	2.452	0.38	0.21
3	1.73	-0.084	0.289	1.73	-0.084	0.289	43	0.87	-0.747	0.202	1.74	0.591	0.219
4	0.605	-1.261	0.273	0.605	-1.261	0.273	44	1.812	1.412	0.166	2.019	0.601	0.273
5	0.829	0.457	0.211	0.829	0.457	0.211	45	1.398	-0.654	0.194	1.776	2.276	0.158
6	1.057	-0.253	0.12	1.057	-0.253	0.12	46	0.357	-1.468	0.159	1.254	0.81	0.232
7	1.345	-2.086	0.194	1.345	-2.086	0.194	47	0.825	1.838	0.18	0.905	1.453	0.164
8	0.888	1.251	0.165	0.888	1.251	0.165	48	0.496	-0.942	0.179	1.034	1.676	0.157
9	0.372	-0.792	0.178	0.372	-0.792	0.178	49	0.782	0.491	0.256	1.624	1.131	0.207
10	0.348	0.564	0.13	0.348	0.564	0.13	50	0.693	-0.704	0.246	1.527	0.216	0.088
11	0.657	1.275	0.281	0.657	1.275	0.281	51	1.237	0.682	0.174	0.583	-2.386	0.289
12	1.272	-0.617	0.175	1.272	-0.617	0.175	52	1.153	-0.789	0.234	1.355	2.644	0.105
13	1.125	0.324	0.142	1.125	0.324	0.142	53	0.995	0.122	0.289	0.783	-0.317	0.141
14	0.827	-0.03	0.231	0.827	-0.03	0.231	54	0.802	-0.17	0.249	0.73	0.7	0.119
15	0.708	-1.041	0.176	0.708	-1.041	0.176	55	0.564	-0.699	0.285	0.905	0.264	0.222
16	0.397	-0.798	0.114	0.397	-0.798	0.114	56	0.853	0.186	0.176	0.343	0.81	0.202
17	0.697	0.925	0.226	0.697	0.925	0.226	57	2.06	0.18	0.195	0.643	1.098	0.134
18	1.668	-0.366	0.262	1.668	-0.366	0.262	58	1.773	-1.017	0.137	1.204	1.228	0.243
19	0.945	-0.822	0.084	0.945	-0.822	0.084	59	0.81	-0.112	0.296	1.291	0.369	0.142
20	0.436	-0.102	0.308	0.436	-0.102	0.308	60	1.072	-0.537	0.132	2.838	1.852	0.166
21	1.404	-0.468	0.181	1.404	-0.468	0.181	61	1.173	0.423	0.188	0.537	-0.312	0.162
22	1.1	0.7	0.196	1.1	0.7	0.196	62	0.704	-1.315	0.144	1.387	-0.087	0.173
23	1.594	0.187	0.302	1.594	0.187	0.302	63	0.794	2.756	0.153	0.452	1.243	0.213
24	2.271	1.11	0.265	2.271	1.11	0.265	64	1.432	-0.691	0.095	1.107	1.52	0.245
25	1.049	-0.71	0.283	1.049	-0.71	0.283	65	0.628	-1.455	0.15	1.396	-0.466	0.102
26	0.917	-0.576	0.331	0.917	-0.576	0.331	66	1.592	0.398	0.165	1.656	1.55	0.234
27	0.925	-1.006	0.213	0.925	-1.006	0.213	67	1.822	0.503	0.211	0.886	1.633	0.16
28	1.881	-0.21	0.206	1.881	-0.21	0.206	68	1.109	1.146	0.191	1.436	-1.034	0.182
29	1.084	0.461	0.119	1.084	0.461	0.119	69	0.575	-1.56	0.12	0.613	0.658	0.164
30	1.105	-0.637	0.216	1.105	-0.637	0.216	70	0.865	-1.198	0.301	0.718	1.164	0.246
31	1.752	-1.411	0.176	1.529	0.898	0.197	71	2.037	1.477	0.144	1.128	0.877	0.349
32	0.646	0.704	0.291	0.537	-0.729	0.268	72	1.017	-0.456	0.156	0.807	-0.216	0.258
33	0.264	0.629	0.189	0.835	0.987	0.255	73	2.281	-0.469	0.218	1.146	-0.172	0.212
34	1.075	0.124	0.235	1.875	1.004	0.307	74	1.032	1.989	0.141	1.385	-0.311	0.17
35	2.5	-0.385	0.163	1.318	0.843	0.161	75	1.143	1.797	0.133	1.339	0.403	0.181
36	0.365	-0.059	0.233	0.655	2.217	0.229	76	1.053	-0.979	0.133	0.855	-0.943	0.233
37	1.062	-0.912	0.281	1.731	0.986	0.201	77	1.341	0.597	0.096	0.648	1.702	0.119
38	0.649	-0.422	0.128	0.649	1.454	0.232	78	0.81	1.04	0.224	1.851	1.305	0.246
39	0.99	0.923	0.155	1.527	-0.601	0.248	79	1.06	-0.504	0.278	0.888	0.579	0.245
40	1.326	0.436	0.264	0.884	0.348	0.285	80	1.982	-0.333	0.156	1.063	0.587	0.134

1-30 arasındaki maddeler ortak maddelerdir.

Tablo 22. 3PLM 3000 kişi normal dağılım 30 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	0.532	2.459	0.18	0.532	2.459	0.18	16	1.874	0.681	0.205	0.832	-0.731	0.21
2	0.615	1.098	0.242	0.615	1.098	0.242	17	1.363	-0.576	0.155	0.98	2.149	0.231
3	0.734	0.467	0.31	0.734	0.467	0.31	18	0.489	0.614	0.147	1.04	1.567	0.164
4	2.35	-0.145	0.283	2.35	-0.145	0.283	19	2.191	-1.065	0.199	1.084	0.35	0.222
5	1.043	-0.388	0.274	1.043	-0.388	0.274	20	0.945	0.064	0.096	0.657	0.103	0.168
6	0.44	0.757	0.203	0.44	0.757	0.203	21	2.339	0.523	0.23	0.967	1.197	0.128
7	1.681	-0.093	0.213	1.681	-0.093	0.213	22	0.789	0.039	0.215	0.977	1.319	0.296
8	0.285	-1.831	0.327	0.285	-1.831	0.327	23	0.577	1.302	0.321	1.344	0.059	0.109
9	1.462	-0.071	0.239	1.462	-0.071	0.239	24	1.738	-0.374	0.125	0.94	-1.119	0.186
10	0.957	0.964	0.293	0.957	0.964	0.293	25	1.305	0.311	0.105	1.462	0.801	0.151
11	1.163	0.959	0.191	0.906	-1.897	0.205	26	1.094	0.879	0.392	0.899	1.703	0.262
12	0.849	0.844	0.153	1.409	-0.066	0.17	27	0.921	0.334	0.258	1.23	-0.194	0.12
13	1.09	-1.113	0.199	0.87	0.904	0.171	28	1.09	0.351	0.147	1.777	-0.959	0.132
14	0.861	-0.033	0.151	0.749	1.302	0.174	29	0.984	-1.03	0.166	0.945	0.037	0.157
15	0.468	-0.009	0.18	0.577	2.212	0.153	30	1.077	-0.972	0.237	2.362	0.889	0.103

1-10 arasındaki maddeler ortak maddelerdir.

Tablo 23. 3PLM 3000 Kişi normal dağılım 60 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	1.096	0.116	0.182	1.096	0.116	0.182	31	0.61	0.113	0.217	0.546	0.489	0.239
2	0.985	1.026	0.11	0.985	1.026	0.11	32	2.125	0.795	0.146	2.008	0.077	0.177
3	1.63	-0.054	0.163	1.63	-0.054	0.163	33	0.923	1.242	0.222	1.692	-1.118	0.258
4	0.586	1.394	0.277	0.586	1.394	0.277	34	1.602	0.718	0.207	0.789	0.895	0.113
5	0.799	-1.48	0.166	0.799	-1.48	0.166	35	1.06	-0.857	0.143	0.549	1.909	0.181
6	1.433	-0.248	0.144	1.433	-0.248	0.144	36	0.674	-0.564	0.219	0.701	0.95	0.224
7	1.954	-0.48	0.119	1.954	-0.48	0.119	37	1.228	0.065	0.324	2.522	1.256	0.218
8	1.934	0.076	0.263	1.934	0.076	0.263	38	0.944	-1.756	0.324	2.134	0.159	0.219
9	1.126	0.836	0.243	1.126	0.836	0.243	39	1.343	-0.4	0.192	0.97	1.136	0.216
10	1.316	-1.007	0.156	1.316	-1.007	0.156	40	1.134	-0.332	0.172	2.175	1.083	0.168
11	0.814	1.503	0.167	0.814	1.503	0.167	41	0.853	-0.692	0.2	0.525	1.532	0.165
12	2.192	-1.761	0.141	2.192	-1.761	0.141	42	1.413	0.844	0.145	0.682	0.391	0.258
13	0.435	-0.894	0.121	0.435	-0.894	0.121	43	3.453	-0.52	0.147	0.492	1.076	0.148
14	0.854	0.332	0.32	0.854	0.332	0.32	44	0.933	-1.58	0.234	1.093	0.241	0.229
15	1.212	0.486	0.127	1.212	0.486	0.127	45	1.099	1.475	0.16	0.647	1.597	0.270
16	0.538	-0.796	0.277	0.538	-0.796	0.277	46	0.596	-0.84	0.188	0.753	1.102	0.259
17	0.647	0.183	0.243	0.647	0.183	0.243	47	0.492	-0.234	0.162	1.288	0.568	0.221
18	1.026	1.152	0.147	1.026	1.152	0.147	48	1.789	0.494	0.229	1.16	-0.564	0.119
19	0.506	-0.219	0.132	0.506	-0.219	0.132	49	0.909	-0.551	0.167	0.413	-1.413	0.316
20	0.606	-0.583	0.177	0.606	-0.583	0.177	50	0.658	1.049	0.225	0.794	-0.152	0.307
21	0.674	-0.227	0.343	0.366	0.377	0.226	51	2.173	-1.455	0.189	1.125	0.949	0.127
22	0.782	-0.477	0.196	0.566	0.813	0.33	52	1.152	-0.313	0.29	0.887	0.744	0.26
23	0.373	1.359	0.159	1.055	1.376	0.221	53	1.523	0.342	0.233	0.866	-0.087	0.218
24	1.076	-1.038	0.165	0.889	-0.625	0.18	54	0.959	0.123	0.119	0.729	-1.897	0.09
25	1.966	0.391	0.164	1.708	1.32	0.211	55	1.339	0.24	0.126	1.244	1.961	0.153
26	0.779	0.025	0.281	1.498	2.114	0.206	56	1.283	0.059	0.232	0.47	0.816	0.168
27	2.007	0.118	0.181	2.059	-0.769	0.2	57	1.158	0.633	0.156	0.97	0.85	0.127
28	0.771	1.584	0.236	0.202	0.001	0.126	58	0.466	-0.221	0.112	0.706	0.361	0.159
29	0.91	-0.324	0.211	1.174	0.166	0.271	59	1.409	0.723	0.152	1.419	0.763	0.141
30	0.724	1.528	0.182	1.236	0.803	0.125	60	0.484	-0.752	0.265	0.784	2.008	0.247

1-20 arasındaki maddeler ortak maddelerdir.

Tablo 24. 3PLM 3000 kişi normal dağılım 80 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	1.901	-0.848	0.111	1.901	-0.848	0.111	41	0.219	0.57	0.132	0.823	-0.553	0.125
2	0.795	-0.109	0.122	0.795	-0.109	0.122	42	0.775	-0.288	0.253	0.685	-0.361	0.211
3	1.612	-0.679	0.3	1.612	-0.679	0.3	43	0.979	0.07	0.235	1.119	1.453	0.28
4	0.883	0.495	0.235	0.883	0.495	0.235	44	0.646	-0.526	0.19	0.943	0.372	0.201
5	1	1.899	0.2	1	1.899	0.2	45	1.236	-0.381	0.213	1.477	-0.114	0.124
6	0.821	0.265	0.31	0.821	0.265	0.31	46	0.701	0.444	0.118	2.037	-0.125	0.248
7	1.438	-0.076	0.315	1.438	-0.076	0.315	47	0.979	-1.859	0.158	1.187	0.658	0.167
8	1.388	0.929	0.222	1.388	0.929	0.222	48	0.748	-0.466	0.219	0.671	0.046	0.219
9	0.764	0.127	0.167	0.764	0.127	0.167	49	0.505	0.364	0.092	1.271	1.257	0.212
10	2.027	1.143	0.149	2.027	1.143	0.149	50	2.038	0.038	0.078	1.254	-0.02	0.234
11	2.385	0.48	0.18	2.385	0.48	0.18	51	0.814	0.149	0.144	1.083	-1.229	0.146
12	0.524	0.427	0.14	0.524	0.427	0.14	52	1.035	-0.074	0.126	0.734	-0.952	0.256
13	1.047	-0.101	0.279	1.047	-0.101	0.279	53	1.231	1.348	0.236	0.376	1.504	0.121
14	2.034	0.574	0.137	2.034	0.574	0.137	54	1.998	-0.269	0.206	0.952	1.52	0.174
15	0.37	0.198	0.22	0.37	0.198	0.22	55	1.302	0.979	0.314	0.929	1.002	0.223
16	1.17	-0.269	0.225	1.17	-0.269	0.225	56	1.631	-0.861	0.175	1.321	1.012	0.18
17	1.06	-0.924	0.201	1.06	-0.924	0.201	57	0.795	0.053	0.164	1.062	1.917	0.121
18	1.113	-2.291	0.234	1.113	-2.291	0.234	58	0.828	0.718	0.291	0.49	-0.712	0.095
19	1.397	-1.021	0.2	1.397	-1.021	0.2	59	0.682	-0.567	0.068	2.617	-1.217	0.217
20	0.909	-0.667	0.133	0.909	-0.667	0.133	60	1.08	-0.346	0.202	0.488	-1.466	0.074
21	1.744	0.908	0.27	1.744	0.908	0.27	61	2.581	-1.644	0.257	2.07	2.157	0.208
22	0.704	-0.714	0.272	0.704	-0.714	0.272	62	2.241	-0.468	0.18	1.144	-0.011	0.193
23	0.443	0.784	0.229	0.443	0.784	0.229	63	1.216	0.263	0.227	3.029	1.171	0.349
24	1.502	0.114	0.182	1.502	0.114	0.182	64	1.107	-0.148	0.235	0.923	1.283	0.166
25	0.868	0.325	0.149	0.868	0.325	0.149	65	0.644	-1.448	0.131	0.451	0.17	0.169
26	0.525	0.995	0.13	0.525	0.995	0.13	66	1.719	1.054	0.282	0.524	-0.59	0.093
27	1.732	-0.892	0.155	1.732	-0.892	0.155	67	1.147	-1.362	0.195	0.899	-1.823	0.173
28	1.494	0.261	0.224	1.494	0.261	0.224	68	0.726	-0.492	0.102	2.024	-0.104	0.223
29	1.962	0.95	0.176	1.962	0.95	0.176	69	0.698	1.276	0.177	0.505	0.832	0.181
30	0.824	-0.229	0.194	0.824	-0.229	0.194	70	0.916	2.151	0.25	0.727	0.06	0.159
31	2.325	0.824	0.237	1.102	-0.701	0.355	71	2.09	-0.03	0.172	2.181	-0.35	0.161
32	0.837	1.918	0.16	1.469	0.444	0.161	72	0.8	-0.016	0.297	1.579	-1.95	0.197
33	4.311	-0.643	0.205	0.813	-0.376	0.269	73	1.159	1.124	0.189	2.749	-0.658	0.186
34	0.726	-0.997	0.216	0.894	0.062	0.262	74	0.996	-0.018	0.151	1.026	-1.13	0.241
35	1.78	0.74	0.26	1.119	-0.517	0.093	75	0.722	-0.17	0.194	0.3	0.013	0.107
36	2.657	-2.349	0.371	0.605	0.972	0.191	76	0.898	0.138	0.166	1.252	0.755	0.332
37	0.577	1.926	0.272	1.564	1.809	0.183	77	0.95	1.543	0.306	2.356	0.505	0.185
38	0.768	0.068	0.178	2.061	-1.077	0.354	78	1.467	-1.185	0.189	0.341	1.619	0.299
39	0.269	-1.083	0.137	0.841	1.547	0.294	79	0.632	-2.375	0.174	1.516	0.07	0.156
40	1.02	0.568	0.199	0.565	1.859	0.129	80	1.168	-0.796	0.153	2.521	1.282	0.273

1-30 arasındaki maddeler ortak maddelerdir.

Tablo 25. 2PLM 1000 kişi pozitif çarpık 30 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	1.94	-0.552	0.292	1.94	-0.552	0.292	16	1.238	0.947	0.13	0.918	-1.65	0.268
2	0.519	-1.424	0.144	0.519	-1.424	0.144	17	1.307	0.262	0.163	0.438	0.365	0.206
3	0.489	-0.971	0.19	0.489	-0.971	0.19	18	0.319	-1.899	0.079	0.399	-0.035	0.118
4	2.133	-0.161	0.398	2.133	-0.161	0.398	19	1.371	-0.175	0.141	0.52	-0.258	0.203
5	1.349	-2.79	0.157	1.349	-2.79	0.157	20	1.368	1.14	0.127	1.974	0.544	0.173
6	0.631	-0.136	0.236	0.631	-0.136	0.236	21	2.742	-0.186	0.073	1.001	0.98	0.151
7	0.694	-1.385	0.204	0.694	-1.385	0.204	22	0.892	-0.414	0.206	1.356	2.329	0.12
8	1.132	0.213	0.356	1.132	0.213	0.356	23	0.521	0.03	0.158	1.9	0.178	0.171
9	1.129	2.106	0.286	1.129	2.106	0.286	24	0.958	0.691	0.213	0.521	0.422	0.285
10	0.949	-1.292	0.148	0.949	-1.292	0.148	25	1.837	-0.738	0.248	2.857	-0.379	0.181
11	0.949	0.618	0.179	0.779	-1.082	0.273	26	1.087	0.219	0.238	0.571	-1.371	0.184
12	1.53	0.97	0.152	0.844	2.652	0.283	27	0.289	-0.182	0.225	1.761	2.861	0.115
13	1.276	0.615	0.079	1.136	-0.273	0.178	28	1.437	1.991	0.347	1.667	-0.97	0.217
14	0.428	-0.896	0.198	0.919	1.767	0.359	29	0.693	0.332	0.311	0.788	-0.644	0.156
15	0.514	-0.947	0.201	0.976	-0.041	0.16	30	0.499	0.34	0.209	1.194	0.736	0.109

1-10 arasındaki maddeler ortak maddelerdir.

Tablo 26. 3PLM 1000 kişi pozitif çarpık 60 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	0.343	-0.422	0.145	0.343	-0.422	0.145	31	0.746	-0.399	0.187	0.6	-1.42	0.294
2	1.194	-0.081	0.155	1.194	-0.081	0.155	32	0.566	-0.3	0.289	1.485	0.924	0.171
3	1.348	-1.035	0.241	1.348	-1.035	0.241	33	1.52	-1.18	0.268	2.469	1.274	0.163
4	0.826	1.405	0.253	0.826	1.405	0.253	34	0.756	0.715	0.206	1.873	0.599	0.175
5	0.905	0.595	0.203	0.905	0.595	0.203	35	0.753	0.6	0.191	0.899	-1.158	0.328
6	0.916	1.864	0.11	0.916	1.864	0.11	36	0.602	-1.461	0.29	0.783	1.117	0.244
7	1.04	0.402	0.171	1.04	0.402	0.171	37	2.661	0.587	0.196	0.645	0.676	0.196
8	0.85	0.296	0.167	0.85	0.296	0.167	38	1.907	0.387	0.143	2.358	-1.172	0.167
9	1.014	-0.72	0.268	1.014	-0.72	0.268	39	0.807	-1.765	0.204	1.315	0.929	0.169
10	1.809	1.033	0.11	1.809	1.033	0.11	40	1.949	1.709	0.206	0.624	0.266	0.124
11	1.028	0.641	0.192	1.028	0.641	0.192	41	1.135	-0.086	0.257	1	-0.352	0.178
12	1.122	1.476	0.162	1.122	1.476	0.162	42	0.686	0.424	0.184	1.599	0.188	0.143
13	0.924	0.399	0.124	0.924	0.399	0.124	43	0.583	0.793	0.142	2.393	1.392	0.314
14	2.006	0.726	0.226	2.006	0.726	0.226	44	1.255	-0.156	0.217	0.863	1.694	0.117
15	0.896	1.046	0.2	0.896	1.046	0.2	45	0.457	0.898	0.157	0.203	0.967	0.224
16	0.748	0.214	0.238	0.748	0.214	0.238	46	2.424	1.967	0.141	0.949	-0.263	0.176
17	0.906	-0.304	0.268	0.906	-0.304	0.268	47	1.073	-1.009	0.139	0.757	0.424	0.131
18	0.931	1.082	0.122	0.931	1.082	0.122	48	1.203	-0.3	0.196	1.587	0.099	0.205
19	0.481	-0.269	0.218	0.481	-0.269	0.218	49	0.718	-1.366	0.234	1.036	-0.214	0.112
20	1.224	1.014	0.187	1.224	1.014	0.187	50	0.569	1.178	0.228	1.085	0.858	0.191
21	0.775	-1.484	0.26	0.779	-0.019	0.155	51	0.91	1.884	0.36	0.885	0.485	0.167
22	0.563	-0.258	0.182	0.653	1.501	0.116	52	0.705	-0.6	0.235	3.323	1.694	0.154
23	2.351	-0.13	0.113	0.639	0.658	0.26	53	1.931	0.302	0.223	0.882	1.434	0.138
24	1.144	0.003	0.206	0.785	-0.074	0.131	54	1.018	1.53	0.2	0.468	1.016	0.099
25	0.918	-1.109	0.317	0.494	2.728	0.163	55	0.663	1.392	0.216	1.141	-0.051	0.183
26	0.675	-0.923	0.129	0.636	2.676	0.258	56	0.749	-1.627	0.178	0.847	-1.903	0.131
27	0.578	0.305	0.111	1.298	1.004	0.179	57	0.997	-0.207	0.344	0.963	-0.389	0.218
28	1.859	-0.963	0.28	1.546	-0.364	0.202	58	0.672	0.335	0.293	1.138	-0.455	0.208
29	0.926	1.481	0.221	1.165	0.126	0.166	59	1.421	-1.509	0.327	0.527	0.989	0.3
30	0.878	-1.222	0.198	0.468	0.98	0.182	60	1.649	-1.048	0.223	0.368	-0.266	0.15

1-20 arasındaki maddeler ortak maddelerdir.

Tablo 27. 3PLM 1000 kişi pozitif çarpık 80 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	0.405	-0.26	0.208	0.405	-0.26	0.208	41	1.276	-0.271	0.229	0.641	0.128	0.221
2	0.952	0.408	0.229	0.952	0.408	0.229	42	0.87	-0.578	0.185	0.936	1.286	0.121
3	0.941	-2.587	0.181	0.941	-2.587	0.181	43	2.058	-1.582	0.322	0.607	1.341	0.179
4	0.551	2.048	0.179	0.551	2.048	0.179	44	0.707	1.719	0.269	0.928	1.45	0.202
5	0.685	0.691	0.225	0.685	0.691	0.225	45	0.49	-1.553	0.197	0.512	-0.081	0.162
6	1.775	-0.019	0.122	1.775	-0.019	0.122	46	1.593	0.013	0.194	0.774	-1.447	0.245
7	0.625	-0.629	0.272	0.625	-0.629	0.272	47	0.62	0.661	0.172	0.533	1.819	0.179
8	0.601	-0.656	0.208	0.601	-0.656	0.208	48	0.423	-0.194	0.214	1.657	-0.524	0.147
9	1.566	1.561	0.185	1.566	1.561	0.185	49	0.954	-0.119	0.139	1.825	2.32	0.226
10	1.195	0.862	0.209	1.195	0.862	0.209	50	0.699	0.436	0.185	1.192	0.518	0.147
11	0.952	-0.306	0.062	0.952	-0.306	0.062	51	0.332	-0.01	0.184	1.108	0.229	0.185
12	0.305	-1.645	0.143	0.305	-1.645	0.143	52	1.417	-1.667	0.148	0.367	0.386	0.296
13	0.688	-0.218	0.205	0.688	-0.218	0.205	53	0.653	-0.033	0.149	0.952	0.21	0.142
14	4.239	-0.8	0.175	4.239	-0.8	0.175	54	1.005	-0.012	0.371	2.116	-1.787	0.213
15	1.354	-0.024	0.164	1.354	-0.024	0.164	55	2.014	0.495	0.231	0.925	0.141	0.152
16	0.468	1.133	0.156	0.468	1.133	0.156	56	2.341	0.106	0.105	1.536	-0.171	0.06
17	0.242	-1.096	0.195	0.242	-1.096	0.195	57	1.739	0.248	0.209	0.758	2.352	0.246
18	0.861	0.162	0.29	0.861	0.162	0.29	58	0.332	0.541	0.169	2.446	-0.048	0.24
19	0.685	-0.113	0.202	0.685	-0.113	0.202	59	0.809	0.675	0.258	1.472	-1.736	0.209
20	0.858	1.021	0.236	0.858	1.021	0.236	60	0.646	1.212	0.373	1.956	0.823	0.169
21	2.462	-0.493	0.209	2.462	-0.493	0.209	61	1.514	1.811	0.163	0.68	-0.244	0.214
22	1.188	-0.144	0.14	1.188	-0.144	0.14	62	0.428	-1.178	0.2	0.742	0.336	0.319
23	0.708	1.935	0.202	0.708	1.935	0.202	63	0.722	1.551	0.179	1.148	0.151	0.142
24	1.887	0.928	0.25	1.887	0.928	0.25	64	0.79	0.84	0.269	0.841	0.645	0.227
25	1.433	0.387	0.273	1.433	0.387	0.273	65	0.814	-0.261	0.192	0.944	-0.472	0.166
26	0.634	1.965	0.188	0.634	1.965	0.188	66	0.55	2.069	0.207	2.357	-1.207	0.213
27	0.871	-1.289	0.15	0.871	-1.289	0.15	67	0.545	0.661	0.143	1.356	3.175	0.204
28	0.987	0.744	0.14	0.987	0.744	0.14	68	1.349	0.326	0.136	0.628	-0.131	0.217
29	1.253	0.589	0.054	1.253	0.589	0.054	69	1.417	0.239	0.216	1.221	2.231	0.197
30	0.679	-0.894	0.299	0.679	-0.894	0.299	70	1.769	0.988	0.094	1.662	0.684	0.151
31	2.23	-1.707	0.127	0.625	-0.704	0.145	71	1.004	-0.098	0.247	0.962	0.263	0.204
32	1.43	0.526	0.164	0.405	0.749	0.153	72	2.175	0.934	0.156	0.506	0.261	0.106
33	1.335	1.298	0.138	1.344	0.349	0.191	73	0.889	-1.164	0.161	1.833	0.103	0.368
34	1.01	-0.498	0.263	2.284	2.326	0.18	74	0.854	-0.608	0.145	0.754	-0.393	0.183
35	0.999	0.784	0.288	0.805	0.963	0.141	75	0.437	-0.735	0.208	1.025	-0.366	0.133
36	0.779	1.263	0.173	1.511	1.266	0.239	76	0.665	1.68	0.22	0.761	1.885	0.136
37	0.582	1.104	0.326	1.251	1.253	0.241	77	0.722	-0.073	0.149	0.576	-2.077	0.147
38	1.275	0.07	0.151	1.058	0.987	0.171	78	1.756	2.042	0.212	1.335	-1.089	0.313
39	1.46	0.83	0.178	0.503	0.872	0.136	79	0.274	-0.471	0.206	2.021	0.137	0.213
40	0.585	-0.308	0.191	0.788	1.089	0.083	80	1.423	0.67	0.228	2.205	1.495	0.175

1-30 arasındaki maddeler ortak maddelerdir.

Tablo 28. 2PLM 3000 kişi pozitif çarpık 30 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	0.979	0.242	0.169	0.979	0.242	0.169	16	0.861	-1.385	0.253	0.53	1.638	0.177
2	1.641	-0.002	0.228	1.641	-0.002	0.228	17	1.493	2.51	0.191	0.525	1.019	0.306
3	0.499	-0.134	0.226	0.499	-0.134	0.226	18	0.776	0.205	0.134	1.162	0.186	0.208
4	0.871	-0.706	0.189	0.871	-0.706	0.189	19	1.314	-0.361	0.182	0.778	0.991	0.271
5	0.675	-0.88	0.336	0.675	-0.88	0.336	20	0.747	0.303	0.129	0.927	-1.038	0.099
6	0.734	-0.033	0.151	0.734	-0.033	0.151	21	0.931	0.309	0.145	1.183	-1.095	0.204
7	2.012	-0.479	0.207	2.012	-0.479	0.207	22	3.442	-0.231	0.29	1.118	-0.272	0.196
8	1.895	-0.052	0.371	1.895	-0.052	0.371	23	0.681	1.077	0.226	1.093	0.713	0.325
9	1.893	0.099	0.258	1.893	0.099	0.258	24	2.415	0.307	0.109	0.873	1.463	0.361
10	0.63	0.968	0.2	0.63	0.968	0.2	25	1.511	0.18	0.134	0.931	-0.018	0.242
11	2.139	0.515	0.134	1.139	0.902	0.192	26	0.884	0.355	0.191	1.957	0.851	0.218
12	0.715	-0.252	0.181	0.916	1.053	0.176	27	1.746	0.753	0.148	1.059	1.024	0.11
13	0.736	-0.104	0.103	3.078	0.531	0.142	28	1.225	-1.367	0.224	0.868	1.595	0.105
14	0.698	0.195	0.218	1.497	-0.487	0.135	29	0.867	0.547	0.247	2.337	0.27	0.22
15	0.776	-0.086	0.224	0.941	0.897	0.152	30	0.995	-0.298	0.25	0.632	2.612	0.149

1-10 arasındaki maddeler ortak maddelerdir.

Tablo 29. 2PLM 3000 kişi pozitif çarpık 60 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	1.777	1.83	0.2	1.777	1.83	0.2	31	0.855	0.143	0.213	1.161	0.292	0.142
2	0.828	-0.506	0.313	0.828	-0.506	0.313	32	0.77	-1.4	0.311	1.116	0.42	0.216
3	1.844	0.378	0.138	1.844	0.378	0.138	33	1.447	0.088	0.341	0.639	0.97	0.221
4	1.439	0.16	0.271	1.439	0.16	0.271	34	1.724	0.158	0.187	1.178	1.048	0.225
5	0.973	-1.326	0.22	0.973	-1.326	0.22	35	2.538	-2.873	0.199	1.843	-0.861	0.28
6	2.936	0.964	0.176	2.936	0.964	0.176	36	0.892	-0.081	0.201	0.623	0.799	0.254
7	0.786	1.78	0.176	0.786	1.78	0.176	37	0.947	-0.443	0.232	0.867	0.452	0.25
8	1.327	-0.338	0.103	1.327	-0.338	0.103	38	0.939	-1.262	0.218	0.451	0.698	0.137
9	0.96	0.63	0.213	0.96	0.63	0.213	39	0.402	-1.17	0.156	2.921	-1.121	0.2
10	0.849	0.438	0.296	0.849	0.438	0.296	40	0.509	-0.485	0.237	1.14	-0.633	0.175
11	0.573	-0.531	0.165	0.573	-0.531	0.165	41	0.849	-0.607	0.119	1.697	1.086	0.132
12	1.894	-0.713	0.135	1.894	-0.713	0.135	42	0.614	-0.023	0.218	1.003	0.001	0.141
13	0.943	0.2	0.074	0.943	0.2	0.074	43	0.645	-0.406	0.167	1.175	0.759	0.254
14	0.962	-0.305	0.235	0.962	-0.305	0.235	44	1.513	-1.16	0.159	1.003	0.054	0.22
15	0.472	0.801	0.3	0.472	0.801	0.3	45	0.911	-0.76	0.155	0.94	0.363	0.092
16	0.895	-1.811	0.09	0.895	-1.811	0.09	46	0.873	0.103	0.213	1.4	0.784	0.201
17	0.95	0.397	0.328	0.95	0.397	0.328	47	1.078	-0.994	0.114	0.957	1.899	0.31
18	1.045	-0.554	0.11	1.045	-0.554	0.11	48	0.806	0.32	0.214	1.329	1.743	0.083
19	1.311	1.832	0.36	1.311	1.832	0.36	49	0.461	-0.263	0.123	0.674	-0.497	0.165
20	1.274	-0.076	0.221	1.274	-0.076	0.221	50	4.741	-0.57	0.218	0.641	0.181	0.293
21	0.932	-0.368	0.163	0.843	-0.531	0.146	51	0.589	0.848	0.15	0.741	-0.264	0.155
22	1.222	-0.451	0.108	1.267	0.738	0.321	52	1.123	0.006	0.249	1.895	2.393	0.29
23	0.984	-0.276	0.329	1.77	2.825	0.094	53	0.615	-0.614	0.281	0.883	-0.017	0.126
24	1.122	0.381	0.168	0.728	-0.504	0.212	54	1.632	1.129	0.195	0.604	-0.824	0.222
25	0.493	0.474	0.081	0.767	1.215	0.15	55	1.661	0.425	0.13	0.849	0.383	0.193
26	0.632	0.911	0.109	0.699	-0.421	0.172	56	1.198	0.032	0.103	0.819	2.601	0.206
27	1.373	-0.023	0.165	2.767	1.999	0.167	57	0.407	-0.171	0.145	1.63	0.262	0.187
28	1.195	-0.554	0.25	0.481	-1.454	0.147	58	1.387	-0.026	0.208	0.769	1.527	0.182
29	0.472	-2.414	0.069	1.23	1.163	0.337	59	0.866	0.247	0.206	1.807	2.204	0.214
30	1.299	1.883	0.267	1.968	-0.963	0.285	60	0.309	-0.671	0.235	0.702	0.474	0.217

1-20 arasındaki maddeler ortak maddelerdir.

Tablo 30. 3PLM 3000 kişi pozitif çarpık 80 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	0.575	-1.713	0.147	0.575	-1.713	0.147	41	1.553	-0.686	0.274	1.009	-0.599	0.156
2	0.64	-0.146	0.233	0.64	-0.146	0.233	42	1.049	0.638	0.179	1.948	0.54	0.152
3	0.948	1.808	0.197	0.948	1.808	0.197	43	1.153	1.333	0.211	0.486	1.041	0.126
4	1.146	-0.398	0.159	1.146	-0.398	0.159	44	1.051	-0.417	0.177	1.113	1.117	0.225
5	1.702	-0.215	0.225	1.702	-0.215	0.225	45	0.664	0.683	0.14	0.711	-1.201	0.098
6	2.656	0.188	0.261	2.656	0.188	0.261	46	1.298	0.629	0.201	1.212	1.471	0.24
7	0.866	-1.658	0.206	0.866	-1.658	0.206	47	1.412	-0.29	0.094	1.995	0.961	0.198
8	1.193	-1.157	0.245	1.193	-1.157	0.245	48	0.897	1.526	0.28	0.514	0.186	0.272
9	1.316	-0.54	0.127	1.316	-0.54	0.127	49	0.319	0.052	0.191	1.558	1.608	0.103
10	0.412	0.238	0.141	0.412	0.238	0.141	50	0.395	-0.167	0.296	2.098	0.869	0.145
11	0.993	0.052	0.232	0.993	0.052	0.232	51	0.881	0.673	0.236	0.381	-0.51	0.147
12	1.366	0.804	0.369	1.366	0.804	0.369	52	0.863	1.192	0.35	2.076	1.196	0.154
13	2.097	1.557	0.091	2.097	1.557	0.091	53	2.076	1.41	0.241	0.7	2.101	0.234
14	1.032	0.208	0.15	1.032	0.208	0.15	54	0.748	0.16	0.2	1.974	-0.179	0.143
15	1.356	1.232	0.241	1.356	1.232	0.241	55	1.197	-0.352	0.338	0.81	2.028	0.289
16	0.548	-1.736	0.162	0.548	-1.736	0.162	56	0.428	0.085	0.267	0.933	1.925	0.152
17	1.683	-0.396	0.156	1.683	-0.396	0.156	57	1.259	1.126	0.078	1.301	2.664	0.205
18	1.155	-0.573	0.151	1.155	-0.573	0.151	58	1.846	-0.042	0.187	1.365	0.984	0.098
19	1.842	-0.377	0.169	1.842	-0.377	0.169	59	0.584	0.125	0.141	0.844	0.558	0.246
20	1.038	-0.039	0.207	1.038	-0.039	0.207	60	2.382	0.012	0.199	2.061	2.103	0.316
21	2.054	-1.126	0.164	2.054	-1.126	0.164	61	1.386	-2.731	0.158	2.133	-1.153	0.35
22	1.173	-0.777	0.164	1.173	-0.777	0.164	62	2.18	0.939	0.079	1.283	0.293	0.315
23	0.904	-0.295	0.24	0.904	-0.295	0.24	63	1.288	0.198	0.215	1.24	0.069	0.171
24	0.602	0.688	0.156	0.602	0.688	0.156	64	0.732	-0.393	0.147	1.097	-1.026	0.184
25	0.749	-0.864	0.148	0.749	-0.864	0.148	65	0.874	-0.252	0.214	1.625	-2.389	0.167
26	1.479	-0.334	0.157	1.479	-0.334	0.157	66	0.974	0.102	0.234	0.713	0.64	0.124
27	0.557	-0.015	0.256	0.557	-0.015	0.256	67	0.486	1.668	0.169	1.443	-0.134	0.118
28	0.83	-0.656	0.189	0.83	-0.656	0.189	68	0.525	0.406	0.202	0.696	0.433	0.088
29	0.857	-0.753	0.218	0.857	-0.753	0.218	69	1.11	0.306	0.156	0.973	1.147	0.271
30	0.684	1.976	0.193	0.684	1.976	0.193	70	1.063	-1.49	0.228	0.89	1.136	0.266
31	1.609	0.169	0.276	0.632	0.759	0.143	71	1.091	1.142	0.191	1.967	0.751	0.134
32	0.755	-0.957	0.14	0.786	1.007	0.099	72	1.579	-0.707	0.143	0.663	2.87	0.193
33	1.019	2.27	0.22	0.951	-0.147	0.105	73	0.371	-1.943	0.275	1.463	-0.236	0.237
34	0.799	1.919	0.276	2.474	1.569	0.145	74	1.504	0.506	0.135	0.718	0.095	0.181
35	0.795	-2.044	0.135	0.876	-1.237	0.115	75	0.686	0.087	0.248	0.539	1.531	0.299
36	0.763	-0.088	0.148	0.438	0.069	0.218	76	0.529	-0.308	0.18	0.403	-1.147	0.159
37	1.602	0	0.154	0.346	0.571	0.126	77	1.068	1.068	0.255	0.643	-0.977	0.101
38	0.848	1.798	0.2	1.76	0.612	0.18	78	0.966	-0.032	0.288	0.75	1.61	0.165
39	1.658	0.197	0.202	2.085	0.709	0.287	79	2.166	2.263	0.166	0.618	0.665	0.273
40	1.058	-0.21	0.279	0.974	0.96	0.117	80	0.895	-0.24	0.151	0.86	0.869	0.133

1-30 arasındaki maddeler ortak maddelerdir.

Tablo 31. 3PLM 3000 kişi negatif çarpık 30 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	0.969	-0.452	0.214	0.969	-0.452	0.214	16	1.129	-0.767	0.228	1.091	1.313	0.205
2	2.041	-0.126	0.199	2.041	-0.126	0.199	17	1.443	0.645	0.258	1.869	-1.009	0.21
3	2.371	0.416	0.187	2.371	0.416	0.187	18	0.995	0.074	0.227	1.006	1.895	0.213
4	1.407	0.342	0.218	1.407	0.342	0.218	19	1.399	0.979	0.238	1.41	1.571	0.09
5	0.912	-0.326	0.15	0.912	-0.326	0.15	20	0.611	0.215	0.203	0.838	2.402	0.206
6	0.622	1.001	0.217	0.622	1.001	0.217	21	0.884	-0.999	0.087	2.332	0.831	0.286
7	0.839	0.465	0.164	0.839	0.465	0.164	22	0.78	-0.629	0.192	2.086	-0.155	0.167
8	2.263	1.182	0.181	2.263	1.182	0.181	23	0.627	-0.832	0.152	0.547	2.401	0.275
9	0.519	0.356	0.116	0.519	0.356	0.116	24	1.482	0.683	0.246	0.965	0.057	0.181
10	0.705	1.153	0.222	0.705	1.153	0.222	25	0.632	-0.42	0.207	0.972	-0.302	0.146
11	0.743	0.273	0.178	1.447	0.304	0.249	26	1.09	0.608	0.368	0.695	-0.749	0.342
12	1.004	0.882	0.169	0.639	-0.01	0.215	27	0.534	0.906	0.179	2.165	-0.102	0.267
13	0.494	0.813	0.168	2.048	1.825	0.269	28	0.915	-1.497	0.255	1.505	1.046	0.235
14	1.967	0.875	0.218	2.689	1.395	0.26	29	0.664	1.806	0.183	1.024	0.723	0.122
15	0.678	-0.486	0.251	0.75	0.572	0.228	30	1.897	0.472	0.26	1.313	0.35	0.129

1-10 arasındaki maddeler ortak maddelerdir.

Tablo 32. 3PLM 1000 kişi negatif çarpık 60 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	1.094	0.41	0.175	1.094	0.41	0.175	31	2.389	0.412	0.169	2.038	2.268	0.212
2	0.519	1.351	0.228	0.519	1.351	0.228	32	1.13	-0.138	0.193	0.963	-0.639	0.318
3	1.241	1.728	0.133	1.241	1.728	0.133	33	1.153	-1.419	0.199	0.514	1.158	0.307
4	0.726	-0.19	0.112	0.726	-0.19	0.112	34	0.746	1.391	0.155	0.717	0.423	0.142
5	1.01	1.982	0.204	1.01	1.982	0.204	35	0.777	0.233	0.236	0.668	1.287	0.2
6	1.644	1.226	0.151	1.644	1.226	0.151	36	3.122	0.826	0.212	0.765	-1.187	0.118
7	1.343	-0.994	0.213	1.343	-0.994	0.213	37	0.885	-0.546	0.196	1.382	0.032	0.18
8	0.587	0.319	0.246	0.587	0.319	0.246	38	1.528	0.887	0.236	0.465	-0.012	0.166
9	0.684	-1.371	0.195	0.684	-1.371	0.195	39	2.36	-0.11	0.191	1.36	1.067	0.174
10	1.353	2.353	0.214	1.353	2.353	0.214	40	0.418	-0.74	0.294	1.5	0.341	0.157
11	0.427	-0.915	0.215	0.427	-0.915	0.215	41	0.401	-2.103	0.275	0.964	1.744	0.253
12	0.684	1.857	0.381	0.684	1.857	0.381	42	1.134	0.661	0.162	0.622	2.708	0.281
13	1.528	-2.212	0.287	1.528	-2.212	0.287	43	0.942	0.422	0.307	0.844	0.813	0.267
14	1.197	-0.455	0.282	1.197	-0.455	0.282	44	0.564	0.353	0.172	0.241	0.099	0.214
15	0.643	-0.561	0.179	0.643	-0.561	0.179	45	1.23	-0.108	0.162	1.218	1.039	0.121
16	1.186	-0.663	0.226	1.186	-0.663	0.226	46	0.603	-0.386	0.164	1.08	-0.057	0.248
17	0.599	0.443	0.152	0.599	0.443	0.152	47	2.498	0.787	0.157	2.459	0.057	0.21
18	0.874	0.367	0.182	0.874	0.367	0.182	48	1.483	-0.175	0.296	0.714	0.293	0.175
19	0.462	-1.202	0.24	0.462	-1.202	0.24	49	2.541	2.308	0.179	0.752	0.865	0.085
20	0.687	0.18	0.237	0.687	0.18	0.237	50	1.818	-0.151	0.164	0.491	1.488	0.268
21	3.158	-0.095	0.214	1.747	0.995	0.164	51	0.638	-0.591	0.216	0.863	1.476	0.173
22	0.701	-1.396	0.216	0.464	0.875	0.214	52	0.769	-1.4	0.139	1.143	-2.27	0.195
23	1.283	-0.308	0.353	0.293	0.383	0.24	53	0.254	-0.424	0.065	1.223	-0.032	0.176
24	1.123	-0.692	0.168	0.591	0.542	0.161	54	1.632	-0.723	0.143	0.889	-0.797	0.189
25	1.357	-0.423	0.222	0.591	-0.933	0.235	55	1.147	-1.636	0.222	0.564	-0.429	0.168
26	0.948	-0.059	0.079	1.239	2.142	0.211	56	1.334	-0.189	0.138	1.089	1.18	0.186
27	0.694	0.493	0.111	1.797	1.573	0.159	57	0.412	0.982	0.173	0.463	-0.249	0.182
28	0.919	0.405	0.126	1.602	0.678	0.209	58	1.127	-0.281	0.163	0.946	0.38	0.324
29	1.014	-0.67	0.253	0.982	0.02	0.278	59	1.234	-0.684	0.265	1.226	0.075	0.203
30	2.332	0.084	0.18	0.758	0.066	0.113	60	1.225	0.892	0.243	1.868	-0.01	0.234

1-20 arasındaki maddeler ortak maddelerdir.

Tablo 33. 3PLM 1000 kişi negatif çarpık 80 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	1.065	-2.951	0.357	1.065	-2.951	0.357	41	0.758	0.708	0.363	0.436	-1.046	0.123
2	1.217	0.973	0.22	1.217	0.973	0.22	42	1.296	-1.384	0.156	0.33	-1.113	0.167
3	1.483	0.651	0.176	1.483	0.651	0.176	43	2.394	1.156	0.216	2.18	-0.594	0.198
4	1.157	-2.723	0.172	1.157	-2.723	0.172	44	0.515	-0.553	0.165	0.714	-1.312	0.225
5	0.777	0.169	0.194	0.777	0.169	0.194	45	1.315	1.292	0.118	1.152	0.752	0.168
6	1.596	2.398	0.308	1.596	2.398	0.308	46	1.426	-1.361	0.216	0.619	-0.188	0.289
7	1.176	-1.811	0.143	1.176	-1.811	0.143	47	0.696	0.614	0.232	0.686	1.64	0.211
8	2.076	0.266	0.237	2.076	0.266	0.237	48	0.336	-0.08	0.352	0.3	-0.398	0.235
9	1.549	-0.485	0.259	1.549	-0.485	0.259	49	2.841	0.227	0.21	0.9	0.929	0.177
10	1.064	-0.337	0.24	1.064	-0.337	0.24	50	0.68	-1.275	0.122	2.595	0.288	0.218
11	0.179	-0.798	0.378	0.179	-0.798	0.378	51	1.191	0.56	0.161	1.07	4.082	0.164
12	0.686	-1.076	0.218	0.686	-1.076	0.218	52	1.1	-1.346	0.212	0.778	-0.269	0.08
13	0.871	1.121	0.14	0.871	1.121	0.14	53	0.881	0.297	0.281	0.852	-0.167	0.339
14	1.271	1.193	0.235	1.271	1.193	0.235	54	0.568	-0.058	0.255	1.869	0.983	0.171
15	1.912	-0.895	0.207	1.912	-0.895	0.207	55	0.411	-1.296	0.203	1.588	0.842	0.195
16	0.71	-0.647	0.092	0.71	-0.647	0.092	56	0.613	0.236	0.154	0.766	-0.956	0.242
17	1.326	-0.738	0.361	1.326	-0.738	0.361	57	1.019	-1.071	0.258	1.301	0.908	0.171
18	1.82	-1.045	0.382	1.82	-1.045	0.382	58	1.089	-0.741	0.229	0.537	1.126	0.173
19	0.916	-2.837	0.159	0.916	-2.837	0.159	59	1.966	0.323	0.199	3.956	-0.468	0.211
20	1.575	0.694	0.108	1.575	0.694	0.108	60	0.31	1.668	0.159	0.858	0.449	0.224
21	0.431	-1.337	0.184	0.431	-1.337	0.184	61	0.99	-0.359	0.142	1.615	0.46	0.232
22	1.338	0.802	0.143	1.338	0.802	0.143	62	0.435	1.632	0.257	0.713	0.763	0.15
23	1.423	-1.095	0.199	1.423	-1.095	0.199	63	0.724	0.136	0.234	1.388	1.94	0.151
24	0.573	0.049	0.206	0.573	0.049	0.206	64	1.669	-1.487	0.165	0.622	0.988	0.221
25	1.28	0.6	0.332	1.28	0.6	0.332	65	0.431	-1.671	0.184	0.698	-0.239	0.143
26	0.755	0.587	0.21	0.755	0.587	0.21	66	1.803	2.6	0.196	0.597	0.755	0.142
27	0.87	1.254	0.288	0.87	1.254	0.288	67	0.96	-0.382	0.204	0.681	1.01	0.18
28	1.603	-1.109	0.26	1.603	-1.109	0.26	68	1.947	-0.363	0.207	0.692	0.082	0.166
29	1.648	-0.7	0.179	1.648	-0.7	0.179	69	0.808	0.325	0.232	1.397	-0.603	0.148
30	1.281	-1.121	0.095	1.281	-1.121	0.095	70	1.223	-0.585	0.308	1.565	1.599	0.119
31	0.654	-0.595	0.188	0.779	0.162	0.24	71	1.236	-1.1	0.197	1.742	-0.402	0.254
32	0.769	-0.809	0.197	2.01	0.801	0.195	72	0.839	-0.267	0.137	0.981	0.305	0.218
33	0.759	-0.587	0.269	1.606	-0.69	0.174	73	0.442	-1.509	0.215	1.013	1.991	0.093
34	1.036	-1.951	0.243	1.361	0.07	0.148	74	1.105	-0.486	0.091	2.196	0.802	0.177
35	1.796	-0.819	0.149	1.471	0.268	0.199	75	1.349	0.162	0.238	0.472	-0.086	0.209
36	1.758	1.631	0.255	1.409	0.602	0.153	76	0.861	-0.241	0.185	2.212	0.296	0.262
37	0.715	0.064	0.246	1.061	-1.072	0.134	77	2.806	0.44	0.17	0.686	0.644	0.205
38	1.168	0.124	0.229	1.622	0.24	0.218	78	0.867	-0.027	0.33	0.704	2.147	0.181
39	0.507	0.057	0.248	0.434	-0.471	0.217	79	0.286	0.415	0.246	0.615	1.455	0.212
40	0.808	-0.644	0.118	1.024	1.455	0.13	80	1.101	-1.602	0.162	1.274	1.094	0.202

1-30 arasındaki maddeler ortak maddelerdir.

Tablo 34. 3PLM 3000 kişi negatif çarpık 30 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	0.82	1.576	0.15	0.82	1.576	0.15	16	0.937	0.24	0.146	1.178	1.74	0.193
2	1.191	-1.5	0.18	1.191	-1.5	0.18	17	1.7	-1.631	0.203	0.668	2.038	0.228
3	1.42	-0.745	0.212	1.42	-0.745	0.212	18	1.728	-0.11	0.152	0.928	-0.431	0.189
4	0.396	0.35	0.186	0.396	0.35	0.186	19	0.724	-1.003	0.125	0.598	0.629	0.198
5	1.027	0.906	0.203	1.027	0.906	0.203	20	0.789	0.522	0.2	0.829	-0.025	0.276
6	0.198	0.408	0.178	0.198	0.408	0.178	21	1.889	-0.114	0.133	1.126	1.137	0.209
7	0.476	0.588	0.263	0.476	0.588	0.263	22	0.863	0.191	0.172	1.641	1.237	0.128
8	0.767	1.044	0.187	0.767	1.044	0.187	23	0.661	-1.133	0.072	0.629	0.57	0.219
9	1.079	0.136	0.267	1.079	0.136	0.267	24	0.505	0.734	0.214	0.643	-0.117	0.267
10	1.019	1.675	0.157	1.019	1.675	0.157	25	0.554	1.847	0.402	0.977	2.124	0.195
11	1.085	-0.481	0.231	0.808	2.331	0.141	26	0.482	-1.911	0.227	0.822	0.921	0.213
12	1.343	-1.204	0.143	0.341	-1.298	0.251	27	0.295	-0.974	0.26	2.073	1.538	0.197
13	0.98	1.085	0.183	1.679	0.941	0.261	28	0.587	0.908	0.16	0.788	0.295	0.299
14	1.102	-0.188	0.27	0.815	2.044	0.221	29	0.993	1.803	0.122	1.289	-0.66	0.183
15	0.93	-0.701	0.308	0.371	-0.311	0.221	30	1.496	0.829	0.187	2.669	-0.361	0.222

1-10 arasındaki maddeler ortak maddelerdir.

Tablo 35. 3PLM 3000 kişi negatif çarpık 60 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	2.349	-1.043	0.318	2.349	-1.043	0.318	31	0.887	-1.61	0.188	0.679	-0.82	0.131
2	0.527	-0.016	0.198	0.527	-0.016	0.198	32	1.491	0.547	0.271	0.851	-0.066	0.166
3	0.277	0.415	0.147	0.277	0.415	0.147	33	0.658	-1.109	0.151	1.105	2.044	0.119
4	0.733	1.766	0.187	0.733	1.766	0.187	34	0.444	0.253	0.299	0.901	1.006	0.203
5	2.135	1.635	0.194	2.135	1.635	0.194	35	1.164	1.184	0.165	0.936	0.888	0.253
6	1.959	0.117	0.15	1.959	0.117	0.15	36	2.455	-0.554	0.355	1.552	2.622	0.096
7	0.901	-1.568	0.199	0.901	-1.568	0.199	37	0.863	1.027	0.141	0.585	-0.787	0.212
8	1.145	0.569	0.391	1.145	0.569	0.391	38	0.749	-0.427	0.182	0.738	1.062	0.133
9	0.595	-0.239	0.27	0.595	-0.239	0.27	39	0.703	0.423	0.202	0.5	1.293	0.321
10	0.652	-0.298	0.226	0.652	-0.298	0.226	40	3.277	-0.42	0.148	0.57	-0.655	0.352
11	1.496	0.835	0.168	1.496	0.835	0.168	41	1.983	2.93	0.143	0.855	0.066	0.183
12	1.077	-0.985	0.144	1.077	-0.985	0.144	42	0.538	-0.712	0.161	1.307	1.319	0.151
13	2.086	1.166	0.18	2.086	1.166	0.18	43	0.822	0.703	0.229	2.327	0.523	0.194
14	0.793	-0.626	0.222	0.793	-0.626	0.222	44	1.389	-0.19	0.133	1.02	3.367	0.176
15	1.231	2.283	0.148	1.231	2.283	0.148	45	0.546	0.173	0.175	0.493	0.965	0.196
16	1.097	-0.232	0.223	1.097	-0.232	0.223	46	0.623	0.593	0.166	1.081	0.301	0.156
17	1.722	0.659	0.197	1.722	0.659	0.197	47	0.718	-1.009	0.308	0.737	-0.036	0.125
18	1.195	0.745	0.412	1.195	0.745	0.412	48	1.659	0.582	0.17	1.637	0.829	0.23
19	0.555	-0.883	0.245	0.555	-0.883	0.245	49	0.742	0.39	0.088	1.183	1.128	0.167
20	0.558	-1.259	0.271	0.558	-1.259	0.271	50	1.447	-0.907	0.213	1.373	0.319	0.196
21	0.798	0.535	0.239	0.548	-0.119	0.205	51	0.956	0.266	0.215	1.648	1.944	0.178
22	0.31	0.938	0.164	0.681	-0.352	0.3	52	1.905	-0.881	0.307	0.472	0.899	0.227
23	0.745	0.485	0.143	1.38	0.99	0.163	53	0.518	0.257	0.293	0.809	1.141	0.226
24	0.949	-0.138	0.113	1.719	1.374	0.264	54	0.843	1.957	0.237	0.94	0.794	0.143
25	0.793	-0.685	0.268	1.006	-0.978	0.179	55	0.738	1.129	0.231	0.673	1.035	0.107
26	0.679	1.742	0.128	1.54	1.44	0.174	56	0.923	0.966	0.133	1.454	1.73	0.181
27	1.305	0.186	0.162	1.355	0.886	0.12	57	0.592	0.466	0.161	0.949	-0.533	0.16
28	1.147	1.023	0.211	0.783	0.673	0.171	58	1.557	-0.745	0.182	0.64	1.661	0.268
29	0.467	0.238	0.144	1.234	-0.079	0.143	59	1.434	-0.551	0.124	1.305	0.189	0.245
30	0.915	-0.703	0.205	1.279	1.589	0.216	60	1.022	-1.216	0.299	0.567	2.16	0.234

1-20 arasındaki maddeler ortak maddelerdir.

Tablo 36. 3PLM 3000 kişi negatif çarpık 80 madde

m	Form1			Form2			m	Form1			Form2		
	a	b	c	a	b	c		a	b	c	a	b	c
1	1.463	1.074	0.078	1.463	1.074	0.078	41	1.307	0.782	0.194	0.628	-0.059	0.144
2	1.15	0.771	0.213	1.15	0.771	0.213	42	0.671	-0.248	0.215	1.428	1.263	0.078
3	0.928	0.395	0.142	0.928	0.395	0.142	43	1.963	-1.013	0.252	1.013	0.354	0.088
4	0.269	-0.694	0.162	0.269	-0.694	0.162	44	0.725	0.543	0.159	0.419	-0.17	0.158
5	1.569	0.792	0.201	1.569	0.792	0.201	45	1.373	0.649	0.262	0.831	-0.062	0.112
6	0.708	-1.347	0.164	0.708	-1.347	0.164	46	0.488	0.112	0.21	0.917	2.02	0.261
7	0.857	-0.509	0.141	0.857	-0.509	0.141	47	1.641	0.466	0.212	1.68	-0.93	0.288
8	0.448	0.706	0.12	0.448	0.706	0.12	48	1.255	-0.076	0.078	1.364	0.51	0.207
9	1.577	0.412	0.266	1.577	0.412	0.266	49	0.918	0.726	0.279	0.621	0.505	0.261
10	1.395	0.972	0.229	1.395	0.972	0.229	50	1.325	1.56	0.189	2.38	0.681	0.278
11	1.965	0.269	0.171	1.965	0.269	0.171	51	1.005	-2.969	0.149	0.546	2.398	0.151
12	1.539	-1.248	0.117	1.539	-1.248	0.117	52	0.54	-0.399	0.165	0.771	-0.098	0.112
13	0.611	-1.718	0.16	0.611	-1.718	0.16	53	1.057	0.136	0.136	0.507	1.034	0.21
14	0.624	0.65	0.295	0.624	0.65	0.295	54	0.332	-1.707	0.259	1.905	0.334	0.226
15	1.758	0.575	0.182	1.758	0.575	0.182	55	0.753	1.116	0.19	0.416	1.058	0.14
16	1.101	-1.349	0.189	1.101	-1.349	0.189	56	2.043	0.159	0.149	0.627	1.108	0.183
17	0.781	1.504	0.171	0.781	1.504	0.171	57	0.566	-0.825	0.151	1.232	0.17	0.157
18	2.114	-0.406	0.177	2.114	-0.406	0.177	58	1.43	-0.061	0.068	0.826	1.837	0.073
19	2.525	1.401	0.208	2.525	1.401	0.208	59	1.479	1.25	0.155	1.4	-0.309	0.165
20	0.564	-1.174	0.152	0.564	-1.174	0.152	60	1.975	-0.876	0.181	1.109	1.243	0.357
21	0.85	2.138	0.194	0.85	2.138	0.194	61	3.081	1.34	0.19	1.387	-0.304	0.167
22	1.315	0.124	0.141	1.315	0.124	0.141	62	1.23	1.178	0.199	0.872	-1.55	0.158
23	0.554	0.94	0.265	0.554	0.94	0.265	63	1.031	0.312	0.286	0.971	-0.434	0.123
24	0.458	-0.788	0.194	0.458	-0.788	0.194	64	1.024	-1.499	0.191	0.505	1.668	0.128
25	0.517	2.148	0.104	0.517	2.148	0.104	65	1.019	0.196	0.18	0.967	-0.132	0.186
26	0.451	-0.024	0.159	0.451	-0.024	0.159	66	1.475	-0.524	0.113	0.508	0.253	0.229
27	0.545	0.379	0.217	0.545	0.379	0.217	67	1.272	2.192	0.187	0.761	0.209	0.162
28	1.49	-0.518	0.147	1.49	-0.518	0.147	68	1.206	0.163	0.249	1.757	0.555	0.192
29	1.179	-0.523	0.225	1.179	-0.523	0.225	69	0.681	-0.581	0.177	1.585	-1.592	0.217
30	1.293	1.958	0.108	1.293	1.958	0.108	70	1.421	1.322	0.249	0.828	0.57	0.225
31	0.543	0.592	0.271	1.056	1.775	0.172	71	0.905	1.862	0.152	1.033	1.527	0.156
32	1.397	-0.253	0.364	1.315	1.077	0.137	72	2.146	-0.346	0.216	0.784	0.584	0.172
33	0.983	1.245	0.201	0.853	-0.68	0.196	73	0.656	0.679	0.129	0.238	0.843	0.181
34	0.945	-0.55	0.393	1.032	1.203	0.196	74	1.18	-1.109	0.327	1.531	0.652	0.133
35	1.295	0.794	0.238	0.581	-0.82	0.104	75	0.731	-0.632	0.173	0.532	0.714	0.085
36	0.523	0.797	0.192	1.462	-1.308	0.196	76	0.699	-1.849	0.166	0.729	-0.146	0.158
37	1.587	0.208	0.149	0.639	-1.127	0.254	77	1.265	0.408	0.274	0.605	0.78	0.189
38	0.839	0.31	0.225	1.592	0.772	0.152	78	1.675	1.212	0.2	2.64	0.155	0.17
39	1.008	0.163	0.226	1.06	-0.153	0.168	79	1.429	-0.461	0.128	2.581	0.937	0.256
40	0.923	0.699	0.264	2.567	0.027	0.14	80	1.022	1.232	0.365	0.834	-1.734	0.128

1-30 arasındaki maddeler ortak maddelerdir.

Ek 2. Parscale syntaksı: 2 PLM 1000 kişi 30 madde

PARSCALE syntax generated by WINGEN

2/10/2011 10:09:02 AM

```
>FILE DFNAME='form1_1.wgr',
    SAV;
>SAVE PAR='form1_1.par',
score='form1_1.sco';
>INPUT NID=8,
NTOTAL=30,
NTEST=1;
(8A1,2X,30A1)
>TEST1 TNAME=WINGEN,ITEM=(1(1)30),NBLOCK=1;
>BLOCK1 BNAME=WINGEN01,NITEOS=30,NCAT=2,
ORIGINAL=(0,1);
>CALIB GRADED,DIAGNOSIS=0,
    CYCLES=(100,5,5,5,5),CRIT=0.001,
    TPRIOR,GPRIO,SPRIOR,RIDGE;
>SCORE EAP;
```

EK 3. 1. replikasyondan elde edilen IRT çıktı örneği: 2 PLM 1000 kişi 30 madde IRTEQ (Han, 2007)

output file (2/10/2011 6:26:50 PM)

Test1:C:\DocumentsandSettings\Administrator\Desktop\normal\1000\30\form1\FORM
1_1.PAR

of iteOS: 30

Test2:C:\DocumentsandSettings\Administrator\Desktop\normal\1000\30\form2\FORM
2_1.PAR

of iteOS: 30

of pairs of linking iteOS: 10

mean discrimination of linking items in Test 1: 0.926
 mean difficulty of linking items in Test 1: -0.257
 standard deviation of linking item difficults in Test 1: 1.115
 mean discrimination of linking items in Test 2: 0.933
 mean difficulty of linking items in Test 2: -0.722
 standard deviation of linking item difficults in Test 2: 1.112
 Correlation coeff. for discrimination of linking items: 0.969
 Correlation coeff. for difficulty of linking items: 0.997
 Correlation coeff. for guessing of linking items: 0.000

Equating coefficient A by Ortalama-ortalama method: 1.008
 Equating coefficient B by Ortalama-ortalama method: 0.471
 Equating coefficient A by Ortalama-standart sapma method: 1.003
 Equating coefficient B by Ortalama-standart sapma method: 0.467
 Equating coefficient A by TCC (Stocking & Lord) method: 1.010
 Equating coefficient B by TCC (Stocking & Lord) method: 0.467
 Minimized Loss Function value by TCC (Stocking & Lord) method: 0.00011

Test 2 Items rescaled to Test 1 using Mean/Mean Method were saved in
1.MM.wgi (2/10/2011 6:27:15 PM)

Test 2 Items rescaled to Test 1 using Mean/Sigma Method were saved in
1.MS.wgi (2/10/2011 6:27:15 PM)

Test 2 Items rescaled to Test 1 using TCC(Stocking & Lord) Method were saved in
1.SL.wgi (2/10/2011 6:27:15 PM)

Test 2 scores equated to Test 1 using Mean/Mean Method were saved in
1.MM.ncs (2/10/2011 6:27:15 PM)

Test 2 scores equated to Test 1 using Mean/Sigma Method were saved in
1.MS.ncs (2/10/2011 6:27:15 PM)

Test 2 scores equated to Test 1 using TCC(Stocking & Lord) method were saved in
1.SL.ncs (2/10/2011 6:27:15 PM)

Test 2 rescaled items using Mean/Mean Method were save in 1_MM.PAR
(2/10/2011 6:27:15 PM)

Test 2 rescaled items using Mean/Sigma Method were saved in 1_MS.PAR
(2/10/2011 6:27:15 PM)

Test 2 rescaled iteOS using TCC(Stocking & Lord) method were saved in 1_SL.PAR (2/10/2011 6:27:15 PM)

Conversion Table using Mean/Mean method was saved in 1.MM.CON (2/10/2011 6:27:15 PM)

Conversion Table using Mean/Sigma method was saved in 1.MS.CON (2/10/2011 6:27:16 PM)

Conversion Table using TCC(Stock&Lord) method was saved in 1.SL.CON (2/10/2011 6:27:16 PM)

IRTEQ run (2/10/2011 6:27:16 PM) has successfully finished.

EK 4. 1. replikasyondan elde edilen IRT çıktı örneği: 2 PLM 1000 kişi 30 madde için ortalama-ortalama yönteminden elde edilen eşitlenmiş parametreler

Eşitlenmiş parametreler					
Maddeler	a	b	Maddeler	a	b
1	0.789	-1.324	16	1.64	-0.418
2	1.805	0.842	17	1.31	1.436
3	0.724	-0.473	18	1.03	-1.661
4	1.139	0.164	19	0.568	0.995
5	0.671	0.931	20	0.587	-0.145
6	0.669	0.772	21	0.905	0.28
7	1.206	-2.15	22	0.532	-0.894
8	0.971	-1.054	23	2.102	0.717
9	0.727	-0.903	24	0.816	0.331
10	0.556	-1.701	25	2.32	-0.443
11	0.793	0.281	26	0.577	-0.288
12	0.821	1.234	27	1.702	-0.224
13	1.479	-0.208	28	1.135	-1.953
14	1.01	0.871	29	0.395	-0.278
15	1.294	1.309	30	1.424	2.021

1-10 arasındaki maddeler ortak maddelerdir.

EK 5. 1.replikasyondan elde edilen IRT çıktı örneği: 2PLM 1000 kişi 30 madde için ortalama-standart sapma yönteminden elde edilen eşitlenmiş parametreler

Eşitlenmiş parametreler					
Maddeler	a	b	Maddeler	a	b
1	0.793	-1.322	16	1.648	-0.417
2	1.814	0.839	17	1.316	1.428
3	0.728	-0.473	18	1.035	-1.654
4	1.145	0.163	19	0.57	0.989
5	0.674	0.927	20	0.59	-0.146
6	0.673	0.768	21	0.91	0.277
7	1.212	-2.146	22	0.535	-0.891
8	0.975	-1.053	23	2.113	0.712
9	0.731	-0.902	24	0.82	0.328
10	0.559	-1.698	25	2.332	-0.442
11	0.797	0.278	26	0.58	-0.288
12	0.825	1.226	27	1.711	-0.224
13	1.486	-0.208	28	1.14	-1.944
14	1.015	0.866	29	0.397	-0.278
15	1.3	1.301	30	1.431	2.01

1-10 arasındaki maddeler ortak maddelerdir.

EK 6. 1. replikasyondan elde edilen IRTEQ çıktı örneği: 2PLM 1000 kişi 30 madde için Stocking-Lord yönteminden elde edilen eşitlenmiş parametreler

Maddeler	Eşitlenmiş parametreler				
	a	b	Maddeler	a	b
1	0.787	-1.102	16	1.637	-0.424
2	1.802	1.077	17	1.307	1.434
3	0.723	-0.245	18	2.316	-0.449
4	1.137	0.395	19	0.815	0.327
5	0.670	1.166	20	2.098	0.714
6	0.668	1.006	21	0.531	-0.901
7	1.204	-1.933	22	0.903	0.276
8	0.969	-0.830	23	0.586	-0.151
9	0.726	-0.679	24	0.566	0.992
10	0.555	-1.481	25	1.028	-1.669
11	0.792	0.276	26	0.576	-0.294
12	0.820	1.231	27	1.699	-0.229
13	1.476	-0.213	28	0.394	-0.284
14	1.008	0.868	29	1.132	-1.962
15	1.291	1.306	30	1.421	2.020

1-10 arasındaki maddeler ortak maddelerdir.