



**TEST EŐİTLEMEDE MADDE TEPKİ KURAMINA DAYALI
YETENEK PARAMETRESİNE YÖNELİK
ÖLÇEK DÖNÜŐTÜRME YÖNTEMLERİNİN KARŐILAŐTIRILMASI**

Tuba Gündüz

**YÜKSEK LİSANS TEZİ
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI**

**GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

KASIM, 2015

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren 24 ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı : Tuba
Soyadı : Gündüz
Bölümü : Eğitimde Ölçme ve Değerlendirme
İmza :
Teslim tarihi :

TEZİN

Türkçe Adı : Test Eşitlemede Madde Tepki Kuramına Dayalı Yetenek Parametresine Yönelik Ölçek Dönüştürme Yöntemlerinin Karşılaştırılması.

İngilizce Adı : Comparison of Scale Transformation Methods In Terms Of Ability Based on Item Response Theory in Test Equating.

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduğumu, yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu beyan ederim.

Yazar Adı Soyadı: Tuba Gündüz

İmza:

Jüri onay sayfası

Tuba Gündüz tarafından hazırlanan “Test Eşitlemede Madde Tepki Kuramına Dayalı Yetenek Parametresine Yönelik Ölçek Dönüştürme Yöntemlerinin Karşılaştırılması” adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: Prof. Dr. Mehtap ÇAKAN

(Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi)

Başkan: Doç. Dr. Nuri DOĞAN

(Eğitimde Ölçme ve Değerlendirme, Hacettepe Üniversitesi)

Üye: Doç. Dr. Şeref Tan

(Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi)

Tez Savunma Tarihi: 20 /11 /2015

Bu tezin Eğitimde Ölçme ve Değerlendirme Anabilim Dalı’nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Tahir ATICI

Eğitim Bilimleri Enstitüsü Müdürü



Yeğenim Defne'ye...

TEŞEKKÜR

Yüksek lisans eğitimim boyunca kıymetli tecrübelerinden yararlanırken göstermiş olduğu hoşgörü ve sabrı ile birlikte bu araştırmanın her aşamasında desteğini benden esirgemeyen, varlığına daima ihtiyaç duyacağım danışmanım ve çok değerli hocam Sayın Prof. Dr. Mehtap ÇAKAN'a çok teşekkür ederim.

Eğitimim sürecinde ve çalışmamda önemli katkılar sağlayan ve yardımını hiç esirgemeyen kıymetli hocam Sayın Doç. Dr. Şeref TAN'a ve jürimde bulunan ayrıca bir dönem boyunca dersini alabilmeme olanak sağlayan hocam Sayın Doç. Dr. Nuri DOĞAN'a teşekkür ederim. Akademik bilgilerini ve yardımlarını benden esirgemeyip sorularıma ivedilikle cevap veren Sayın Prof. Michael J. KOLEN'a da teşekkürü borç bilirim.

Çalışmamın tamamlanması sürecinde yanımda olup yardımlarını esirgemeyen değerli arkadaşlarım Yıldız YILDIRIM, Gül İNCE, Sibel ADA, Ayşenur ERDEMİR ve Esra OYAR'a teşekkür ederim.

Her konuda benden yardımını ve sabrını esirgemeyen biricik eşim Fahrettin GÜNDÜZ'e, yalnızca eğitimimde değil, tüm hayatım boyunca bana inanan, güvenen, benim bugünlere gelmemi sağlayan annem Gülseren KARACAN'a ve babam Murat KARACAN ile her yönüyle örnek aldığım desteğini daima hissettiren abim Mustafa KARACAN'a sonsuz teşekkürlerimi sunarım.

**TEST EŞİTLEMEDE MADDE TEPKİ KURAMINA DAYALI
YETENEK PARAMETRESİNE YÖNELİK ÖLÇEK DÖNÜŞTÜRME
YÖNTEMLERİNİN KARŞILAŞTIRILMASI
(Yüksek Lisans Tezi)**

**Tuba Gündüz
GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
Kasım, 2015
ÖZ**

Bu araştırmada, test eşitlemede Madde Tepki Kuramı'na dayalı ortalama-ortalama, ortalama-standart sapma, Stocking-Lord ve Haebara ölçek dönüştürme yöntemlerini yetenek parametresine yönelik olarak gerçek veriler üzerinden karşılaştırmak amaçlanmıştır. İkili puanlanan maddeler için 3PLM altında yürütülen araştırmanın çalışma grubunu TIMSS 2011 dördüncü sınıf düzeyinde seçkisiz seçilen iki matematik formunu cevaplayan (535 ve 529 kişi) Türk öğrenciler olmaktadır. Bu formlar 14'ü ortak olan 26'sar maddeden oluşmaktadır. Eşitleme desenlerinden, formların yapısına ve veri toplama yöntemine uygun olan "ortak maddeli eşdeğer olmayan gruplar deseni" kullanılmıştır. Bu desen kullanımında, farklı formlardan gelen ve ayrı ölçek üzerinde olan yetenek ve madde parametrelerini aynı ölçek üzerine koymak için ST 2.0 programından yararlanarak ölçek dönüştürme yapılmıştır. Bu program ile dört yöntem için ölçek dönüştürmede kullanılacak katsayılar ve sabitler bulunmuştur. Hangi yöntemin daha uygun olduğuna karar vermek için yetenek parametreleri üzerinden "RMSD (Root mean square difference)" ölçütü kullanılmıştır. Araştırmadan elde edilen bulgular doğrultusunda yapılan yetenek parametresine yönelik ölçek dönüştürmede en uygun yöntemden (en az hata içeren yöntem) en az uygun olan yönteme doğru yapılan sıralamanın; (1) Ortalama-ortalama, (2) Stocking-Lord, (3) Haebara ve (4) Ortalama-standart sapma şeklinde olduğu görülmüştür. En uygun bulunan yöntem belirlendikten sonra da bu yöntemin dönüşümünden elde edilen parametreler ile PIE 2.0 programını kullanarak MTK gerçek puan eşitlemesi yapılmıştır. Bunun sonucunda ise Form Y'nin Form X'e göre daha zor olduğu sonucuna ulaşılmıştır.

Bilim Kodu : 10113990

Anahtar Kelimeler : Test Eşitleme, Madde Tepki Kuramı (MTK), Ölçek Dönüştürme Yöntemleri, Ortak Maddeli Eşdeğer Olmayan Gruplar Deseni

Sayfa Adedi : 115

Danışman : Prof. Dr. Mehtap ÇAKAN

**COMPARISON OF SCALE TRANSFORMATION METHODS IN
TERMS OF ABILITY BASED ON ITEM RESPONSE THEORY IN
TEST EQUATING**

(M.S Thesis)

Tuba Gündüz

GAZI UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

November, 2015

ABSTRACT

In this study, it was aimed that the mean-mean, the mean-standard deviation, Stocking-Lord and Haebara scale transformation methods which based on Item Response Theory in test equating were compared in terms of ability parameters (tetas) on real data. For dichotomous data under the 3PLM model, the sample of the study composed of Turkish students who participants of TIMSS 2011 fourth grade level in math chosen two forms answer (535 and 529 students). These forms have 26 items and both of forms contain 14 common item. Choosing from equating design, according to the structure of forms and methods of data collection "common item nonequivalent groups design" is used. This design of use in which to put tetas and item parameters from different forms on different scales on the same scale with the software to be used in the ST scale transformation coefficients and constants were found for all four methods. RMSD (Root mean square difference) criteria were used to decide which method is more appropriate. According to findings obtained from research on the transformation scale for the teta to do the most appropriate parameter (containing at least error method) to the method of ranking made the least favorable; (1) Mean-mean, (2) Stocking-Lord, (3) Haebara and (4) Mean-standard deviation were found. After determining the most appropriate method by which parameters are derived from the transformation of this method using the software PIE 2.0 points made IRT true score equating and interpreted. As a result of this, it is concluded that Form Y is more difficult than Form X.

Science Code : 10113990

Key Words :Test Equating, Item Response Theory (IRT), Scale Transformation Methods, Common Item Non-equivalent Groups Design

Page Number : 115

Supervisor : Prof. Dr. Mehtap ÇAKAN

İÇİNDEKİLER

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU	iv
ETİK İLKELERE UYGUNLUK BEYANI.....	ii
TEŞEKKÜR	v
ÖZ.....	vi
ABSTRACT	vii
İÇİNDEKİLER	vi
TABLolar LİSTESİ.....	ix
ŞEKİLLER LİSTESİ	xi
SİMGELER VE KISALTMALAR CETVELİ	xiii
BÖLÜM I.....	1
GİRİŞ	1
Problem Durumu.....	1
Test Eşitleme	2
Yatay ve Dikey Eşitleme.....	3
Test Eşitlemenin Koşulları.....	4
Test Eşitleme Aşamaları	5
Eşitleme Desenleri.....	6
Klasik Test Kuramı ve Geleneksel Test Eşitleme Yöntemleri.....	12
Madde Tepki Kuramı Modelleri ve Madde Tepki Kuramına Dayalı Eşitleme Yöntemleri.....	16
Ölçek Dönüştürme Yöntemleri (Kalibrasyon).....	22

TIMSS ile ilgili Kısa Bilgi	28
Problem İfadesi.....	30
Araştırma Soruları.....	30
Araştırmanın Amacı ve Önemi	31
Sayıtlar	31
Sınırlılıklar	32
Tanımlar.....	32
BÖLÜM II	35
LİTERATÜR TARAMASI.....	35
BÖLÜM III.....	43
YÖNTEM.....	43
Araştırmanın Modeli.....	43
Çalışma Grubu	43
Veri Toplama Deseni (Eşitleme Deseni)	43
Verilerin Düzenlenmesi ve Analizi.....	44
MTK Sayıtlarının Test Edilmesi.....	50
Tek Boyutluluk	52
Yerel Bağımsızlık	58
Madde Karakteristik Eğrisinin Monotonik Artışı	58
Model Veri Uyumu	58
Araştırmada Kullanılan Ölçek Dönüştürme Yöntemleri	60
Değerlendirme Kriteri	60
BÖLÜM IV	63
BULGULAR VE YORUMLAR	63
Araştırma Sorusu 1’e ait Bulgu ve Yorum:.....	63
Araştırma Sorusu 2’ye ait Bulgu ve Yorum:	65
Araştırma Sorusu 3’e ait Bulgu ve Yorum :	66

BÖLÜM V.....	69
SONUÇ VE TARTIŞMA	69
Araştırma Sorusu 1’e ait Sonuç ve Tartışma:	69
Araştırma Sorusu 2’ye ait Sonuç ve Tartışma:	70
Araştırma Sorusu 3’e ait Sonuç ve Tartışma:	71
BÖLÜM VI.....	73
ÖNERİLER	73
KAYNAKLAR	75
EKLER.....	83
EK 1- ST 2.0 Output Dosyası	84
EK 2 - Form X ve Form Y için Yen’in Q_3 Testi	84
EK 3- PIE 2.0 Input ve Output Dosyası	84
EK 4- Form X ve Form Y’ye Ait Madde Karakteristik Eğrileri	84
EK 5 - Form X ve Form Y’ye ait Her Bir Model Altındaki Madde Uyumu.....	84

TABLÖLAR LİSTESİ

Tablo 1. Form X ve Form Y'nin Ortalamaları, Standart Sapmaları, KR-21 Güvenirlik Katsayıları, Basıklık Katsayıları ve Çarpıklık Katsayıları.....	44
Tablo 2. Form X ve Form Y'ye Ait Maddelerin Parametreleri.....	45
Tablo 3. Formlara Ait Ortalama Ayırıcılık, Ortalama Güçlük ve Ortalama Şans Başarısı Parametreleri.....	46
Tablo 4. Formlarda Bulunan Ortak Madde Setinin Ayırıcılığa (a parametresine) göre İlişkisiz t-testi İle Karşılaştırılması.....	47
Tablo 5. Formlarda Bulunan Ortak Madde Setinin Güçlüğe (b parametresine) göre İlişkisiz t-testi ile Karşılaştırılması.....	49
Tablo 6. Formlarda Bulunan Ortak Madde Setinin Şans Parametresine (c parametresine) göre İlişkisiz t-testi İle Karşılaştırılması.....	50
Tablo 7. Form X'e ait Açıklayıcı Faktör Analizi Sonuçları.....	55
Tablo 8. Form Y'ye ait Açıklayıcı Faktör Analizi Sonuçları	55
Tablo 9. Form X ve Form Y için $-2\ln$ Değerleri	59
Tablo 10. Form X ve Form Y için Uyum Sağlamayan Madde Sayıları.....	60
Tablo 11. OO ile Ölçek Dönüştürme İçin Elde Edilen Dönüştürme Katsayısı ve Sabiti.....	63
Tablo 12. OS ile Ölçek Dönüştürme İçin Elde Edilen Dönüştürme Katsayısı ve Sabiti.....	64
Tablo 13. HA ile Ölçek Dönüştürme İçin Elde Edilen Dönüştürme Katsayısı ve Sabiti....	64
Tablo 14. SL ile Ölçek Dönüştürme İçin Elde Edilen Dönüştürme Katsayısı ve Sabiti.....	65
Tablo 15.Farklı Yöntemler ile Ölçek Dönüştürme Sonucunda Elde Edilen RMSD Ortalamaları.....	66

Tablo 16. MTK Gerçek Puan Eşitleme Yöntemi Kullanılması Sonucu Elde Edilen Eşitlenmiş Puanlar.....	67
--	----



ŞEKİLLER LİSTESİ

Şekil 1. Test eşitlemenin testlerin güçlük düzeyleri bakımından sınıflandırılması.....	4
Şekil 2. Randum grup deseni.....	7
Şekil 3. Tek grup deseni.....	8
Şekil 4. Dengelenmiş tek grup deseni.....	9
Şekil 5. İç ortak maddeli (internal) eşdeğer olmayan grup deseni.....	10
Şekil 6. Dış ortak maddeli (external) eşdeğer olmayan grup deseni.....	10
Şekil 7. Geleneksel test eşitleme yöntemleri.....	12
Şekil 8. Eşit yüzdelikli eşitleme.....	15
Şekil 9. 3PLM için madde karakteristik eğrisi örneği.....	19
Şekil 10. Araştırmada kullanılan test eşitleme yöntemleri.....	24
Şekil 11. Ortak maddelerin a parametrelerinin grafiği.....	47
Şekil 12. Ortak maddelerin b parametrelerinin grafiği.....	48
Şekil 13. Ortak maddelerin c parametrelerinin grafiği.....	49
Şekil 14. Form X'in Q-Q plot grafiği.....	51
Şekil 15. Form Y'nin Q-Q plot grafiği.....	51
Şekil 16. Form X'in öz değerleri saçılma grafiği.....	53
Şekil 17. Form Y'nin öz değerleri saçılma grafiği.....	53
Şekil 18. Form X'e ait tetrakorik faktör analizi öz değer grafiği.....	56
Şekil 19. Form Y'ye ait tetrakorik faktör analizi öz değer grafiği.....	56

Şekil 20. TIMSS 2011 4. sınıf matematik başarı testini oluşturan konu alanları.....	57
Şekil 21. MTK gerçek puan eşitleme kullanarak Form X gerçek puanlarının kestirilen Form Y gerçek puan eşdeğerleri.....	68



SİMGELER VE KISALTMALAR CETVELİ

MTK	Madde Tepki Kuramı
KTK	Klasik Test Kuramı
3PLM	Üç parametrelı lojistik model
2PLM	İki parametrelı lojistik model
1PLM	Bir parametrelı lojistik model
TIMSS	Trends in International Mathematics and Science Studies
RMSD	Root mean square difference (Eşitleme Hatası)
BIAS	Eşitleme yanlılıđı
OO	Ortalama-ortalama yöntemi
OS	Ortalama-standart sapma yöntemi
HA	Haebara yöntemi
SL	Stocking-Lord yöntemi

BÖLÜM I

GİRİŞ

Bu bölümde, test eşitleme ve Madde Tepki Kuramı ile ilgili temel kavramlara, araştırmanın problem durumuna, amacına, önemine, sayıltılarına ve sınırlılıklarına yer verilmiştir.

Problem Durumu

Bir test (sınav), belli davranış örnekleminin belirlenmesi ve bu örneklemin kategorilenmesi ya da bu örneklemin belli puanlarla standartlaştırılması işlemi olarak tanımlanır (De Gruijter ve Leo, 2008). Ancak test kavramı oldukça geniş bir yelpaze altında kullanılabilir öyle ki Erkuş (2012), Thorndike’ın “test” sözcüğünün 40 farklı anlamda kullandığını ileri sürmüştür. Test puanları bireysel ve kurumsal düzeylerde önemli kararlar alırken bilgi sağlamak için sıklıkla kullanılır (Kolen ve Brennan, 2014). Örneğin bir öğrencinin hangi liseye gitmesine veya bir kurum ya da kuruluşta bir meslek grubundaki kabul edilecek gerekli şartları sağlayan bireyler arasında karar verilmesi için test puanlarına ihtiyaç duyulur. Alınan karar, türüne bakmaksızın mümkün olduğunca daha kesin ve daha iyi bilginin üzerine dayandırılmalıdır. Diğer bütün koşullar eşit olmak şartıyla bilgi ne kadar fazla doğru olursa daha iyi karar verilir (Kolen ve Brennan, 2014). Karar vermede kullanılan bilgilerin elde edildiği testler, özellikle geniş ölçekli uygulamalar olduğunda, test gizliliği açısından tekrarlı olarak kullanılamamaktadır. Bu testlerin defalarca kullanılamamasından dolayı da Klasik Test Kuramı (KTK)’na göre aynı amaçlarla yapılan bir teste paralel olması gereken başka formlarla ölçme yapılması gerekir. Aiken (2000)’in de belirttiği gibi, pratikte paralel testlerin oluşturulması oldukça zordur (aktaran, Kilmen, 2010). Benzer özellikteki katılımcı gruplarda aynı yapının ölçülmesi istenilirken farklı zamanda ve/veya testin farklı formu ile yapılan ölçümlerin sonucunun eş olması beklenir. Böyle durumlardan başka; aynı yapıyı ölçen ancak benzer özellikte olmayan katılımcı grupların, farklı güçlükteki iki test ile ölçümlerinin ardından bireylerin yeteneklerinin karşılaştırılması istenen durumlar arasında olabilir. Bu durumların üstesinden test eşitleme çalışmalarıyla gelinebilir.

Ülkemizde geçmişten günümüze kadar yapılan; temel eğitimden orta öğretime geçiş, orta öğretime yüksek öğretime geçiş ve yüksek öğretim sırasında ve sonrasında kurumlara girişlerde bireyler kriter dayanaklı ve/veya norm dayanaklı değerlendirmenin yapılacağı sınavlara katılırlar. Aynı amaç için yapılan testlerde, gizlilik ve diğer sebeplerden dolayı farklı soruların sorulması, testi alan bireylere bazı formların kolay, bazı formların zor gelmesine yol açabilir. Örneğin YDS (Yabancı Dil Bilgisi Seviye Tespit Sınavı) kriter dayanaklı değerlendirmede kullanılabilen bir İngilizce başarı testidir. Bu sınavın geçerlilik süresi, kurumların kendilerine uygun geçerlilik süresi belirleyeceği durumlar hariç 5 yıldır (ÖSYM, 2014). Bu geçerlilik süresi göz önüne alındığında yılda iki dönem yapılan bu sınavın, toplamda on farklı test formundan alınan puanların aynı değerlendirme sürecinde karşılaştırılması olası bir durumdur. Örneğin; bireylere yüksek lisans eğitimi veren bir üniversitede başvuru şartlarından biri YDS puanı 50 ve üzeri alma olsun. Bu durumda farklı zamanlarda yapılan sınavlardan 50 puan alan bireylerin, ölçülen özellik bakımından aynı yeteneğe sahip olması gerekir. Ya da aynı yeteneğe sahip iki bireyin farklı sınavlardan aldığı puanların birbiriyle eşit olması veya eş değer olup olmamağı kontrol edilmesi gerekir. Farklı bir test formunu alan katılımcının diğer bir katılımcıdan daha zor bir test formu almasının doğuracağı sonucu engellemek ve aynı testin pek çok formu üzerinde karşılaştırılabilir puanlar elde etmek için test formlarının eşitlenmesine ihtiyaç duyulur.

Test Eşitleme

Birçok durumda, bir değerlendirmenin farklı formlarından veya farklı zamanlarda uygulanan formlar üzerinden katılımcı performanslarını karşılaştırmak arzu edilir (Muraki, Hombo ve Lee, 2000). Böyle durumlarda test eşitleme, iki veya daha fazla formdan gelen puanları uygun istatistiksel yöntemlerden geçirdikten sonra ortak bir ölçeğe yerleştirerek, yani formlardan elde edilen puanları birbiri yerine kullanılmasını sağlayarak, bunun sonucunda da karşılaştırma yapmaya olanak vererek tatmin edici cevaplar sunabilir.

Eşitleme işlemi, farklı formlardan elde edilen puanların birbirleriyle karşılaştırmaya ihtiyaç duyulduğunda gereklidir (Meng, 2012). Test eşitleme yapılmadan, test formları güçlük bakımından farklılık gösteriyorsa bir uygulamadaki katılımcılar diğer bir uygulamadaki katılımcılara göre adaletsiz bir şekilde avantajlı ya da dezavantajlı olabilirler (Hagge, 2010).

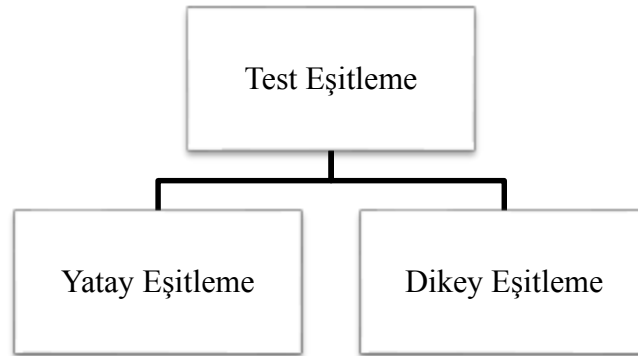
Test eşitlemede odak noktası farklı test formlarının ölçekleri arasındaki ilişkiyi bulma problemidir (Bränberg, 2010). Angoff (1984), test formlarının eşitlenmesini, bir formun

birim sistemini diğer formun birim sistemine dönüştürmek olarak tanımlamış ve dönüştürmeden sonra iki formdan gelen puanların direkt eşdeğer olacağını belirtmiştir. Kolen ve Brennan (2014) ise eşitlemeyi istatistiksel ve içerik olarak benzer geliştirilen test formları arasındaki farklılıkları düzenleyen ve formlardan elde edilen puanların birbiri yerine kullanılmasını sağlayan istatistiksel bir süreç olduğunu belirtmişlerdir. Braun ve Holland (1982) güçlük düzeyi farklılık gösteren formlardan elde edilen puanların birbiri yerine kullanılabilmesini sağlama için yapılan sayısal düzenlemeler olarak tanımlamışlardır. Haertel (2004) 'e göre test eşitleme ortak test formunun bulunduğu diğer alternatif formdaki puan ile puan eşleştirme fonksiyonunu türetme sürecidir, öyle ki eşitlemeden sonra hangi test formunun uygulandığına bağlı olmaksızın verilen herhangi ölçek puanı aynı anlama sahip olur.

Test eşitlemeyle, aynı amaç için hazırlanmış test formlarının güçlüklerinden kaynaklanan farklılıkların giderilmesi sonucunda öğrencilerin daha kolay veya daha zor bir test formunu almış olmasından dolayı bir avantaj veya dezavantaj sağlaması önlenmiş olunur (Tan, 2015). En genel tanımıyla eşitleme, aynı yapıyı ölçen testler arasındaki güçlülük farklarını düzenleyerek elde edilen puanların birbiri ile karşılaştırılmasını sağlayan sürece verilen addır.

Yatay ve Dikey Eşitleme

Bir test geliştiricinin amacı eşit güçlükte alternatif formlar oluşturabilmek olsa da çoğu zaman formlar farklı güçlükte olurlar (Dvorak, 2009). Alternatif test formundaki bu farklılığı ayarlamak için test eşitleme yöntemleri geliştirilmiştir. İki veya daha fazla test formunun puanlarını eşitlemede, testlerin zorluk düzeyleri bakımından sınıflandırılması Şekil 1'de gösterilmiştir (Cook ve Eignor 1985, Crocker ve Algina, 1986; Hambleton ve Swaminathan, 1985).



Şekil 1. Test eşitlemenin testlerin zorluk düzeyleri bakımından sınıflandırılması

Yatay eşitleme (horizontal), aynı yapıyı ölçen ve güçlük düzeyleri birbirine eşit ya da benzer olan iki farklı form puanları arasında yapılan eşitlemedir. Yatay eşitlemede, testlerin eşit ya da benzer güçlükte olması ve testi alan grupların benzer yetenek düzeyinde olması beklenir. Örneğin ÖBBS (Öğrenci Başarısını Belirleme Sınavı) gibi yapılan sınavlarda güvenlik sebebiyle bir testin farklı formlarının aynı sınıf düzeyindeki öğrencilere uygulanması sonucu her bir formu alan gruplardaki öğrenci puanlarının birbirleriyle karşılaştırılabilmesi için yatay eşitlemeye başvurulabilir. TIMSS (Trends in International Mathematics and Science Studies), PISA (Programme for International Student Assessment), PIRLS (Progress in International Reading Literacy Study) gibi geniş ölçekli sınavlarda da test gizliliği ve madde kalibrasyonu gibi çalışmalar açısından aynı amaçla oluşturulan farklı test formları kullanılmaktadır. Bu uygulamalarda benzer güçlükteki test formlarını alan farklı bireylerin puanları yatay eşitleme ile birbirleriyle karşılaştırılabilir.

Dikey eşitleme (vertical) ise, aynı yapıyı ölçen, farklı yetenek düzeylerindeki bireylere uygulanan ve güçlük düzeyleri farklı olan iki veya daha fazla test puanları arasında yapılan eşitlemelerdir (Aiken, 2000) . Farklı düzeydeki bireylere uygulanan testler, özel bir testin çoklu formları değildir ve genel olarak güçlüğü farklı düzeyde olan testlerdir (Hambleton ve Swaminathan, 1985). Örneğin Milli Eğitim Bakanlığı'nın 2007 ve 2010 yılları arasında liselere geçiş için getirilen, 6., 7.,ve 8. sınıflarda yapılan sınavların puanlarının belli yüzdeleri alınarak liseye geçişte etki eden SBS (Seviye Belirleme Sınavı)'ye ait matematik alt testlerinin tamamı matematik başarısını ölçer ancak testler farklı güçlük düzeylerindedir. Burada dikey eşitleme tek bir ölçek üzerinde farklı sınıf düzeylerindeki öğrencilerin yeteneklerinin karşılaştırılmasına izin verir. Dahası yatay eşitlemenin tersine, testleri alan bireylerin puanlarının dağılımı birbirinden farklı olacaktır (Hambleton ve Swaminathan, 1985). Bu aşamadan sonra bu çalışma kapsamında test eşitleme demekle yatay eşitleme kast edilmektedir.

Test Eşitlemenin Koşulları

Dorans ve Holland (2000) test eşitlemenin başarılı sonuçlar verebilmesi için beş varsayımın sağlanmasının gerekli olduğunu özetlemiştir (aktaran Dorans, Pommerich ve Holland, 2007). Aşağıdaki varsayımlar literatürdeki önemlilik derecesine göre kabaca sıralanmıştır (Dorans vd, 2007).

- a) Aynı yapıyı ölçme özelliği: Testlerin her ikisi aynı yapıyı ölçmelidir (gizil özellik yetenek, beceri).

- b) Eşit güvenilirlik düzeyleri özelliği: Testlerin ölçümlerinin güvenilirliğinin aynı düzeyde olması gerekir.
- c) Simetri özelliği: Eşitleme dönüşümünün simetrik olmasını gerektiren simetri özelliği (Lord, 1980) eşitleme yapılacak bir formdan alınan bir puanı diğer bir formdaki puana dönüştürmede kullanılan fonksiyon, diğer bir formdan alınan bir puanı önceki forma dönüştürmek için kullanılan fonksiyonun tersinin alınmasını gerektirir. Örneğin bu özellik ile eşitleme denkleminde X formundaki 26 ham puanını Y formundaki 27 ham puana dönüştürüyor ise Y formundaki 27 ham puanını X formundaki 26 ham puana dönüştürmesi gerektiği anlamına gelir. Bu simetri özelliği regresyon denklemi gibi değildir (Kolen ve Brennan, 2014). Çünkü regresyonda, iki değişkenden hangisinin bağımsız, hangisinin bağımlı olduğuna göre regresyon katsayısı ve kesim noktası değeri değişir.
- d) Eşitlik özelliği: Eşitlenmiş testlerde, bireylerin iki testten birini almış olması ilgilenmeye değmeyen bir durum olmalıdır.
- e) Gruplar arası değişmezlik özelliği: Test eşitleme sürecinde test eşitleme fonksiyonunun kestirimi için seçilen örneklem grubunun eşitlemeden bağımsız olmasını gerektirir. Bununla birlikte test eşitleme süreci sonunda elde edilen eşitleme denkleminin bulunduğu örneklemde bağımsız olarak diğer örneklem için de aynı olması beklenir.

Bazı araştırmacılara göre bu koşulların bazıları diğerlerine göre daha fazla veya daha az önem taşımakta veya bazılarının gereksinim duymamaktadırlar. Hambleton ve Swaminathan (1985) eşitleme koşullarının; eşitlik, simetri, gruplar arası değişmezlik ve tek boyutluluk olduğunu belirtmiştir. Lord (1980) da benzer şekilde; eşitleme için aynı yapıyı ölçme özelliği, eşitlik özelliği, simetri özelliği ve gruplar arası değişmezlik özelliklerini eşitlemenin varsayımları olarak sıralamıştır. Angoff (1984) ise eşitleme koşulları için her iki testte yer alan maddelerin aynı özelliği ölçmesi ve dönüştürme sonrasında tüm durumlarda gözlenen puanların birbiri yerine kullanılabilmesine işaret etmiştir.

Test eşitlemenin koşullarına bakıldıktan sonra eşitleme çalışması uygun görülüyorsa eşitleme sürecine geçilebilir. Çalışmanın bu kısmında sürecin aşamalarından bahsedilmiştir.

Test Eşitleme Aşamaları

Eşitleme uygulaması için adımların listesi aşağıdadır (Kolen ve Brennan, 2004).

1. Eşitlemenin amacına karar verme.
2. Alternatif test formları oluşturma, içerik ve istatistiksel özellikleri benzer formlar oluşturulur.
3. Veri toplama desenini seçme. Eşitleme, test formlarının istatistiksel olarak nasıl farklılaştığına yönelik bilgi sağlama için veri toplamayı gerektirir.
4. Veri toplama deseni doğrultusunda uygulanma yapılması. Desen tarafından belirlendiği gibi test uygulanır ve veri toplanır.
5. Eşitliğin bir veya daha fazla işlemsel tanımını seçme. Eşitleme formları arasında ne tür ilişkilerin olacağına yönelik tercihlerin seçimini gerektirir.
6. Bir veya daha fazla istatistiksel tahmin yöntemi seçme.
7. Eşitleme sonuçlarını değerlendirme.

Test eşitleme çalışmalarında eşitlemenin amacına karar verilip eşitlenecek formlar belirlendikten sonra veri toplamak için nasıl bir yol izleneceğine karar verilir. Eşitlemenin bu aşaması eşitleme deseninin seçimini gerektirir.

Eşitleme Desenleri

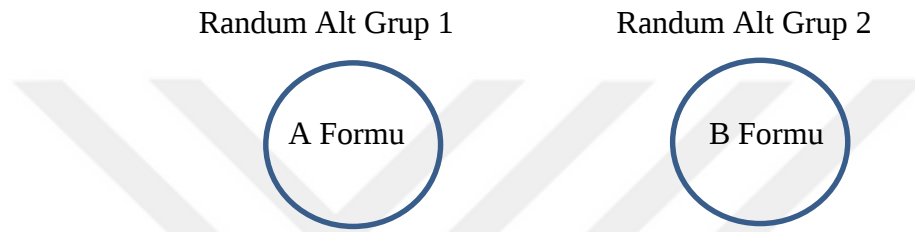
Veri toplama, eşitlemedeki iyi uygulamaların en önemli unsurlarından biridir (Dorans, Moses ve Eignor, 2010). Eşitlemede farklı veri toplama yöntemlerine “eşitleme deseni” denilebilir. Desen seçimi uygulamayı ve istatistiksel durumun her ikisini kapsar (Kolen ve Brennan, 2014). Farklı testleri alan katılımcıların puanlarını eşitleme veya maddeleri ölçeklendirme yalnızca belli koşullar altında yapılır, örneğin iki farklı gruba uygulanmış iki farklı test eşitlenemez (Hambleton ve Swaminathan, 1985).

Dört temel eşitleme deseni vardır.

Random Grup Deseni (Random Group Design)

Random grup deseninde uygulanacak olan formlar katılımcılara random olarak dağıtılır. Formlar, random olarak oturma düzeninde olan katılımcılara spiral (sarmal) yöntemle dağıtılmalıdır (Kolen ve Brennan, 2014). Örneğin iki form uygulaması yapılacak bir uygulamada ilk kişiye A formu veriliyorsa ikinci kişiye B formu üçüncü kişiye A formu verilerek ve bu spiral metot izlenerek devam edilmelidir. Çünkü bu spiral metot genel

anlamda karşılaştırma yapabilmeyi sağlar, random olarak eşdeğer gruplar A formunu ve B formunu alırlar (Kolen ve Brennan, 2014). Sarmal yöntemle form uygulamalarında katılımcıların sistematik olarak oturmaları engellenmelidir. Örneğin bir kız bir erkek şeklinde oturan bir sınıf düzeninde spiral metotla form dağıtıldığında A formunun tamamı kızlara, B formunun da tamamı erkeklere ya da A formunun tamamı erkeklere, B formunun da tamamı kızlara denk gelmiş olabilir. Bu yüzden de desenin random gruplar olmasının önüne geçilmiş olur. Şekil 2’de random grup desenine dayalı olarak yapılan uygulama özetlenmiştir.



Şekil 2. Random grup deseni

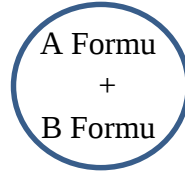
Bu desende iki form arasındaki fark, formların arasındaki güçlük farkının doğrudan göstergesi olarak alınır. Örneğin A formu ile B formunu eşitlemek için büyük bir grup katılımcıdan random grup deseni kullanılarak veriler toplansın. A formunu alan katılımcıların ham puan ortalamaları 86, B formunu alan katılımcıların ham puan ortalamaları ise 80 olsun. Burada A formu için hesaplanan ortalama B formu için hesaplanan ortalamadan 6 puan daha fazla olduğundan A formu, B formundan 6 puan daha kolaydır.

Random grup deseninin avantajları tüm formların aynı anda uygulanması ve böylece de zamandan tasarruf edilmesidir. Dezavantajı ise büyük örneklem gerektirmesi ve random grup oluşturmadaki hatalar yüzünden desenin varsayımının bozulmasıdır.

Tek Grup Deseni (Single Group Design)

Tek grup deseni en basit veri toplama desendir. Tek grup deseninde bir evrenden gelen tek bir örneklem her iki testi de alır. Bu desen göreceli olarak daha küçük örneklemle doğru eşitleme sonuçları sağlayabilir (Dorans vd, 2010). Şekil 3’te gösterildiği gibi; tek grup deseni, bir gruba iki formunda uygulanmasını içerir.

Grup

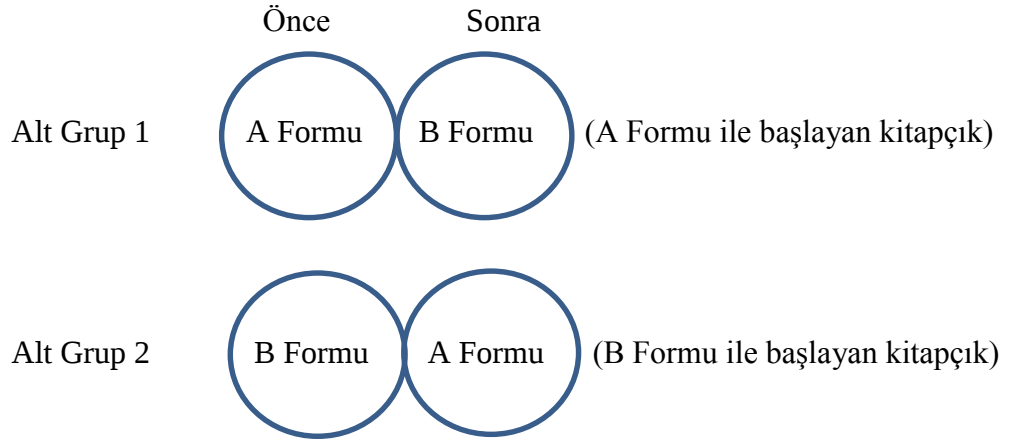


Şekil 3. Tek grup deseni

Her iki form tek bir gruba uygulandığı için, sıra etkileri kontrol edildiği sürece, puanlar arasındaki farklılıklar form farklılıklarına atfedilebilir (Kim, 2014). Tek grup deseninin avantajları, aynı kişilerin aynı anda testlerin tamamını almalarından dolayı yalnızca bir gruba ihtiyaç duyulması ve aynı kişilerin her iki testi de alması nedeniyle testlerin güçlüklerinin kişilerin yetenek düzeylerinden etkilenmemesinden dolayı testi alan bu grubun hedef evreni temsil etmesine gerek olmayışdır. Dezavantajı ise kişilere birden fazla test verilmesinden kaynaklı zaman problemi ve ikinci alınan formun birinci formdan etkilenebilmesidir.

Dengelenmiş Tek Grup Deseni (Single Group Design with counterbalancing)

Bu desen tek grup deseninde ortaya çıkan sıra etkisini giderebilmek için geliştirilmiştir. Çünkü sıra etkisi, eşitlenmek istenen formlar gerçekten iki ayrı formda tek grup deseninin kullanılmasını elverişsiz kılar (Kolen ve Brennan, 2014). Bu desende dengeleme, test kitapçıkları ile sağlanır. Test kitapçıklarının yarısı A, yarısı B formunu içerir. Katılımcılar random olarak ikiye bölünür ve bunlar alt grupları oluşturur. Alt grupların birindeki bir birey önce A sonra B formunu (A Formu ile başlayan kitapçık) alırken diğer alt grubundaki bir birey önce B sonra A formunu alır (B Formu ile başlayan kitapçık). Bu süreç A formunu ilk alan gruba, B formunu ilk alan grubun karşılaştırılmasını sağlar (Kolen ve Brennan, 2014). Şekil 4'te dengelenmiş tek grup deseni özetlenmiştir.



Şekil 4. Dengelenmiş tek grup deseni

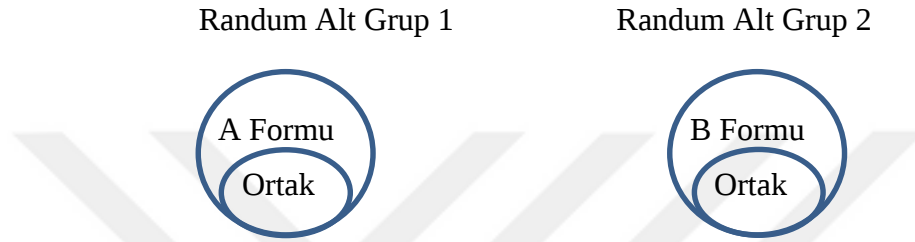
Bu desenin kullanımında sıra etkisinin bertaraf edilmesi ve bir katılımcının iki formu da cevaplamasından ötürü random grup desenine göre iki kat fazla veri toplanması avantaj olarak görülürken, uzun formların eşitlenmesinde sorunlar yaşanabilmesi dezavantaj olarak görülebilir. Özellikle denek sayısı az olduğunda ve test formları çok uzun olmadığında dengelenmiş tek grup deseninin kullanımı önerilmektedir (Tan, 2015).

Ortak Maddeli Eşdeğer Olmayan Grup Deseni (Common Item Nonequivalent Groups Design)

Bu desende A Formu ve B Formu ortak madde seti içerir. Bu iki form farklı gruptaki katılımcılara uygulanır. Yani bir gruba ortak madde içeren A Formu, diğer bir gruba da yine ortak madde seti içeren B Formu uygulanır. Uygulama sonunda test puanlarındaki farklılıklar form farklılıkları ve grup farklılıklarının bir kombinasyonunun sonucu olabilir (Kim, 2014). Form ve grup etkilerinden kurtulmak için, orantılı içerik ve istatistiksel özellikleri açısından toplam test formları temsil eden ortak madde seti iki alternatif formun içine gömülürler (Livingston, 2004).

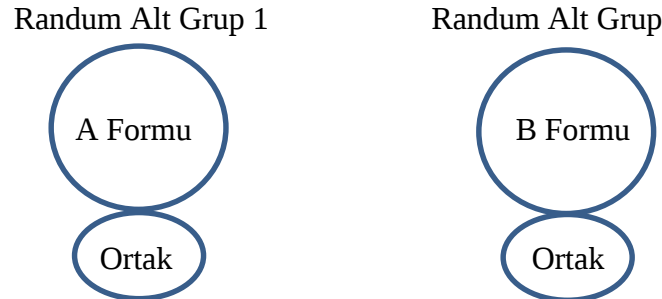
Bu desenin iki çeşidi vardır (Kolen ve Brennan, 2014). Formların içine gömülü olan ortak madde setinden alınan puan bireyin test puanına katkı sağlıyorsa bu desen iç ortak maddeli (internal) eşdeğer olmayan grup deseni olarak adlandırılır; katkı sağlamıyor yani ortak maddeden alınan puan bireyin test puanını etkilemiyorsa dış ortak maddeli (external) eşdeğer olmayan grup deseni olarak adlandırılır.

Şekil 5’te iç ortak maddeli eşdeğer olmayan grup deseninin şemalaştırılmış hali vardır. İç ortak maddeli eşdeğer olmayan grup deseninde ortak maddeler puanlamaya dâhil olurlar. Örneğin her biri bir puan değerinde olan, 5 maddesi ortak 20 soruluk formlardan birinde tamamını doğru yapan bireyin puanı 20 dir. Livingston (2004), iç ortak maddeli eşdeğer olmayan grup deseninin dezavantajlarını; ortak maddelerin test formunun içeriğine göre özenle seçilmesi gerektiği ve referans grubun (farklı tarihlerde yapılan sınavlarda eski formun uygulandığı grup) maddeleri açığa vurduğunda formun kullanılamaz olması olarak sıralamıştır.



Şekil 5. İç ortak maddeli (internal) eşdeğer olmayan grup deseni

Şekil 6’ da dış ortak maddeli eşdeğer olmayan grup deseninin şemalaştırılmış hali vardır. Dış ortak maddeli eşdeğer olmayan grup deseninde ortak maddeler puanlamaya dâhil olmazlar. Örneğin her biri bir puan değerinde olan, 5 maddesi ortak 20 soruluk formlardan birinde ortak maddelerden ikisini yanlış yapan ancak diğer maddeleri doğru yapan bir bireyin puanı da formdaki tüm soruları doğru yapan bir bireyin puanı da 15 olur. Pratikte karşılaşılabilecek bu durumun KTK’da dikkate alınması gerekebilir. Livingston (2004), dış ortak maddeli eşdeğer olmayan grup deseninin dezavantajlarını; eşitlenecek testlerle aynı beceriyi ölçen ortak maddelerin bulunmasının zor olması ve ortak maddelerin ek süre gerektirebilecek olması olarak sıralamıştır.



Şekil 6. Dış ortak maddeli (external) eşdeğer olmayan grup deseni

Ortak maddeli eşdeğer olmayan grup deseni büyük idarî esneklik sağlar ve muhtemelen, uygulamada en yaygın desen olarak kullanılır (Cook, 2007). Ancak bu desen uygulaması en

zor olan desenlerden biridir (Cook, 2007). Farklı test veya test formlarını alan grupların eşdeğer olduğu kabul edilmez. Ortak madde setinin asıl amacı; grup farklılıklarını form farklılıklarından ayırmaktır (Tian, 2011).

Gruplardan gelen farklılığı doğru olarak yansıtmak için, ortak madde seti test formlarının bütününe yapısal ve istatistiksel özelliklerini orantılı olarak temsil etmesi iyi olur (Kolen ve Brennan, 2014). Formlarda bulunan ortak madde seti test formunun bütününe mini versiyonu olmalıdır.

Ortak maddeler tam anlamıyla birbirinin aynısı olmalıdır (Kolen ve Brennan, 2014). Ortak maddelerde kullanılan kelimeler kaldırılmamalı veya alternatifleriyle değiştirilmemelidir. Ek olarak Kolen ve Brennan (2014), ortak madde setinin formlarda aynı yerde (madde numaraları aynı) olmasının benzer davranışlar sağladığını belirtmiştir. Ancak bu çalışmadaki iki formun ortak madde seti her ne kadar kendi içinde aynı sıralamada olsa da ilk formun son on iki sorusu ile ikinci formun ilk on iki sorusu olarak konumlandırılmıştır.

Bu desende iki gruptan elde edilen ortalama ve standart sapmaların farklı olmalarının bir önemi yoktur. Yani grupların eşdeğer olmalarının önemi yoktur. Burada eşitleme çalışmalarının amacı her iki testten elde edilen madde parametrelerini ortak bir ölçeğe yerleştirmek ve böylelikle her iki testten elde edilen puanları karşılaştırmaktır (Embreston ve Reise, 2013).

Örneklem genişliği büyük ve formlar, güvenlik problemleri olmadan yeniden kullanılabilir olması için, her bir katılımcının her iki formu da aldığı tek grup desende ortaya çıkabilen sıra etkisi sorununun oluşmasını önlediğinden ortak maddeli eşdeğer olmayan grup deseni genellikle iyi bir seçim olarak kabul edilir. Son yıllarda yapılan eşitleme çalışmalarında bu desen sıklıkla kullanılmıştır (Hagge, 2010; Hagge vd, 2011; He, 2011; Jurich, DeMars ve Goodman, 2012; Kim, 2014; Meng, 2012; Tian, 2011; Wang, 2013; Wolf, 2013). Peterson (2014)'a göre de tek bir sınav tarihinde birden fazla test formu uygulanamadığında da ortak maddeli eşdeğer olmayan grup deseni en iyi seçenek olabilir.

Çalışmanın devam eden kısmında geleneksel test eşitleme yöntemleri ve Madde Tepki Kuramı ile eşitleme yöntemlerinden bahsedilecektir.

Klasik Test Kuramı ve Geleneksel Test Eşitleme Yöntemleri

Geleneksel test eşitleme yöntemi Klasik Test Kuramı (KTK)'na dayanır. KTK; gözlenen puan, gerçek puan ve hata puanı gibi kavramlarla açıklanır (Crocker ve Algina, 1986).

$$x = \tau + e$$

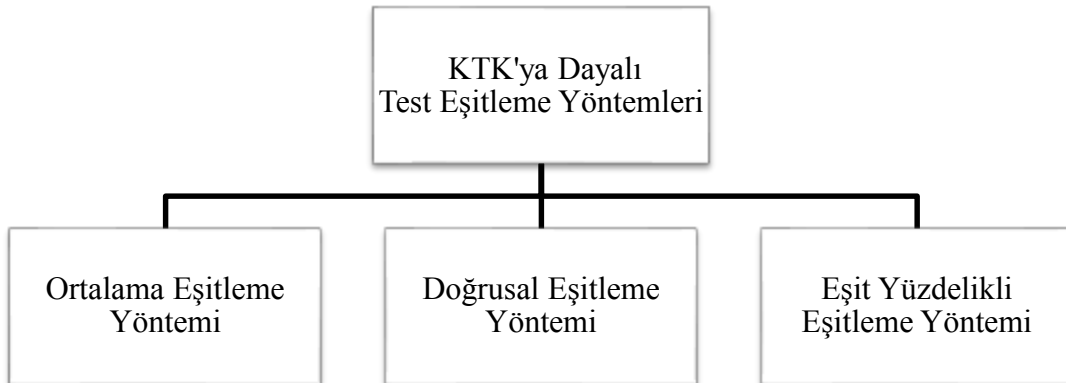
x : Gözlenen puan

τ : Gerçek puan

e : Hata

Gözlenen puan, yukarıdaki eşitlikte olduğu gibi gerçek puan ve hata puanının toplamına eşittir. Gerçek puan, bireyin ölçülen özelliğinin hatalardan arınık olan puanıdır; bireyin ölçülen özelliğinin sonsuz sayıda ölçme yapıldığında elde edilen ölçümlerin ortalaması olarak kabul edilir. Hata puanı ise randum (tesadüfi) hata olarak adlandırılabilir (Crocker ve Algina, 1986). Eşitleme çalışmaları dâhil ölçme uygulamalarında ölçmeye karışan hata ne kadar azsa sonuçlar daha doğru olur.

Eşitleme işleminde formdan forma olan farklılıkları ayarlamak için istatistiksel metotlar kullanılır (Tong ve Kolen, 2005). Geleneksel test eşitlemede bu metotlar güçlü bir puanlama modeline dayanmayan ve gözlenen puan aralığı için eşitleme ilişkisini belirleyen; ortalama eşitleme yöntemi, doğrusal eşitleme yöntemi ve eşit yüzdelikli eşitleme yöntemidir (Peterson, 2014; Tian, 2011). Şekil 7’de bu yöntemler özetlenmiştir.



Şekil 7. Geleneksel test eşitleme yöntemleri

Ortalama Eşitleme Yöntemi

Ortalama eşitlemede iki testin ortalaması belli bir grup öğrenci için eşit şekilde düzenlenir. İki test arasındaki güçlüğün tüm puan ölçeğinde sabit olduğu varsayılır. Ortalama eşitleme için aşağıdaki eşitlikten yararlanılır.

$$x - \mu(X) = y - \mu(Y)$$

x : X testindeki bir kişinin puanı,

$\mu(X)$: X testinin ortalaması,

y : Y testindeki bir kişinin puanı

$\mu(Y)$: Y testinin ortalaması.

Eşitlik y için düzenlendiğinde, $m_Y(x)$, ortalama eşitlemeyi kullanarak, Y testinin ölçeğine dönüştürülen X testi üzerindeki x puanıdır, aşağıdaki eşitlik bu ifadeyi gösterir (Kolen ve Brennan, 2014).

$$m_Y(x) = y = x - \mu(X) + \mu(Y)$$

Kısaca ortalama eşitlemede, eşitlenmiş puan dağılımları güçlüklerindeki farklılıkları ayarlamak için tüm puanlama ölçeği boyunca bir sabiti ekleyerek ya da çıkararak oluşturulan eşit ortalamalara sahiptir (Peterson, 2014).

Doğrusal Eşitleme Yöntemi

İki form arasındaki farklılığın sabit olmadığı düşünüldüğünde, doğrusal eşitleme farklı zorluktaki iki test arasında eşitlemeye izin verir (Kolen ve Brennan, 2014). Doğrusal eşitleme yöntemi güçlükteki farklılıkları puanlama ölçeği boyunca standart sapma ile değişen seviyelerde düzenlenir (Peterson, 2014). Doğrusal eşitlemede iki testin ortalaması ve standart sapması eşit şekilde düzenlenir. Eğer testlerden elde edilen standart sapmalar birbirine eşitse doğrusal eşitleme, ortalama eşitleme ile aynı eşitleme sonuçlarını verecektir. Doğrusal dönüştürmede aşağıdaki eşitlikten yararlanılır (Kolen ve Brennan, 2014).

$$\frac{x - \mu(X)}{\sigma(X)} = \frac{y - \mu(Y)}{\sigma(Y)}$$

x : X testindeki bir kişinin puanı,

$\mu(X)$: X testinin ortalaması,

y : Y testindeki bir kişinin puanı

$\mu(Y)$: Y testinin ortalaması.

$\sigma(X)$: X testinin standart sapması

$\sigma(Y)$: Y testinin standart sapması

Eşitlik y için düzenlendiğinde, $l_Y(X)$, doğrusal eşitlemede Y testinin ölçeğine dönüştürülen X testi üzerindeki x puanıdır. Aşağıda bu ifadenin eşitliği yer almaktadır.

$$l_Y(X) = y = \sigma(Y) \left[\frac{x - \mu(X)}{\sigma(X)} \right] + \mu(Y)$$

Yukarıdaki eşitlik doğrusal regresyon denklemi bağlamında yeniden düzenlendiğinde aşağıdaki gibi “eğim” ve “sabit” elde edilebilir.

$$l_Y(X) = y = \frac{\sigma(Y)}{\sigma(X)} x + \left[\mu(Y) - \frac{\sigma(Y)}{\sigma(X)} \mu(X) \right]$$

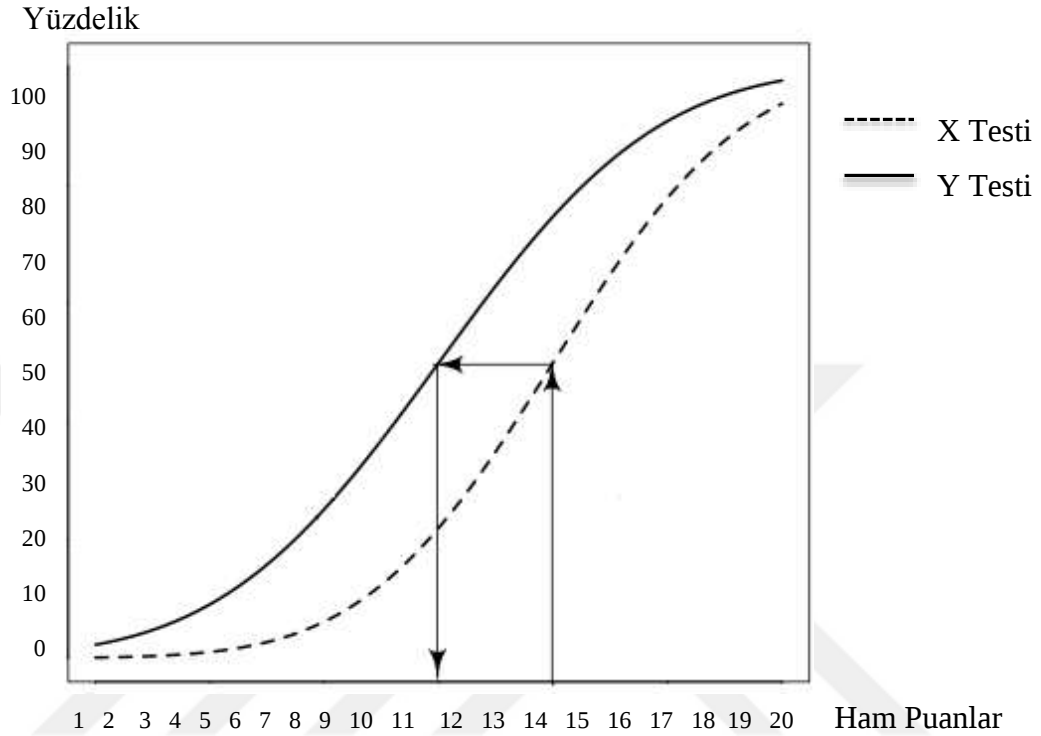
Kolen and Brennan (2014), ortalama eşitleme sürecinin doğrusal eşitleme sürecinden daha az karmaşık olduğunu belirtmiştir. Ancak doğrusal eşitleme dönüştürmede daha fazla bilgi kullanır ve daha çok esneklik sağlar.

Eşit Yüzdelikli Eşitleme Yöntemi

Eşit yüzdelikli eşitlemede, X testinin dağılımı Y formunun dağılımına eşit düşecek şekilde düzenlenmiştir. Dağılımlar eşitlenirken, ortalamanın ve standart sapmanın yanında basıklık ve kayışıklıkta eşit yüzdelikli eşitlemede kullanılır. Kayışıklık ve basıklık bağlamında testin dağılımlarının şekli aynıysa eşit yüzdelikli eşitlemenin ve doğrusal eşitlemenin sonuçları aynıdır.

Eşit yüzdelikli eşitlemede iki formda da aynı yüzdelik dilime sahip puanlar eşit kabul edilirler. Eşit yüzdelikli eşitleme süreci grafiksel metotlar kullanılarak gösterilebilir. Eşit yüzdelikli eşitlemenin ilk adımı eşitlenecek iki testin her birinde puan dağılımları için yüzdelik dilimleri belirlemektir (Peterson, 2014). Daha sonra yüzdelik dilimler iki testin her biri için ham puanlara karşı kullanılır ve yüzdelik dilim ham puan eğrileri oluşturulur (Peterson, 2014). Şekil 8’de her biri 20 maddelik iki testin (X testi ve Y testi) puanlarının

yüzdelik dilim ham puan eğrileri oluşturulmuştur. Burada grafiksel olarak puanları dönüştürmek oldukça kolaydır. Örneğin X testindeki 15 puanı, Y testindeki 12 puana denktir.



Şekil 8. Eşit yüzdelikli eşitleme

Eşit yüzdelikli eşitlemedeki esas sorun teste giren öğrencilerin verdikleri gerçek testlerdeki puan dağılımlarının genel olarak düzensiz olmasıdır. Teste giren öğrencilerin puanlarının yüzdesi puan arttıkça artmaz, dalgalanma gösterir. Puan dağılımındaki düzensizlikler Eşit yüzdelikli eşitlemede sorunlara neden olur. Eşit yüzdelikli eşitleme uyarlamasında düzensizliklere neden olur ve bu düzensizlikler var olan evrene genellenemez (Peterson, 2014). Bu yüzden puan dağılımı ve eşit yüzdelikli ilişkinin mantıklı bir şekilde olması için çok büyük bir örnekleme ihtiyaç duyulur (Kolen ve Brennan, 2004). Düzgünleştirme işlemleri de bu sorunun önüne geçebilir. Düzgünleştirme ile ilgili daha fazla bilgi için Kolen ve Brennan (2014)'ın kitabından faydalanılabilir.

Anlatılan bu üç yöntem KTK'ya dayalıdır, ancak KTK'nın pek çok sınırlılığı bildirilmiştir. KTK'da bir öğrencinin yeteneği testlerle ölçülür ve bir testteki maddelerin zorluğu maddeye doğru cevap veren aynı gruptaki öğrencilerin oranı olarak tanımlanır. (Hambleton, Swaminathan ve Rogers, 1991). Bu şu anlama gelmektedir, maddelerin güçlüğü bir dereceye kadar teste tabi tutulan öğrencilerin yeteneklerinden etkilenmektedir. Madde ayırt ediciliği

ise bir testin testte yüksek başarı gösteren öğrenci ile düşük performans gösteren öğrenciyi ne kadar iyi ayırt ettiği olarak tanımlanabilir. Madde ayırt ediciliği değeri de teste tabi tutulan öğrencilerin yeteneklerinden etkilenir. Ek olarak, testin ölçümlerinin geçerlilik ve güvenilirlik derecesi belli bir gruba özgüdür. Bundan dolayı istatistikler şişme eğilimindedir çünkü madde güçlüğü ve ayırt ediciliği grup bazlıdır. Örneğin, ayırt edicilik indeksleri ve heterojen öğrenci güvenilirlik katsayısı homojen öğrenci örneklerine göre daha yüksek olma eğilimindedir. (Hambleton, 1989). KTK'nın başka bir sınırlılığı da eşit hata varsayımıdır. KTK' da her bir öğrencinin eşit ölçme hatasına sahip olduğu varsayılır ancak bu varsayım düşük ve yüksek yetenekli öğrenci gruplarında her zaman doğru olmayabilir. Sonuç olarak KTK' da paralel testler oluşturulabilir ama gerçekte bu varsayımı karşılamak imkânsızdır. KTK' nın son bir eksikliği de öğrencinin verilmiş bir maddeye verdikleri cevabın göz önünde bulundurulmamış olmasıdır. Bu şu demektir ki KTK' da öğrencilerin ister kişi olsun ister grup verilen maddeyi yordama performanslarını yordamak imkânsızdır (Hambleton vd, 1991).

Madde Tepki Kuramı Modelleri ve Madde Tepki Kuramına Dayalı Eşitleme Yöntemleri

Araştırmalarda kullanılan değişkenlerin kontrol edilmesi için gerekli olan yöntemleri ve değişkenlerin ölçülmesinde kullanılacak araçların oluşturulması için gereken bilgiyi test kuramı sağlar (Crocker ve Algina, 1986). Zekâ, başarı, yetenek, tutum ve ilgiyi ölçme girişimleri ile psikoloji ve eğitim teknolojilerinin gelişimiyle birlikte ölçmeye yeni kuramlar kazandırılmıştır. Klasik Test Kuramı'nın sınırlılıklarını gidermek için Modern Test Teorisi'nin kullanılması da önemli gelişmelerdendir. Modern Test Teorisi literatürde “Madde Tepki Kuramı”, “Madde Yanıt Kuramı” ve “Örtük Özellikler Teorisi” olarak da adlandırılmaktadır. Bu çalışmada “Madde Tepki Kuramı (MTK)” ifadesinin kullanımı tercih edilmiştir.

Bir bakıma MTK, KTK'da tanımlanmış olan zayıflıkları çözmek için geliştirilmiş bir alternatif teoridir. Hambleton (1989) MTK kullanmanın pek çok avantajından söz etmiştir. İlk olarak öğrencilerin yeteneğini ortaya koyan bir puan üretir. İkinci olarak da test düzeyinden daha çok madde düzeyinde uygulandığından daha anlamlı bir teoridir. Üçüncü olarak KTK'daki keskin paralel test varsayımını gerektirmeyen bir teoridir. Son olarak da

test maddelerini yetenek düzeyleriyle eşleştirme olanağı verir (Hambleton, 1989). Embreston ve Reise (2013)'a göre de MTK'nın birçok avantajı vardır:

- a) Bir MTK yetenek (özellik) seviyesi kestirimi, özellikleri bilinen herhangi bir maddeden elde edilebilir.
- b) Madde özellikleri testin durumu ile direkt ilişkilendirilir.
- c) Bağımsız değişkenler, yetenek (özellik) seviyesi ve madde özellikleri, ek bilgisiz ayrı olarak tahmin edilebilir.

MTK'ya dayanan modeller bu özellikleri olmayan KTK'ya dayanan modellere kıyasla daha çok tercih edilir (Embreston ve Reise, 2013). Bir kişinin maddeye doğru cevap verme olasılığı madde karakteristik eğrisinin şekline bağlı olduğundan, kişinin yeteneği grubun yetenek dağılımından bağımsız olacaktır. Aynı şekilde madde parametrelerinin değeri de madde karakteristik eğrisinin şekline bağlı olduğundan madde parametreleri de gruptan bağımsız olarak kestirilebilir (Doğan ve Tezbaşaran, 2003; Hambleton, 1989; Lord ve Novick, 1968). MTK modelleri bir maddeye verilen yanıt ile araç tarafından ölçülen yetenek ya da özellik (θ ile gösterilir) arasındaki ilişkiyi gösterir (DeMars, 2010). Madde tepkisi; doğru-yanlış, evet-hayır, katılıyorum-katılmıyorum gibi dikotomus (ikili puanlanan- iki kategorili) veya Likert tipi ölçeklerden gelen politomus (çoklu puanlanan- iki kategoriden fazla) derecelendirme olabilir. Bu kurama göre örneklemden bağımsız madde ölçeklemesi ve maddeden bağımsız olarak bireyin yetenek ölçeklemesi matematiksel modellerle kestirim yapılabilir. Temel olarak, MTK bir bireyin ölçülen yetenek düzeyi ile maddelere verdiği yanıtlar arasındaki ilişkiyi açıklamak için matematiksel fonksiyon kullanarak açıklayan bir model sunar.

MTK altında çalışmaların yapılabilmesi için bazı varsayımların karşılanması gerekir. Bu çalışmada tek boyutlu MTK'nın varsayımlarından bahsedilmiş ve çalışmanın sonraki kısmında bu kuramın varsayımları test edilmiştir. Kuramın en temel varsayımı olan tek boyutluluk varsayımı testin tek bir gizil yeteneği ölçmesiyle ilgilidir (Speron, 2009). Tek boyutluluk, bireyin maddeye vereceği tepkisini etkileyen bir özellik olması durumudur denilebilir.

Bir diğer varsayım olan yerel bağımsızlık ise testi alan bireylerin yetenekleri göz önünde tutularak maddelere verdikleri tepkilerin birbirinden bağımsız olmasıyla ilgilidir. Yerel bağımsızlık varsayımı bir maddenin cevaplanmasının diğerine bağımlı olmadığını belirten bir yaklaşımı ima eder (Kolen ve Brennan, 2014). Yerel bağımsızlık, ölçülen özellik sabit

tutulduğunda bir bireyin her bir maddeye verdiği tepkinin istatistiksel olarak birbirinden bağımsız olması demektir (Crocker ve Algina, 1986). Bir örnek verilirse, maddeler pasajı okuma ve grafikleri içeren şekilde hazırlanırsa yerel bağımsızlık varsayımı ihmal edilmiş olur (Kolen ve Brennan, 2014).

Tek boyutluluk ve yerel bağımsızlık ve madde karakteristik eğrisinin monotonik artışı varsayımları test edildikten sonra eşitleme dâhil MTK'ya bağlı çalışmalar sürdürülebilir.

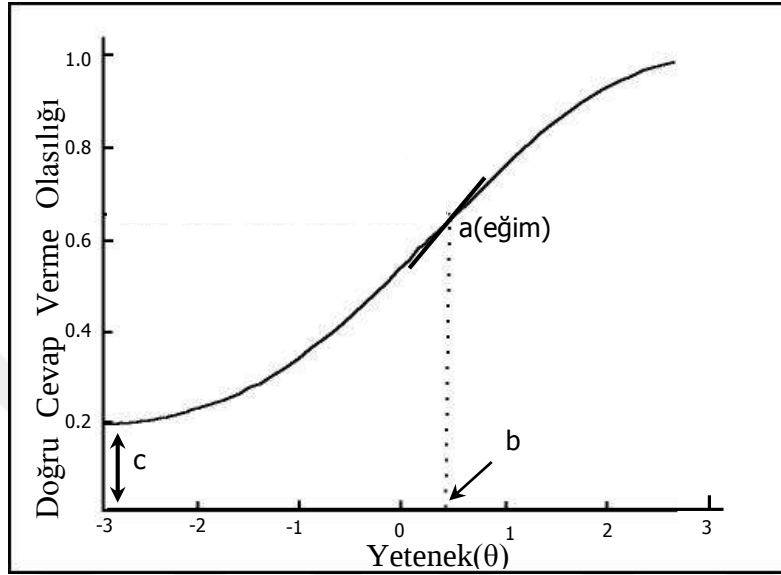
Madde Tepki Kuramı'nın ana kavramlarından biri olan "Madde Karakteristik Eğrisi" terimi 1946 yılında Ledyard Tucker'a atfedilebilir (Hambleton ve Swaminathan, 1985). Ölçülen örtük özellik tek boyutlu olduğunda, fonksiyonu belirleyen parametreler sabit parametreler olur ve fonksiyon madde karakteristik eğrisi adını alır. (Lord ve Novick, 1968). Madde karakteristik eğrisi bireyin ölçülen özelliği ile maddeye ilişkin gösterdiği performans arasındaki ilişkinin matematiksel olarak ortaya konulabileceğini varsaymaktadır.

Madde karakteristik fonksiyonuna göre tanımlanan iki model grubu vardır. Bunlar "normal ogive" ve "lojistik" gruplarıdır. Normal ogive modeli Lord tarafından geliştirilmiş ve adını birim normal dağılım fonksiyonundan almaktadır. Lojistik modeller Birnbaum tarafından geliştirilmiştir. MTK üzerinde yapılan ilk araştırmalarda "normal ogive" MKE için daha baskın olmasına rağmen, günümüzde yeri büyük ölçüde daha kolay hesaplama gerektiren üç lojistik model ile değiştirilmiştir (Crocker ve Algina, 1986). Bu modeller kullanılan parametre sayılarına göre farklılık gösterir. Dikotomus veriler kullanılılan modeller; Bir Parametrelili Lojistik Model, İki Parametrelili Lojistik Model ve Üç Parametrelili Lojistik Model 'dir. Politomus verilerde ise Samejima'nın ağırlıklandırılmış tepki modeli; Muraki'nin genelleştirilmiş kısmi puanlama modeli; Andrich'in derecelendirme ölçeği modeli vb. kullanılmaktadır (Embretson ve Reise, 2013).

Çalışmanın devam eden kısmında dikotomus veriler için üç model açıklanmaktadır. Bu modellerde yer alan madde parametrelerinden b parametresi madde güçlük indeksini belirtir ve değeri arttıkça maddenin zorlaştığını gösterir. a parametresi madde ayıricılık indeksini belirtir ve değeri arttıkça maddenin daha ayırıcı olduğunu gösterir. C parametresi ise literatürde (lower asymptote, pseudo-guessing, pseudo-chance) şans parametresi olarak adlandırılabilen bir indekstir ve değeri arttıkça yetenek düzeyi düşük olan bireylerin ($\theta = -\infty$) maddeye doğru cevap verme olasılığının arttığını gösterir (He, 2011).

MTK'da madde karakteristikleri grafiksel bir madde karakteristik eğrisi (MKE) aracılığıyla gösterilebilir (Peterson, 2014). MTK'nın birçok uygulamasında MKE'nin "S biçimi (eğik

S)” nde olduđu varsayılır (Crocker ve Algina, 1986). MKE tanımında yetenek ölçeđi, gizil özelliktir ve çoğunlukla Yunan harfi θ ile gösterilir. MKE'nin bir avantajı, madde karakteristiđini aynı anda bütün θ ölçeđi ile iliřkisini tasvir ediyor olmasıdır (Peterson, 2014). řekil 9’da bir MKE örneđi bulunmaktadır (3 parametrelı lojistik model için).



řekil 9. 3PLM için madde karakteristik eđrisi örneđi

Bir Parametrelı Lojistik Model (1PLM)

Bir madde parametresi içerdıđinden dolayı bir parametrelı lojistik model olarak bilinir. Sadece madde güçlük indeksleri (b) madde karakteristik eđrisinde incelenir. 1PLM’de ortalama bir ayırıcılık deđeri hesaplanır ve bütün maddelerin eřit madde ayırıcılık indekslerine sahip olduđu kabul edilir ve 2PLM’nin özel bir durumu olarak görülebilir (Algina ve Crocker, 1986). Rasch Modeli, 1PLM’ye matematiksel olarak eřittir fakat ayrı bir řekilde geliřtirilmiř ve 2PLM’nin özel bir durumu olarak belirtilmemiřtir. Rasch Modeli’nde ayırıcılık parametresi “1” kabul edilir.

Ařađıda 1PLM’ye ait j kiřisinin i maddesine dođru cevap verme olasılıđının matematiksel gösterimi verilmiřtir (Crocker ve Algina, 1986).

$$P(\theta_j) = \frac{e^{D(\theta_j - b_i)}}{1 + e^{D(\theta_j - b_i)}}$$

θ_j : j kiřisinin yetenek düzeyi

b_i : i maddesinin güçlüğü

D : 1,7 değerindeki sabit sayı

e : 2,718 değerindeki sabit sayı

Lord ve Novick (1968) formüllerdeki sabit değerin 1,7 olarak alınmasının geleneksel haline gelmesinin sebebinin, normal ve lojistik modellerin sayısal olan detaylarının arasındaki uyuşmayı maksimize edilmesi olduğunu belirtmiştir.

İki Parametrelili Lojistik Model (2PLM)

Bu modelde, maddeyi doğru cevaplandırma olasılığının hesaplanması için madde güçlük indeksinin yanı sıra, 1PLM modelden farklı olarak, madde ayırt edicilik indeksi kullanılır. Şans parametresi 1PLM modelde olduğu gibi sıfırdır (Hambleton ve Swaminathan 1985). Aşağıda 2PLM modele ait j kişinin i maddesine doğru cevap verme olasılığının matematiksel gösterimi verilmiştir (Crocker ve Algina, 1986).

$$P(\theta_j) = \frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)}}$$

θ_j : j kişinin yetenek düzeyi

a_i : i maddesinin ayırt ediciliği

b_i : i maddesinin güçlüğü

D : 1,7 değerindeki sabit sayı

e : 2,718 değerindeki sabit sayı

Bir MTK modeline madde ayırtıcılığı eklendiğinde, yetenek düzeyi kestirimleri madde setinde özel bir başarı ve başarısızlık örüntüsüne bağlı olur; toplam puanları aynı olan iki cevap örüntüsü farklı yetenek düzeyi kestirimlere sahip olabilir. (Embretson ve Reise, 2013). Örneğin 25 sorudan 15 soruyu doğru cevaplayan iki birey ele alalım. Bu bireylerden ilkinin ayırtıcılığı yüksek olan maddeleri doğru yaptığı, diğerinin de madde ayırtıcılığı daha düşük olan maddeleri doğru yaptığı kabul edilsin. Ölçülen özellik bakımından ilk bireyin yeteneği diğer bireyin yeteneğinden daha yüksek olarak kestirilir. 2PLM’de, yetenek düzeyi kestiriminde, maddeler eşit ağırlıklara sahip değildirler (Embretson ve Reise, 2013). Bir başka deyişle maddelerin, ölçülen özellik ile eşit seviyede ilişkisi bulunmuyorsa 2PLM’nin kullanılması uygundur.

Üç Parametrelili Lojistik Model (3PLM)

Çoktan seçmeli sınavlarda katılımcıların şans ile maddeyi doğru cevaplaması olası bir durumdur. Böylece doğru cevap olasılığı şanstın dolaylı küçük bir bileşeni içerir (Baker, 2001). Çok seçenekli veya doğru-yanlış tipi maddelerden elde edilen veriye normal ogive modeli, bir ve iki parametrelili lojistik modelleri uygulamada problem çıkabilir (Crocker ve Algina, 1986). Bu problem şöyle açıklanmaktadır: θ düzeyi küçüldükçe $P(\theta)$ 'nın sıfıra yaklaşması gerekir, oysaki çok düşük yetenek düzeyindeki bir birey, doğru cevabı tahmin yoluyla bulabileceğinden bireyin maddeye doğru cevap verme olasılığı sıfırdan büyük çıkacaktır. Bu olasılığa izin vermek için üç parametrelili lojistik model kullanılabilir (Crocker ve Algina, 1986). 3PLM'de (Birnbaum, 1968) madde güçlüğü indeksi (b) ve madde ayırtıcılığı indeksi (a) yanında şans parametresi (c) de vardır. c parametresi maddeyi yalnızca şans yolu ile doğru bulma olasılığıdır. Baker (2001), şans parametresinin tanımını yaparken c değerinin yetenek düzeylerinin fonksiyonu gibi değişmediğinin önemli olduğunu belirtmiştir. Yani düşük veya yüksek yetenekli katılımcıların her birinde maddeyi şans ile doğru cevaplandırma olasılığı aynı değildir.

Aşağıda 3PLM'ye ait j kişinin i maddesine doğru cevap verme olasılığının matematiksel gösterimi verilmiştir (Crocker ve Algina, 1986).

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

θ_j : j kişinin yetenek düzeyi

a_i : i maddesinin ayırt ediciliği

b_i : i maddesinin güçlüğü

c_i : i maddesi için şans başarısı

D : 1,7 değerindeki sabit sayı

e : 2,718 değerindeki sabit sayı

Bundan sonraki devam eden kısımda, test eşitlemede MTK' ya dayalı ölçek dönüştürme yöntemlerinden bahsedilecektir.

Baker ve Al Karni (1991)'de Madde Tepki Kuramı kapsamında testlerini eşitlemek amacıyla uygun bir lineer dönüşümün eğimi ve kesişim katsayıları edinilmesi gerektiğini belirtmiştir. Dönüştürülmüş parametre kestirimleri, bazen “kalibre edilmiş” olarak da ifade edilebilen,

Form X ve Form Y'den elde edilen ham puanlar arasındaki denklikleri kurmada ve daha sonra da puanları ölçeklemede kullanılabilir (Kolen ve Brennan, 2014). Yani dönüşüm ile parametre kestirimleri yapılarak gruplardan elde edilen parametreler ve parametre kestirimlerini aynı ölçeğe yerleştirme süreci kalibrasyon ya da ölçek dönüştürme olarak adlandırılabilir. Bu çalışmada kalibrasyon ile parametre kestirim yöntemleri ölçek dönüştürme olarak değerlendirilmiştir. Ölçek dönüştürme için kullanılan eşitlikte katsayılar çalışmanın ilerleyen kısmında A ve B olarak değerlendirilecektir.

Ölçek Dönüştürme Yöntemleri (Kalibrasyon)

MTK'na dayalı olarak yapılan eşitlemelerde model seçildikten sonra (dikotomus veriler için tek boyutlu lojistik modellerden 1PLM, 2PLM veya 3PLM seçimi gibi) madde ve yetenek parametreleri kestirilir. Parametre kestirimleri MTK'ya dayalı eşitleme yapılmadan önce iki formun parametrelerinin kıyaslanması için aynı ölçek üzerinde olması gerekir (Lee ve Ban, 2010). Yani ölçek dönüştürme, iki forma ait parametreler aynı ölçek üzerinde olmadığında eşitleme yapmak için bir önkoşuldur (Kolen ve Brennan, 2014). Tek grup deseni, dengelenmiş tek grup deseni kullanılarak eşitleme yapıldığında madde parametreleri ve yetenek parametreleri aynı ölçek üzerinde olduğundan ölçek dönüştürmeye gerek yoktur. Ortak maddeli eşdeğer olmayan gruplar deseni ile eşitleme yönteminde ise, formları alan gruplar farklı olduğundan 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 ve Brennan'dan aktaran Gök ve Kelecioğlu, 2014). Burada ölçek dönüştürme ortak maddeler sayesinde yapılmaktadır.

Test eşitlemede MTK'ya dayalı ölçek dönüştürme yöntemleri parametrelerin kestirimlerine göre farklılık gösterir. En popüler iki metot, eş zamanlı (concurrent) ve parametrelerin ayrı kestirildiği ayrı kestirim (separate) yöntemleridir (Hanson ve Béguin, 2002). Eş zamanlı ölçek dönüştürme yönteminde ortak maddeli iki formun madde parametreleri birlikte kestirilir. Ayrı ölçek dönüştürme yönteminde ise iki formun madde parametreleri ayrı ayrı kestirilir. Bu yöntemden elde edilen parametreler farklı ölçek üzerinde gösterilemeyeceğinden, madde parametrelerini aynı ölçek üzerine koyacak bir lineer denklem elde edilir.

Her iki formdan alınan puanların a ve b parametrelerine dayalı olarak aynı ölçekte olmasının sağlayan bir lineer denklem elde edilebilir. Form J ve Form I'nın lineer bir fonksiyonla

farklılaşan 3 PLM ölçeği olarak tanımlanırsa iki ölçekteki θ değerleri aşağıdaki gibi ilişkilidir (Kolen ve Brennan, 2014).

$$\theta_{ji} = A\theta_{ii} + B,$$

θ_{ji} : i bireyinin form J'deki yetenek düzeyi

θ_{ii} : i bireyinin almadığı form I'daki kuramsal yetenek düzeyi

A: Dönüştürme denkleminin eğimi

B: Dönüştürme denkleminin sabiti

Yukarıdaki lineer dönüşümle Form I'daki i kişisi için θ değerinin almadığı form J'deki karşılığı bu denklem ile bulunabilir. Bu dönüşüme benzer şekilde madde parametreleri de doğrusal dönüştürülür. Bu madde parametrelerinin dönüşümü şöyledir:

$$a_{jj} = \frac{a_{ij}}{A}$$

$$b_{jj} = Ab_{ij} + B$$

$$c_{jj} = c_{ij}$$

a_{jj} : Form J'deki j maddenin ayıricılık parametresi

a_{ij} : Form I'daki j maddenin ayıricılık parametresi

b_{jj} : Form J'deki j maddenin güçlük parametresi

b_{ij} : Form I'daki j maddenin güçlük parametresi

c_{jj} : Form J'deki j maddenin şans parametresi

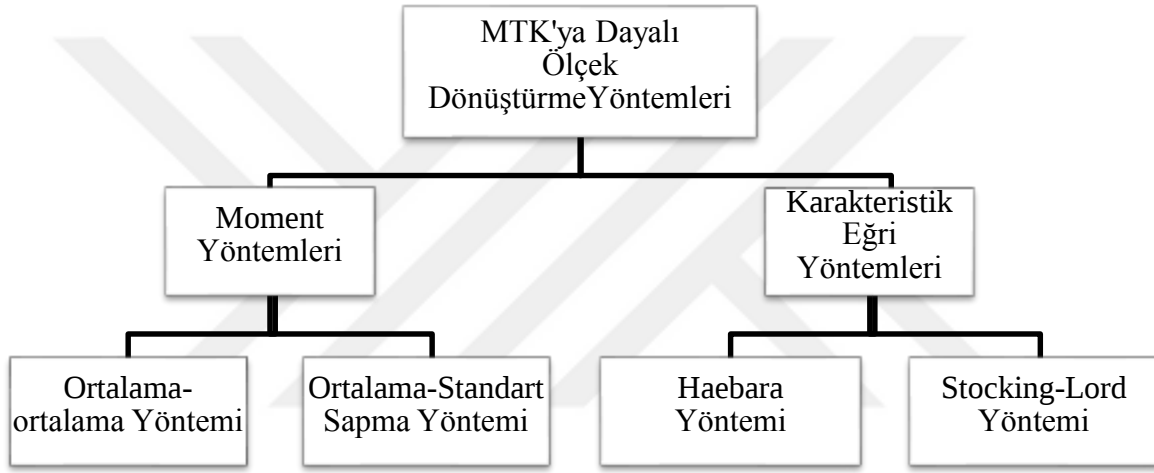
c_{ij} : Form I'daki j maddenin şans parametresi

A: Dönüştürme denkleminin eğimi

B: Dönüştürme denkleminin sabiti

Son eşitlikte görüldüğü gibi j maddesine ait c parametresi aynı kalmıştır, c parametresi ölçek dönüşümünden bağımsızdır (Kolen ve Brennan, 2004).

Bu çalışmada kullandığı gibi ölçek dönüşümü için sıklıkla kullanılan dört tane yöntem vardır. Bazı çalışmalarda eşitleme yöntemi olarak (Kilmen, 2010; Meng, 2012; Uysal, 2014) ve bazı çalışmalarda bağlama yöntemleri olarak (Lee ve Ban, 2010) değerlendirilen bu yöntemler Şekil 10’da gösterildiği gibi; ortalama- standart sapma (Macro, 1977), ortalama-ortalama (Loyd ve Hoover, 1980), Haebara (Haebara, 1980) ve Stocking-Lord (Stocking ve Lord, 1983)’ tur (Meng, 2007). Ortalama-ortalama (OO) ve ortalama-standart sapma yöntemleri (OS) moment yöntemi; Stocking-Lord (SL) ve Haebra yöntemleri (HA) de karakteristik eğri yöntemleri olarak adlandırılmaktadırlar (Meng, 2007).



Şekil 10. Araştırmada kullanılan ölçek dönüştürme yöntemleri

Moment Yöntemleri

Moment yöntemleri, ölçek dönüşümü denkleminde A ve B terimlerinin hesaplanmasında madde parametrelerinin momentlerinden faydalanır. Aşağıda moment yöntemlerinden OS ve OO ‘nun kullandıkları momentler ve kullandıkları madde parametreleri hakkında bilgi verilmiştir.

Ortalama-Standart Sapma Yöntemi

OS yönteminde denklemdeki A ve B’nin hesaplanmasında ortak maddelere ait b parametrelerinin standart sapması ve ortalaması kullanılır. Aşağıda form I ve form J’den elde edilen momentler ile A ve B’nin hesaplanması için formüller bulunmaktadır.

$$A = \frac{\sigma(b_J)}{\sigma(b_I)}$$

$$B = \mu(b_J) - A\mu(b_I)$$

$\sigma(b_J)$: Form J'nin güçlük parametrelerinin standart sapması

$\sigma(b_I)$: Form I'nin güçlük parametrelerinin standart sapması

$\mu(b_J)$: Form J'nin güçlük parametrelerinin ortalaması

$\mu(b_I)$: Form I'nin güçlük parametrelerinin ortalaması

A = Dönüştürme denklemindeki eğim

B = Dönüştürme denkleminde sabit

Yukarıdaki denklemler incelendiğinde, ölçek dönüştürme denkleminde A terimi Form J' den elde edilen b parametrelerinin standart sapmasının, Form I'dan elde edilen b parametrelerinin standart sapmasına bölünmesiyle bulunmaktadır. Ölçek dönüştürme denkleminde B terimi, Form J'den elde edilen b parametrelerinin ortalamasından Form I'dan elde edilen b parametrelerinin ortalamasının A terimi ile çarpımının çıkarılmasıyla bulunur.

Ortalama-Ortalama Yöntemi

OO yönteminde denklemdaki A ve B 'nin hesaplanmasında b parametrelerinin yanında a parametreleri de kullanılır. Aşağıda form I ve form J'den elde edilen momentler ile A ve B 'nin hesaplanması için formüller bulunmaktadır.

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

$$B = \mu(b_J) - A\mu(b_I)$$

$\mu(a_I)$: Form I'nin ayıricılık parametrelerinin ortalaması

$\mu(a_J)$: Form J'nin ayıricılık parametrelerinin ortalaması

$\mu(b_I)$: Form I'nin güçlük parametrelerinin ortalaması

$\mu(b_J)$: Form J'nin güçlük parametrelerinin ortalaması

A = Dönüştürme denklemindeki eğim

B = Dönüştürme denkleminde sabit

Yukarıdaki denklemler incelendiğinde, ölçek dönüştürme denkleminde A terimi Form I' dan elde edilen a parametrelerinin ortalamasının, Form J'den elde edilen a parametrelerinin

ortalamasına bölünmesiyle bulunmaktadır. Ölçek dönüştürme denkleminde B terimi, Form J'den elde edilen b parametrelerinin ortalamasından Form I'dan elde edilen b parametrelerinin ortalamasının A terimi ile çarpımının çıkarılmasıyla bulunur.

Karakteristik Eğrileri Yöntemi

Moment yöntemlerinde olduğu gibi madde parametrelerinin topluca değerlendirmesi ölçek dönüştürmede maddelerin formlar üzerindeki yerinin ve katkısının kısıtlanmasına neden olur. Örneğin, b parametresi kestirimleri çok farklı olan bir maddenin iki kestiriminde maddeler çok benzer madde karakteristik eğrilerine sahip olabilirler; Bu durumda, OS yöntemi bir maddenin iki kestirimdeki madde karakteristik eğrileri çok benzer olsa da b parametreleri arasındaki farklılıktan fazlaca etkilenmiş olabilir (Kolen and Brennan, 2014). Böylelikle b parametresi farklı olsa da a parametresinin farklılığını ya da benzerliğinin dikkate alınmamasıyla birlikte sorun ortaya çıkar. Bu sorun, madde parametre kestirimlerinin tamamının eş zamanlı kestirildiği dikkate alınmadığından dolayı doğar (Kolen and Brennan, 2014). Karakteristik eğri yöntemleri, ölçek dönüştürme katsayılarını elde etmek için bütün madde parametre kestirimlerinden gelen bilgiyi değerlendiren parametrik bir MTK ölçek dönüştürme yöntemidir (Dvorak, 2009). Bu metodların MTK ölçek dönüştürmenin güçlü bir yöntemi olduğu gösterilmiş ve uygulamada sıklıkla kullanılmıştır (Kaskowitz ve De Ayala'dan aktaran Dvorak, 2009).

Heabara ve Stocking–Lord yöntemleri sırasıyla madde karakteristik eğrileri veya test karakteristik eğrilerini kullanarak ölçek dönüşümünü kestirir (Lee ve Ban, 2010). Speron (2009); yaptığı çalışmada Haebara yöntemini “ madde karakteristik fonksiyonu yöntemi” diye; Stocking-Lord yöntemini ise “test karakteristik fonksiyonu yöntemi” diye atfetmiştir. Madde karakteristik eğrisi ile test karakteristik eğrisi yöntemlerinin güçlü yanı, her ikisinin de tüm madde parametrelerini eşzamanlı olarak dikkate almasıdır; bu yöntemlerin potansiyel zayıf yanı ise bir miktar keyfi olan tetaların sayısının seçimini gerektirir (Speron, 2009).

Haebara Yöntemi

Haebara (1980), madde karakteristik eğrileri arasındaki farkı belirlemek için, her bir madde karakteristik eğrilerinin farkının kareleri toplamını alır. Ancak belli bir yetenek düzeyi (θ_i) için ve yalnızca ortak maddelerin ($j=V$) karakteristik eğrilerinin farkının karesinin toplamını alır. Aşağıda Haebara yönteminin matematiksel ifadesi yer almaktadır (Kolen ve Brennan, 2014).

$$Hdiff(\theta_i) = \sum_{j=V} \left[p_{ij}(\theta_{ji}; \hat{a}_{Jj}, \hat{b}_{Jj}, \hat{c}_{Jj}) - p_{ij} \left(\theta_{ji}; \frac{\hat{a}_{Ij}}{A}, A\hat{b}_{Ij} + B, \hat{c}_{Ij} \right) \right]^2$$

A = Dönüştürme denklemindeki eğim

B = Dönüştürme denkleminde sabit

$p_{ij}(\theta_{ji}; \hat{a}_{Jj}, \hat{b}_{Jj}, \hat{c}_{Jj})$: Madde karakteristik eğrisi

$p_{ij} \left(\theta_{ji}; \frac{\hat{a}_{Ij}}{A}, A\hat{b}_{Ij} + B, \hat{c}_{Ij} \right)$: Kestirilen madde karakteristik eğrisi

$Hdiff(\theta_i)$: Verilen bir θ_i yetenek düzeyi için testteki her bir madde karakteristik eğrilerinin farkın karelerinin toplamı

$Hdiff(\theta_i)$ bulunduktan sonra kestirim sürecine aşağıdaki kritere göre A ve B 'yi minimum yapan değer hesaplanır.

$$Hcrit = \sum_i Hdiff(\theta_i)$$

Yukarıdaki eşitlikteki toplam, katılımcılar üzerinden yapılır.

Stocking Lord Yöntemi

Bu yöntemin ana amacı; iki formdan gelen test karakteristik eğrileri arasındaki farklılığı en aza indirmektir. En aza indirmeyi sağlayan kayıp fonksiyonun (loss function) değerleri; X formundaki bir puanı Y formu üzerindeki bir forma çevirmek için ölçek dönüştürme katsayılarına karar vermede kullanılır (Dvorak, 2009).

Stocking ve Lord (1983), test karakteristik eğrileri arasındaki farkı belirlemek için, tüm maddeler üzerinden, her bir madde karakteristik eğrisindeki farkın toplamlarının karesini alır. Aşağıda Stocking Lord yönteminin matematiksel ifadesi yer almaktadır (Kolen ve Brennan, 2014).

$$SLdiff(\theta_i) = \left[\sum_{j=V} p_{ij}(\theta_{ji}; \hat{a}_{Jj}, \hat{b}_{Jj}, \hat{c}_{Jj}) - \sum_{j=V} p_{ij} \left(\theta_{ji}; \frac{\hat{a}_{Ij}}{A}, A\hat{b}_{Ij} + B, \hat{c}_{Ij} \right) \right]^2$$

A = Dönüştürme denklemindeki eğim

B = Dönüştürme denkleminde sabit

$p_{ij}(\theta_{ji}; \hat{a}_{jj}, \hat{b}_{jj}, \hat{c}_{jj})$: Madde karakteristik eğrisi

$p_{ij}(\theta_{ji}; \frac{\hat{a}_{ij}}{A}, A\hat{b}_{ij} + B, \hat{c}_{ij})$: Kestirilen madde karakteristik eğrisi

$SLdiff(\theta_i)$: Her bir θ_i yetenek düzeyi için testteki toplam tüm madde karakteristik eğrilerinin toplamalarının farkının karesi

$Hdiff(\theta_i)$ 'in yalnızca verilen belli bir teta düzeyi için bulunmasının tersine $SLdiff(\theta_i)$ toplam katılımcılar üzerinden bulunur (Kolen ve Brennan, 2014). Bir sonraki aşama ise test karakteristik eğrileri arasındaki farklılığı en aza indirmek için seçilecek kritik A ve B değerinin belirlenmesidir.

$$SLcrit = \sum_i SLdiff(\theta_i)$$

Bu yaklaşımlar ($Hcrit$ ve $SLcrit$) bilgisayar programlarıyla bulunması mümkün iteratif bir yaklaşımdır (Kolen ve Brennan, 2014).

Çeşitli alanlarda ve çeşitli amaçlarla yapılan, katılımcılar için bir hayli öneme sahip olan bu sınav formlarının eşitlenmesi ve ölçek dönüştürme çalışmalarının yapılan test eşitleme araştırmalarının dikkate alınarak yapılması gereklidir. Bu anlamda ülkemizde de yaygınlaşması gereken kalibrasyon çalışmalarına yardım edebilecek bir araştırma ortaya koyulması için kalibrasyon çalışmaları zaten yapılan (eş zamanlı kalibrasyon) TIMSS 2011 verilerden uygulama yapmak için matematik testinin ortak madde içeren iki formuna (ikinci ve üçüncü kitapçık) ait puanların Madde Tepki Kuramı'na dayalı hangi ölçek dönüştürme yönteminin uygun olduğu araştırılmaktadır.

TIMSS ile ilgili Kısa Bilgi

TIMSS (Trends in International Mathematics and Science Study - Uluslararası Matematik ve Fen Eğilimleri Çalışması), merkezi Hollanda'da bulunan Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (IEA- International Association for the Evaluation of Educational Assessment) tarafından yürütülen bir tarama çalışmasıdır (MEB YEĞİTEK, 2014). Aynı zamanda TIMSS matematik ve fen alanında en uzun süredir ve en yaygın kullanılan karşılaştırmalı eğitim çalışmasıdır (Tian, 2011). 1995'teki ilk bilgi toplama çalışmasından

itibaren her 4 yılda bir yürütülen ve katılan ülkelerin 4. ve 8. sınıf öğrencileri dâhil edilen TIMSS’te öğrencilerin çok yönlü bilgi ve becerilerinin ölçülmesi amaçlanmaktadır.

TIMSS 2011’de yer alan maddelerin yarısı çoktan seçmeli, yarısı uzun/kısa cevaplı sorulardan oluşmaktadır ve her iki sınıf düzeyinde (4. ve 8. sınıf) de 14’ü fen ve 14’ü matematik sorularından oluşan 28 bloktan oluşmaktadır (MEB YEĞİTEK, 2014). TIMSS 2011’e Türkiye’den 4. sınıf düzeyinde 257 okuldan toplam 7479 öğrenci katılmıştır.

TIMSS sonuçları ilk kez 1995’te ortalaması 500, standart sapması 100 olan bir puan dağılımına göre rapor edilmiştir. Zaman içerisinde ölçümleri etkinleştirmek için, ardışık TIMSS değerlendirmelerinden gelen başarı verileri aynı ölçeğe dönüştürülmüştür. Bu dönüştürme “eş zamanlı kalibrasyon” süreci ile yapılmaktadır (Mullis, Martin, Ruddock, O’Sullivan ve Preuschoff, 2009). Bu süreç yapılan bir değerlendirmenin önceki değerlendirmenin sonucu ile doğrusal dönüşüm yapılarak aynı ölçeğe yerleştirilmesi için uygulanır. Bu işlem ile TIMSS’in beş değerlendirme döngüleri arasında eğilimleri ölçmeyi ve karşılaştırma yapmayı sağlar (1995, 1999, 2003, 2007 ve 2011) (Mullis vd, 2009).

Bu araştırmada, MTK’da ortak maddeli eşdeğer olmayan gruplar deseni altında eşitleme uygulaması gerçekleştirilmiştir. Bu eşitleme sürecinin uygulanması için öncelikle farklı gruplardan gelen farklı ölçekteki parametrelerin aynı ölçeğe yerleştirilmesi gereklidir. Bu işlem için iki ölçekten birinin diğerinin ölçeğine çevrilmesini yani ölçek dönüştürmeyi gerektirir. Bu çalışmada ilgilenilen dört ölçek dönüştürme yöntemlerini karşılaştırıp, üzerinde çalışılan veriler için en uygun bulunan yöntemi bulmak ve eşitleme sürecinin bulunan en uygun yöntemden yararlanarak tamamlanması istenmiştir. Bazı araştırmalarda (Gök, 2012; Uysal, 2014; Kilmen, 2010) ölçek dönüştürme yöntemleri eşitleme yöntemleri olarak değerlendirilse de bu çalışmada yalnızca farklı ölçekteki parametreleri aynı ölçek üzerine koymak anlamında kullanılmıştır. Bir diğer ifade ile bu araştırmada aranan cevap TIMSS 2011 verilerinden seçilen iki forma ait veriler üzerinde ortak maddeli eşdeğer olmayan gruplar deseni altında MTK’ya dayalı eşitleme yapabilmek için gerekli olan ölçek dönüştürme yöntemlerinden ortalama-ortalama, ortalama-standart sapma, Stocking-Lord ve Haebara yöntemlerinden hangisinin en uygun olduğudur. Bu durumda cevaba ulaşmak için izlenmesi gereken adımlar şöyle sıralanabilir: Birinci adım; farklı formları alan iki öğrencinin puanları arasında karşılaştırma yapıp yapılamayacağını ortaya çıkarmaktır (varsayımların kontrolü). İkinci adım; iki forma ait veriler ile ölçek dönüştürme yöntemleri ile ölçek dönüştürme denklemlerinin bulunmasıdır. Üçüncü adım; dönüşüm ile kestirilen

eşitleme hatalarına bakarak hangi yöntemin en uygun olduğunun bulunması ve son olarak; bulunan en uygun yöntem ile MTK gerçek puan eşitlemesinin yapılmasıdır.

Problem İfadesi

TIMSS 2011 verilerinden yararlanılarak, ortak maddeli eşdeğer olmayan gruplar deseni altında üç parametrelili lojistik modelde ortalama-ortalama, ortalama-standart sapma Stocking-Lord ve Haebara yöntemleri ile yetenek parametresi için yapılan ölçek dönüşümlerinden hangisi en uygun olanıdır ve en az eşitleme hatasını içerir?

Araştırma Soruları

1. TIMSS 2011 dördüncü sınıf düzeyi ortak madde içeren iki matematik alt testinin puanlarının
 - 1.1. OO ile yetenek parametresine yönelik ölçek dönüşümü sonucunda elde edilen dönüştürme katsayısı ve sabiti nedir?
 - 1.2. OS ile yetenek parametresine yönelik ölçek dönüşümü sonucunda elde edilen dönüştürme katsayısı ve sabiti nedir?
 - 1.3. HA ile yetenek parametresine yönelik ölçek dönüşümü sonucunda elde edilen dönüştürme katsayısı ve sabiti nedir?
 - 1.4. SL ile yetenek parametresine yönelik ölçek dönüşümü sonucunda elde edilen dönüştürme katsayısı ve sabiti nedir?
2. TIMSS 2011 dördüncü sınıf düzeyi ortak madde içeren iki matematik alt testinin puanlarının OO, OS, HA ve SL ile bulunan yetenek parametresine yönelik ölçek dönüştürme denklemlerinden yararlanarak; yetenek bakımından eşitleme hataları karşılaştırıldığında RMSD ortalamaları nedir?
3. TIMSS 2011 dördüncü sınıf düzeyi ortak madde içeren iki matematik alt testinin puanlarının yetenek parametresine yönelik bulunan en uygun ölçek dönüştürme ile MTK gerçek puan eşitleme yöntemi kullanılması sonucu elde edilen eşitlenmiş puanlar nasıldır?

Araştırmanın Amacı ve Önemi

Bu araştırmanın amacı MTK'ya dayalı OO, OS, SL ve HA ölçek dönüştürme yöntemlerini TIMSS 2011 verilerinden yararlanarak eşitleme hataları açısından karşılaştırmak ve bu ölçek dönüştürme çalışmasındaki yöntemlerden en uygun olanını önermektir. Ayrıca bu işlemde sonra da en uygun bulunan ölçek dönüştürme yöntemi kullanarak eşitleme sürecini tamamlamak araştırmanın amacını oluşturmaktadır.

Literatür taraması yapıldığında ölçek dönüştürme ve eşitleme ile ilgili çalışmaların çoğunluğunun simülasyon verisiyle veya simülasyon verisiyle desteklenen gerçek veriyle yapıldığına rastlanmıştır (Cohen ve Kim, 1998; French, 1996; Hanson ve Beguin, 1999, 2002; Kim ve Cohen, 2002; Kim ve Lee, 2006; Lee ve Ban, 2010; Meng, 2012; Speron, 2009; Tian, 2011; Uysal 2014; Wolf, 2013). Simülasyon verileri gerçek veride olması mümkün olmayan mükemmel bir uyum sağlar. Bu ölçek dönüştürme ve eşitleme çalışmasında ise gerçek veri setinden elde edilen veriler kullanılmıştır. Gerçek veri setiyle yapılan eşitleme çalışmalarına (Çetin, 2009; Hagge vd, 2011; He, 2011; Kelecioğlu, 1994; Peterson, 2014; Tan, 2015) daha az rastlandığından, bu eşitleme çalışması gerçek veri setiyle yapılmasından dolayı alana katkı getireceği düşünülmektedir.

Sayıtlar

1. Sınava giren iki gruptaki bütün öğrenciler için, eşit koşullar sağlanmış olduğu varsayılmıştır.
2. Jurich, DeMars ve Goodman (2012), ortak maddeli eşdeğer olmayan gruplar eşitleme deseni altında kopya çekmenin test karakteristik eğrisi yöntemi (Stocking-Lord yöntemi) ve Madde Tepki Kuramı gerçek puan eşitlemeye etkisini incelemiş ve simülasyon çalışması olan bu araştırma sonucunda kopya çekilen (taviz verililen) formun uygulandığı gruptaki tüm katılımcıların eşitlenmiş puanlarını suni olarak artırmıştır. TIMSS'de bulunan "trend madde (Mullis vd. (2009)'nin tanımı ile trend maddeler; bir TIMSS değerlendirmesinden bir sonraki TIMSS değerlendirmesine muhafaza edilen maddelerdir)." lerin paylaşılmamış olması yani; çalışılan veriler üzerinde gizlilik ihlalinin yapılmadığı varsayılmıştır.

3. Çalışmada kullanılan formların ortak maddeleri dışındaki maddelerin içeriksel olarak eşit olduğu varsayılmıştır.

Sınırlılıklar

Bu çalışma;

1. 2011 yılı TIMSS dördüncü sınıf düzeyindeki Türk öğrencilerinin cevaplandığı kitapçıklardan, ikinci kitapçığı alan 535 kişinin ve üçüncü kitapçığı alan 529 öğrencinin verileri ve ilgili kitapçıkların matematik formları ile sınırlıdır.
2. MTK'ya dayalı ikili puanlanan maddeler için bulunan modellerden yalnızca 3PLM ve dört ölçek dönüştürme yöntemi (OO, OS, HA ve SL) ile sınırlıdır.
3. Kitapçıklarda bulunan 1 tanesi ortak olan 3 madde de geçen “kısmî doğru” tepkisi “yanlış” olarak kabul edilmiş olması ile sınırlıdır.

Tanımlar

Ölçek Dönüştürme: Eşdeğer olmayan gruplar deseni altında, eşitleme sürecinin başında uygun bulunan MTK modeli parametreleri kestirildiğinde farklı test formlarından elde edilen parametreler aynı ölçekte olmayacaktır. İki test formunun parametrelerini lineer bir dönüşüm denklemi ile aynı ölçeğe (önceden uygulanmış olan formun parametre ölçeğine) yerleştirme işlemine ölçek dönüştürme denir.

Dönüştürme Katsayısı: Ölçek dönüştürme yapılması için elde edilen lineer denklemin katsayısına dönüştürme katsayısı denir.

Dönüştürme Sabiti: Ölçek dönüştürme yapılması için elde edilen lineer denklemdeki sabit terime dönüştürme sabiti denir.

Yetenek (Teta): Bireyin bir ölçek tarafından ölçülen özelliği, performansı, yeterliliği olarak tanımlanabilir.

Özdeş Eşitleme: Bir formdan elde edilen puanın diğer formda da aynı puana eş gelmesidir.

Form X: Ölçek dönüştürme ve eşitleme yapılan formlardan yeni form olarak kabul edilen, TIMSS 2011 dördüncü sınıf düzeyi ikinci kitapçıktaki matematik formu.

Form Y: Ölçek dönüştürme ve eşitleme yapılan formlardan eski form olarak kabul edilen, TIMSS 2011 dördüncü sınıf düzeyi üçüncü kitapçıktaki matematik formu.





BÖLÜM II

LİTERATÜR TARAMASI

Bu bölümde araştırmayla ilgili olabilecek, farklı desen ve kuramlar altında yer alan eşitleme, bağlama ve madde parametreleri ve/veya yetenek parametreleri bakımından ölçek dönüştürme (kalibrasyon) çalışmaları incelenmiş ve özetlenmiştir.

Kelecioğlu (1994)'nın yaptığı araştırmada ÖSS'nin Sözel ve Sayısal bölümlerinden alınan test ve madde puanlarından yararlanmış, araştırma evreni için 1990, 1991 ve 1993 yıllarında yapılan ÖSS'ye giren öğrencilerden random olarak seçilen 2000'er kişiden oluşturmuştur. Sınavın alt testlerini KTK'ya dayalı olan eşit yüzdelikli ve doğrusal; MTK'ya bağlı Rasch modeli ve 2 parametrelili lojistik model ile eşitleme yapmıştır. Bu araştırma Türkiye'de yapılan eşitleme çalışmalarından ilki olduğu kanısıyla, önem arz etmektedir. Araştırma sonucunda ise bu testlerde gerçekte herhangi bir ortak madde olmamasına rağmen, araştırmacının faktör analiziyle aynı değişkeni ölçtüğü düşünülen maddelerle oluşturduğu ortak testlerle eşitleme yapılmıştır. Sonuçlara göre uygun bulunan yöntemlerin Türkçe alt testi için doğrusal eşitleme, Fen bilimleri alt testi için eşit yüzdelikli eşitleme, Sosyal bilimler ve Matematik içinse Rasch modeli olduğu görülmüştür.

French (1996), yaptığı çalışmada kriter dayanaklı test ve norm dayanaklı testler üzerinde Stocking-Lord Yöntemi'nin kullanılabilirliğini araştırmıştır. Karma modeller üzerinde yürüttüğü bu çalışmada; simülasyon verilerini ortak madde seti 11 maddede (10'u dikotomus 1'i politomus) olan toplamda 42 maddeden (40'ı dikotomus 2'si politomus) oluşan araç ile üretmiştir. Değerlendirme yapmak için bilinen a parametresi, b parametresi ve teta değerleri (yetenek parametresi) ile eşitlenmiş (dönüştürülmüş) a parametresi, b parametresi ve teta değerleri üzerindeki RMSE ve BIAS kriter olarak alınmıştır. Araştırmacı SL Yönetimi'nin her iki (test ve norm dayanaklı) test için de uygun bir yöntem olduğu sonucuna ulaşmıştır.

Kilmen (2010), MTK'ya dayalı test eşitleme yöntemlerinden OO, OS, HA ve SL'yi farklı örneklem büyüklükleriyle (500 ve 1000) ve farklı yetenek dağılımı durumlarına göre

(yetenek ortalamaları 0 standart sapmaları 1 olan iki grup ile yetenek ortalaması 0 standart sapması 1 olan ve diğeri yetenek ortalaması 1 standart sapması 1 olan iki grup) karşılaştırılma yapılan araştırmada, WINGEN programını kullanarak 10 maddesi ortak 50 maddeli simülatif veriler üretilmiş, 3PLM altında yapılan bu çalışmada, RMSD değerlendirme ölçütüne göre SL yöntemi ile yapılan test eşitleme çalışmalarının daha az hatalı sonuçlar verdiği bulunmuştur. Ayrıca araştırmada benzer yetenek dağılımlarına ait test eşitleme hataları farklı yetenek dağılımlarına ait test eşitleme hatalarına nazaran daha az bulunmuştur.

Karkee ve Wright (2004), verilerini 31,813 öğrencinin 78 maddeye verdiği cevaplardan elde ettiği çalışmalarında, karakteristik eğri yöntemleri (Stocking Lord ve Haebara), moment yöntemleri (Ortalama-ortalama ve Ortalama-standart sapma) ve yetenek Regresyonu (Theta Regression) ile 3PLM ile kestirilen parametreleri 1PLM ile kestirilen parametrelere eşitlemişlerdir. Test karakteristik eğrisi, standart hata, ölçek puanlarının ortalaması ve standart sapması ve farklı yetenek düzeylerindeki öğrencilerin sınıflandırılması kriterlerine göre 3PLM'den 1PLM'ye yapılan eşitlemede en iyi uyum karakteristik eğri yöntemlerinde bulunmuştur. Bu kriterlere göre Karakteristik eğri yöntemleri arasındaki farklılığın önemsiz olarak bulunsa da; eğer ölçek puan dağılımlarının ortalama ve standart sapmanın sırasının daha önemli olduğu kabul edilirse Stocking Lord yöntemi tercih edilebilir olduğunu ve farklı yetenek düzeylerindeki öğrencilerin sınıflandırılmasını daha önemli görülürse Haebara yönteminin tavsiye edildiği belirtilmiştir.

Kim ve Kolen (2006), ayrı kestirim yöntemleri (iki karakteristik eğri yöntemi ve iki moment yöntemi) ile eş zamanlı kestirim yöntemini karşılaştırmıştır. Dikotomus puanlanan maddeler için 3 PLM model ile, açık uçlu olan maddeler içinse Genelleştirilmiş Kısmi Puanlama modeli ile madde parametreleri bulunmuştur. Ortak maddeli eşdeğer olmayan gruplar deseni altında, 36 çoktan seçmeli ve 15 açık uçlu maddeden oluşan iki formun 12 çoktan seçmeli 5 açık uçlu maddesi ortak (tüm testin %33,3'ü) olan iki formdan üretilen simülasyon verileriyle yapılan bu araştırmada, değerlendirme kriteri olarak MSE kullanılmıştır. Eş zamanlı kestirim yönteminin diğer yöntemlere göre daha az hatalı sonuçlar vermesiyle birlikte karakteristik eğri yöntemlerinin de moment yöntemlerine göre daha az hata verdiği görülmüştür. En fazla hataya sahip olan yöntem ise OO olarak bulunmuştur. Ayrı kestirim yöntemlerinden de en az hataya sahip olan yöntemin SL olduğu bulunmuştur.

Çetin (2009), araştırmasını 2005 yılı ÖBBS (Öğrencilerinin Başarılarının Belirlenmesi Sınavı)'ye ait 6. sınıfta 13401, 7. sınıfta 5368 ve 8. sınıfta da 11200 öğrenci nin Türkçe ve matematik test puanları ile gerçekleştirmiştir. Ortak maddeli eşdeğer olmayan gruplar deseni altında yürütülen çalışmada KTK'ya dayalı Thurstone Ölçekleme ve MTK'ya dayalı dikey ölçekleme tekniği kullanılmıştır. Bu araştırma sonucunda KTK'ya dayalı Thurstone ölçeklemede, ortalama örüntüsünden bağımsız olarak her iki testte sınıf seviyesi ile birlikte standart sapmalar artış göstermektedir. MTK'ya dayalı 2PLM model altındaki ölçekleme sonunda ise, matematik testinde, sınıf seviyesi ile ortalamalar arasında tutarlı bir artış yokken, standart sapmalar ile sınıf seviyeleri arasında artış göstermiştir. Fakat Türkçe testinde ise ne ortalama ne de standart sapmalarda düzenli bir artış veya azalış örüntüsü bulunamamıştır.

Tian (2011), eş zamanlı kestirim yöntemi ile ayrı kestirim yöntemlerinden Stocking Lord yöntemini, üç farklı ortak madde seti uzunluğu (5 madde, testin %20'sinden az; 10 madde, testin %20'si ve 15 madde, testin %20'sinden fazla) ve ortak madde setinin üç farklı tipinde (tamamı çoktan seçmeli, tamamı yapılandırılmış cevaplı, her iki madde tipi) koşulları altında karşılaştırma yapmıştır. Ortak maddeli eş değer olmayan gruplar deseni altında toplamda 18 durumu test eden bu çalışma TIMSS 2003 8. Sınıf düzeyinin matematik formunun madde parametrelerine dayanan simülasyon verileri ile yürütülmüştür. a parametreleri, b parametreleri ve c parametrelerinin MSE(ortalama hata kare) 'si ve bireylerin gerçek yetenek ile kestirilen yetenekleri arasındaki RMSE 'si kriterlerine göre değerlendirme yapılmış ve araştırmanın genelinde yöntemler arasındaki çok küçük farklılıklarla, Eş zamanlı kestirimin SL yöntemine göre daha iyi performans gösterdiğini belirtmiştir.

Meng (2012), yaptığı çalışmada, eşitleme sonuçlarını etkilediği bilinen birkaç faktörü manipüle ederek Kernel eşitleme yöntemi ve MTK'ya dayalı gerçek puanı eşitlenmesi doğruluğunu ve kararlılığını değerlendirmek istemiştir. MTK'ya dayalı yöntemlerden OS ve SL yöntemlerini kullanmıştır. Genel olarak; örneklem büyüklüğündeki farklılık (200, 500 ve 2000), ortak madde sayısındaki farklılık (5 ve 15) ve gruplar arası yetenek farklılığı (0 ve 0,5) koşulları altında daha iyi eşitleme yöntemini RMSE kriterine göre değerlendirmiştir. Her bir faktör altında MTK'ya dayalı yöntemlerin Kernel Eşitleme yöntemine göre daha doğru sonuçlar ortaya çıkardığı sonucuna varılmıştır. Simülasyon verileri ile yürütülen bu çalışmada toplamdaki on iki farklı koşulda da SL yöntemi OS yöntemine göre daha az hatalı sonuçlar vermiştir. Ayrıca ortak madde sayısı ve örneklem büyüklüğü arttıkça eşitleme hatası azalmıştır.

Hanson ve Béguin (1999), ayrı kestirim yöntemleri ile eşzamanlı kestirim yöntemini kullanarak eşitleme çalışmasını yürütmüşlerdir. Çalışmada, dikotomus olarak cevaplanan 60 maddeli iki forma ait simülasyon verileri kullanılarak, dört faktörün farklı koşulları altında incelenmiştir. Bunlar; farklı kestirim programları (MULTILOG ve BILOG-MG), farklı grup genişliklikleri (3,000 öğrenci ile 1,000 öğrenci), farklı ortak madde sayısı (20 ve 10) ve eşdeğer gruplar ile eşdeğer olmayan grupların (ortalamaları aynı olan iki grup ve ortalamaları 1 standart sapma kadar farklı olan iki grup) formları almasıdır. BIAS, varyans ve MSE değerlendirme kriterlerinin dikkate alındığı toplamda çalışılan 80 (5x2x2x2x2) durumda, Stocking Lord Yöntemi ve Haebara Yöntemi, Ortalama-ortalama ve ortalama-standart sapma yöntemlerinden daha az hata vermektedir ve istisnalar dışında, OO, OS'ye göre ve SL; HA'ya göre daha az hata vermektedir. Bu yöntemlerin tamamında ise grup genişlikleri daha büyük olduğunda ve eşdeğer olan gruplara göre eşdeğer olmayan gruplarda MSE daha az bulunmuştur. Ancak yetenek bakımından gruplar eşdeğer olmadıkça, karakteristik eğri yöntemleriyle yapılan ayrı kestirimler, eşzamanlı kestirim yöntemine göre daha iyi sonuçlar verdiği bulunmuştur.

Speron (2009), eşdeğer olmayan gruplarda ortak maddeli eşitleme deseni altında sıklıkla kullanılan MTK bağlama yöntemleri yanı sıra alan küçültme yaklaşımını (Area-Minimization Approach) değerlendirmiştir. Değerlendirme yaptığı yöntemler: ortalama-ortalama, ortalama-standart sapma, test karakteristik fonksiyonu (SL), madde karakteristik fonksiyonu (HA) ve alan küçültme yaklaşımıdır. Bu yöntemler ile de dört farklı koşul altında çalışılmıştır. 40 maddelik dikotomus puanlanan ve 2PLM modele uyumlu iki form için üretilen simülasyon verileriyle yürütülen bu çalışmadaki faktörler şöyledir: grup büyüklükleri (250 kişi ve 1000 kişi), ortak madde sayısı (%10, %25, %100), grup farklılıkları (ortalama ve standart sapmaları aynı, ortalamaları farklı standart sapmaları aynı, ortalamaları aynı standart sapmaları farklı, ortalamaları ve standart sapmaları farklı) ve ortak madde temsiliyeti (Formdaki parametrelerinin aralığından düzgün bir şekilde örneklenmiş ve böylece ideal parametre aralığında dağıtılan ortak maddeler ve herhangi bir sınırlandırılma yapılmamış, parametre aralığında dağıtılmamış ortak maddeler). Araştırma sonucunda ise RMSE ve BIAS kriterleriyle yapılan değerlendirmeler sonucu karakteristik eğri yöntemleri ve ortalama-ortalama yöntemlerinin yakın sonuçlar verdiği ve daha uygun kestirimlere neden olduğu bulunmuştur. Genel anlamda grup büyüklüğü ve ortak madde sayısı arttıkça, ortak madde temsiliyeti de ideal oldukça ve grup farklılıkları azaldıkça eşitlemeler daha az hatalı sonuçlar vermiştir.

Lee ve Ban (2010), random gruplar deseni üzerinde yaptıkları simülasyon çalışmasında, örneklemelerin aynı evrenden geldiği durumlar hariç ayrı kalibrasyon işlemleri (Stocking - Lord ve Haebara) eş zamanlı kalibrasyon ve yeterlilik dönüşümüne göre genel olarak, daha iyi performans sağladığını ve iki ayrı kalibrasyon işlemleri arasında ise Haebara metodunun Stocking - Lord metoduna göre daha düşük hata verdiği sonucuna ulaşmıştır. Ancak formlardaki madde sayısı azaldıkça ve grup büyüklükleri küçüldükçe; yeterlilik dönüşümleri ayrık kalibrasyonlar kadar ve hatta bazen daha az hatalı sonuçlara rastlanıldığı belirtilmiştir.

Gök (2012), eşdeğer olmayan gruplarda ortak madde deseni altında MTK'ya dayalı OO, OS, SL yöntemlerini iki model (2PLM ve 3PLM) altında, üç farklı test uzunluğunda (30-60 ve 80 soruluk), iki farklı örneklem büyüklüğüne (1000 ve 3000 kişi) ve üç farklı yetenek koşulları altında toplam 36 farklı durumda karşılaştırmıştır. Araştırma verilerini; iki kategorili madde cevapları Wingen3 simülasyon programını kullanarak üretmiştir. RMSE ve Bias kriterlerine göre yaptığı değerlendirmelerde 3PLM modele uyumlu veriler için en az hata içeren yöntem genel olarak SL yöntemi olurken, en fazla hata içeren yöntem ise OS yöntemidir.

Uysal (2014), ortak maddeli eşdeğer olmayan gruplar eşitleme deseni altında, MTK'ya dayalı test eşitleme yöntemlerinden OO, OS, Sağlam OS, Haebara ve Stocking Lord yöntemlerini karma modeller altında ve simülatif veriler üzerinde karşılaştırmıştır. Çalışmada yöntemleri değerlendirme kriteri olarak RMSD seçilmiştir. Farklı ortak madde uzunluğunda olan (%10, %20 ve %30) ve grupların yetenek dağılımının benzer (normal-normal, sola çarpık-sola çarpık, sağa çarpık-sağa çarpık) ya da farklı olduğu (sola çarpık-normal, sağa çarpık-normal, sağa çarpık-sola çarpık) koşullar altında yapılan karşılaştırmada SL yönteminin genellikle daha az hata verdiği, sağlam OS yönteminin ise genellikle en fazla hatayı verdiği görülmüştür.

Kim ve Lee (2006), karma testler üzerinde MTK'ya dayalı Ortalama-ortalama, Ortalama-standart sapma Haebara ve Stocking-Lord yöntemlerini beş MTK modeli altında karşılaştırmıştır: üç parametrelili, kademeli tepki (GR), Genelleştirilmiş Kısmi Kredi (GPC), nominal tepki (NR) ve çoktan seçmeli (MC) modelleridir. Yapılan bu simülasyon çalışmasında değerlendirme kriteri olarak MSE (ortalama hata kareleri) seçilmiştir. Araştırmanın bulgularında genel olarak karakteristik eğri yöntemleri moment yöntemlerine göre daha az MSE üretmiştir. Haebara yöntemi ise kalan diğer üç yönteme göre en az MSE'yi vermiştir.

Kim ve Kolen (2006), MTK varsayımlarının ve model veri uyumunun ihlali altında karakteristik eğri yöntemleri moment yöntemlerini ve eş zamanlı yöntemi karşılaştırmıştır. Çoktan seçmeli ve kısa cevaplı maddeleri içeren karma test simülasyon verileriyle yürütülen çalışmada dikotomus maddeler için 3PLM model, politomus maddeler içinse GPC model ile parametre kestirimleri yapılmıştır. Eşdeğer olmayan gruplar ve testteki madde türlerinin arasındaki korelasyonların 3 farklı düzeyi altında, yöntemleri karşılaştırmak için, kestirilen ve gerçek gözlenen puan dağılımı arasındaki farklılığa dayalı olan, gözlenen puan dağılımı (OSD) ve MSE kriter olarak seçilmiştir. Araştırmanın bulgularında çoktan seçmeli ve kısa cevaplı soruların arasındaki korelasyon arttıkça yöntemlerin içerdiği hatanın azaldığı ve moment yöntemlerinin karakteristik eğri yöntemleri ve eş zamanlı kestirime göre daha fazla hata içeren sonuçlar verdiği sonucuna ulaşılmıştır.

He (2011), ortak maddeli eşdeğer olmayan grup deseni altında, çoktan seçmeli ve kısa cevaplı maddelerin bir arada bulunduğu karma testlerin eşitlenmesi üzerine bir çalışma yürütmüştür. Gerçek veriler üzerinden yürütülen çalışma da 39 karma test kullanılmıştır. Dört tane eşitleme yöntemi kullanılmıştır; bunlar frekans kestirimi, zincir eşit yüzdelikli eşitleme, MTK gerçek puan eşitleme ve MTK gözlenen puan eşitleme. Sonuçlar üç eşitleme özelliğine göre değerlendirilmiştir; birincil eşitlik (first-order equity), ikincil eşitlik(second-order equity) ve eşit özellik dağılımı. Araştırma sonucunda MTK eşitleme yöntemleri arasında gerçek puan eşitlemede birincil eşitlik gözlenen puan eşitlemeye göre daha iyi korunurken, gözlenen puan eşitlemede ikincil eşitlik ve eşit özellik dağılımı gerçek puan eşitlemeye göre daha iyi korunduğu ortaya çıkmıştır. KTK eşitleme yöntemleri arasında zincir eşit yüzdelikli eşitlemede birincil eşitlik frekans kestirimi yöntemine göre daha iyi korunurken, frekans kestirimi yöntemi ile eşitlemede ikincil eşitlik ve eşit özellik dağılımı zincir eşit yüzdelikli eşitlemeye göre daha iyi korunduğu ortaya çıkmıştır. Çoktan seçmeli ve kısa cevaplı maddelerin arasındaki yüksek korelasyon her iki MTK modeli için birincil eşitlik özelliğinin daha iyi korunmasıyla ilişkilendirilmiştir. Literatürdeki rastlanan çalışmaların aksine (Speron, 2009; Tian, 2011; Meng; 2012; Uysal, 2014) bu çalışmada ortak madde oranındaki değişimin her bir eşitleme sonucunda önemli bir etki oluşturmadığını belirtmiştir.

Wolf (2013), dört eşitleme yöntemini kullanılarak, eşitlik özelliklerinin korunmasını incelenmiştir. Bu yöntemler MTK gerçek puan eşitleme, MTK gözlenen puan eşitleme, frekans kestirimi (frequency estimation) ve zincirleme eşit yüzdelikli (chained equipercentile) eşitlemedir. Simülasyon verilerinin kullanıldığı ve ortak maddeli eşdeğer

olmayan gruplar deseni altında yapılan bu çalışmanın amacı; her bir eşitleme yönteminin eşitlik özelliklerinin korumasındaki; grup yetenek dağılımlarındaki farklılık (eş değer ve eşdeğer olmayan) ile ortak madde setinin özelliği (biçimsel olarak temsil eden ve temsil etmeyen) arasındaki farklılık durumlarında test yapısının (tek boyutlu ve çok boyutlu) etkisini incelemektir. Araştırmanın ana bulguları şöyle özetlenebilir: MTK eşitleme yöntemleri tüm koşullar arasında eşitlik özelliklerinin korunması açısından geleneksel eşitleme yöntemlerinin üzerinde bir performans göstermiştir. Gruplar eşdeğer olduğunda geleneksel eşitleme yöntemleri benzer performans göstermiştir. Ancak ortalama grup yeteneklerindeki artış, yöntemler arasında büyük farklılık yaratmıştır. MTK gerçek puan eşitleme yöntemi, test yapısı ne olursa olsun birinci dereceden eşitliğin korunması açısından en başarılı olarak bulunmuştur. Gruplar eşdeğer olduğunda her yöntem için eşit dağılım özelliği açısından benzer sonuçlar elde edilmiştir. MTK gözlenen puan yöntemi ortalama grup yetenek farklılıkları arttığında en iyi performansı göstermiştir.

Yapılan çalışmalar, farklı eşitleme yöntemlerini karşılaştırmaya ek olarak KTK ve MTK eşitleme metotları arasındaki karşılaştırmaları da içermektedir (Cook ve Eignor, 1985; Kolen, 1981; Marco, Petersen ve Stewart, 1983; Petersen, Cook ve Stocking, 1983). Özetle araştırma sonuçları göstermiştir ki; MTK eşitleme yöntemlerinin KTK eşitleme yöntemlerine göre bazı üstünlükleri vardır. Örneğin Marco ve arkadaşları (1983) güçlükleri eşit olmayan düzeydeki testler eşitlendiğinde MTK yöntemlerinin en az hata verdiğini bulmuştur. Peterson ve arkadaşları (1983), MTK ve KTK yöntemlerini SAT SÖZEL sınavı üzerinden kıyaslamış ve MTK yöntemlerinin KTK yöntemlerine göre çok daha az hata verdiğini göstermiştir. Başka bir çalışmada, Cook ve Eignor (1985), GRE başarı testleri üzerinden MTK ve KTK testlerini kıyaslamış ve eşit yüzdelikli eşitlemenin her bir uygulamada yetersiz olduğunu bulmuştur.



BÖLÜM III

YÖNTEM

Bu bölümde araştırmanın modeli, veri toplama deseni, verilerin analizi ve değerlendirme kriterine yer verilmiştir. Ayrıca formlara ait betimsel istatistiklere, MTK varsayımlarının ve test eşitleme koşullarının test edilmesine yer verilmiştir.

Araştırmanın Modeli

Bu çalışmada, TIMSS 2011 verilerinden yararlanılarak, ortak maddeli eşdeğer olmayan gruplar deseni altında üç parametrelili lojistik modelde ortalama-ortalama, ortalama-standart sapma, Stocking-Lord ve Haebara yöntemlerinden, belli bir kritere göre, araştırma kapsamında en uygun olanı seçmek amaçlanmıştır. Araştırmada var olan yöntem ve tekniklerin gerçek veri üzerinden sınanması yapıldığından ve daha çok durum belirlemeye yönelik bir çalışma olduğu için araştırma temel araştırma niteliğindedir (Karasar, 2009).

Çalışma Grubu

Çalışmanın evrenini TIMSS 2011 çalışması için 4. sınıf düzeyinde Türkiye’deki belirlenen 260 okuldan ulaşılabilen 257 okul oluşturmuştur. Bu araştırmanın çalışma grubunu TIMSS 2011’e katılan dördüncü sınıf Türk öğrencilerinden random (seçkisiz) olarak seçilen bir formu alan (ikinci kitapçık) ve araştırma doğrultusunda ortak madde içeren bu kitapçıkla ardışık olan kitapçıklardan birini (bir önceki veya bir sonraki kitapçık) cevaplayan gruptan da madde sayılarının eşit olması için üçüncü formu almış olan toplam 1064 öğrenci oluşturmaktadır.

Veri Toplama Deseni (Eşitleme Deseni)

Araştırmada kullanılan veriler 2011 yılı TIMSS’in Türkiye’den dördüncü sınıf düzeyindeki katılımcıların cevaplandığı iki ve üç numaralı kitapçığının matematik test formlarından elde edilen verilerden oluşmaktadır. Her bir form 14’ü ortak (iki numaralı kitapçığın

matematik formunun son 14 maddesi ile üç numaralı kitapçığın matematik formunun ilk 14 maddesi ortak) ve 12'si farklı olmak üzere toplam 26'şar maddeden oluşmaktadır. Bu iki matematik formu iki farklı gruba uygulanmıştır. Böylece veriler, ortak madde içeren iki farklı formun iki farklı gruba uygulanması ile elde edilmiştir. Çalışmanın veri toplama şekline uygun olarak, bu araştırmada eşitleme desenlerinden, eşdeğer olmayan gruplarda ortak madde deseni kullanılmıştır.

Verilerin Düzenlenmesi ve Analizi

Araştırmada kullanılan veriler <http://timssandpirls.bc.edu/timss2011/international-database.html> SPSS formatında alınıp ülkemizin öğrencilerinin verilerinin bulunduğu dosyadan iki ve üç numaralı kitapçıktan matematik başarı formları için veriler 1-0 şeklinde düzenlenmiştir. Doğru tepkiler 1, kısmen doğru ve yanlış cevaplar ile diğer durumlarda 0 şeklinde kodlanmıştır. Kısmen doğru cevap verilebilecek olan maddelerin iki tanesi bir testte diğeri ikinci testte olduğundan, istatistiksel ve içeriksel bütünlüğünün sağlanması için atılmamıştır. Ayrıca araştırmada dikotomus puanlanan verilerin kullanıldığı modeller altında çalışıldığından, kısmen doğru cevap tepkisi yanlış kategorisine alınmıştır. Çalışmada kolaylık bakımından iki numaralı kitapçığın matematik formunu Form X olarak; üç numaralı kitapçığın matematik formu Form Y olarak adlandırılmıştır. Form X ve Form Y'ye ait bazı istatistikler hesaplanmış ve Tablo 1'de verilmiştir.

Tablo 1: Form X ve Form Y' nin Ortalamaları, Standart Sapmaları, KR-21 Güvenirlik Katsayıları, Basıklık Katsayıları ve Çarpıklık Katsayıları

	Form X	Form Y
Kişi Sayısı (N)	535	529
Soru Sayısı (K)	26	26
Ortalama ($\bar{X}$)	11,20	11,04
Mod	8	8
Medyan	11	11
Standart Sapma (ss)	5,62	5,61
KR-21	0,86	0,86
Basıklık Katsayısı (Krt)	-0,73(0,211*)	-0,68(0,191*)
Çarpıklık Katsayısı (Sk)	0,22(0,106*)	0,33(0,139*)

(*momente ait standart hata)

Çarpıklık katsayısının ± 1.0 sınırları içinde kalması puanların normalden aşırı bir sapma göstermediği şeklinde yorumlanabilirken, basıklık katsayısının 0'dan küçük olması dağılımın basık olduğunu gösterir (Büyüköztürk, Çokluk ve Köklü, 2012). Ancak çarpıklık

ve basıklık momentleri ile standart hatalarının oranına bakılınca her bir değer 1.96'dan büyük çıkmaktadır. Buradan 0,05 manidarlık düzeyi ile dağılımın normal olmadığı kanısına varılabilir.

Tablo 1'in ışığı altında Form X ve Form Y ye ait puanları, bir yoruma göre normalden aşırı bir sapma göstermemişse de başka bir bulgu aksine yorumlanabilir. Çalışmanın ileriki kısımlarında normallik varsayımı farklı yöntemlerle de test edilmiştir. Formlardan elde edilen puanların güvenilirlikleri ise araştırma için yeterli bulunmuştur(KR-21=0,86>0,70).

Testlerin KTK'ya göre genel özelliklerine bakıldıktan sonra BILOG MG 3.00 ile parametre kestirimi yapılmıştır.

Ölçek dönüştürme ve eşitleme çalışmaları yapılacak olan iki formun kestirilen madde parametrelerine Tablo 2'de yer verilmiştir.

Tablo 2: Form X ve Form Y'ye Ait Maddelerin Parametreleri

X Formuna Ait Maddelerin Parametreleri				Y Formuna Ait Maddelerin Parametreleri			
Madde no	a	b	c	Madde no	a	b	c
1	1,304	0,577	0,318	1	2,512	0,05	0,267
2	2,671	0,79	0,22	2	1,381	2,865	0,161
3	1,966	1,234	0,096	3	1,416	1,412	0,325
4	1,513	2,193	0,281	4	1,48	0,107	0,232
5	1,639	0,037	0,178	5	1,981	0,962	0,092
6	2,041	0,908	0,095	6	2,363	1,974	0,048
7	1,457	1,126	0,15	7	2,903	0,47	0,271
8	1,881	0,16	0,169	8	1,114	0,567	0,262
9	2,274	1,192	0,116	9	1,879	0,804	0,283
10	1,586	0,282	0,261	10	2,591	0,774	0,086
11	2,96	0,248	0,102	11	2,764	-0,378	0,272
12	2,074	0,856	0,325	12	1,64	2,464	0,047
13	2,946	-0,022	0,244	13	1,644	-0,760	0,367
14	1,864	2,296	0,208	14	1,918	-0,567	0,373
15	1,425	1,876	0,306	15	1,503	-0,277	0,218
16	1,694	0,199	0,203	16	2,935	0,45	0,105
17	2,478	0,857	0,096	17	2,352	1,713	0,038
18	2,232	1,833	0,050	18	1,713	2,166	0,06
19	2,381	0,406	0,285	19	2,505	1,276	0,217
20	1,36	1,119	0,182	20	1,781	1,187	0,2
21	1,575	0,742	0,229	21	1,392	0,406	0,339
22	1,92	0,984	0,108	22	2,054	0,638	0,253
23	2,396	-0,113	0,247	23	2,045	0,465	0,139
24	2,293	2,3	0,036	24	1,016	1,754	0,225
25	1,45	-0,726	0,332	25	2,12	0,991	0,08
26	1,871	-0,313	0,348	26	2,221	1,019	0,177

Form X ve Form Y ‘nin a, b ve c madde parametrelerinin ortalamalarına Tablo 3’te yer verilmiştir.

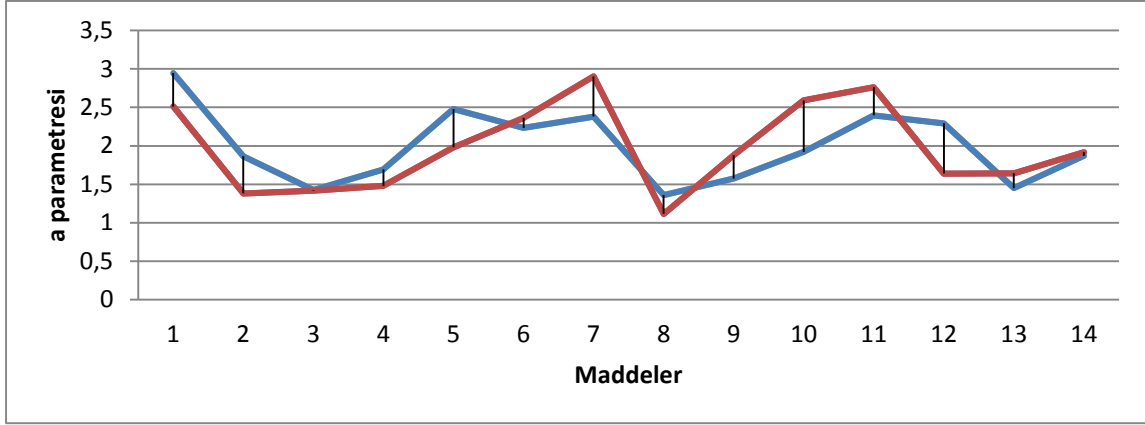
Tablo 3: Formlara Ait Ortalama Ayıricılık , Ortalama Güçlük ve Ortalama Şans Başarısı Parametreleri

	N	Ortalama Ayıricılık (a _{ort})	Ortalama Güçlük (b _{ort})	Ortalama Şans Başarısı (C _{ort})
Form X	26(12 farklı +14 ortak madde)	1,971	0,809	0,199
Form Y	26(12 farklı +14 ortak madde)	1,970	0,866	0,198

Tablo 3’te de görüldüğü gibi test formlarına ait maddelerin ayırt ediciliklerinin ortalamaları, güçlüklerinin ortalamaları ve şans başarılarının ortalamaları birbirine benzerdir. Bir testin farklı formlarının ortalama güçlük düzeylerinin ve testi alan grupların yetenek dağılımlarının yaklaşık olarak eşit olduğu durumlarda yatay eşitlemeye başvurulur (Kellecioğlu,1994). Bu araştırmada da Form X ve Form Y puanları arasında yatay eşitlemeye başvurulmuştur.

Form X’in 13-26 numaralı maddeleri ile Form Y’nin 1-14 numaralı maddeleri sırasıyla ortak maddelerdir. MTK’da madde parametreleri örneklemden bağımsız kestirildiği için (parametre değişmezliği) Tablo 2’deki koyu renklerle vurgulanmış ortak maddelerin parametrelerinin yakın olması beklenmektedir. Örneğin Form X’in 13. maddesinin parametre değerleri ile Form Y’nin 1. maddesinin parametre değerlerinin birbirine yakın olması gerekmektedir (parametre değişmezliği). Ancak örneklem büyüklüğünün küçük olması gibi nedenlerden dolayı parametreler bire bir örtüşmemektedir (N<500). Parametrelerin örtüşmelerine bakmak için, ortak maddelerin a, b ve c parametrelerine grafiksel olarak bakılmıştır.

Ortak maddelerin a parametrelerine Şekil 11’deki gibi grafiksel olarak bakıldığında yaklaşık olarak örtüştüğü yorumu yapılabilir. Şekil 11’de mavi ile gösterilen dize Form X’in 12. maddeden 26. Maddeye kadar olan 14 ortak maddenin ayırt ediciliği; kırmızı ile gösterilen dize Form Y’nin 1. maddeden 14. maddeye kadar olan 14 ortak maddenin ayırt ediciliğidir.



Şekil 11. Ortak maddelerin a parametrelerinin grafiği

Şekil 11’de görüldüğü üzere ortak maddelerin ayırt edicilikleri arasındaki farkların en fazla olanı ortak maddelerden onuncu maddeye (Form X’te 22. Madde ve Form Y’de 10. madde) ait (0,671); en az fark ise üçüncü maddeye (Form X’te 15. madde ve Form Y’de 3. madde) aittir (0,009). Ortak maddelerin ayırtıcılık parametrelerinin ortalamaları arasında manidar bir fark olup olmadığını belirlemek için İlişkisiz Örneklem t testi yapılmıştır. Bu analizden önce t testi sayıtları test edilmiştir. Form X ve Form Y’den elde edilen a parametrelerinin normal dağılım gösterip göstermediğine ilişkin, örneklem büyüklüğü 30’dan küçük olduğundan, Shapiro-Wilk istatistiği kullanılmıştır. Buna göre her iki dağılımın da 0,05 manidarlık düzeyinde normal dağıldığı görülmüştür ($p=0,493>0,05$ ve $p=0,473>0,05$). Tablo 4’te formlarda bulunan ortak madde setinin ayırtıcılığına (a parametresine) göre ilişkisiz t-testi ile yapılan analizin tablosu vardır.

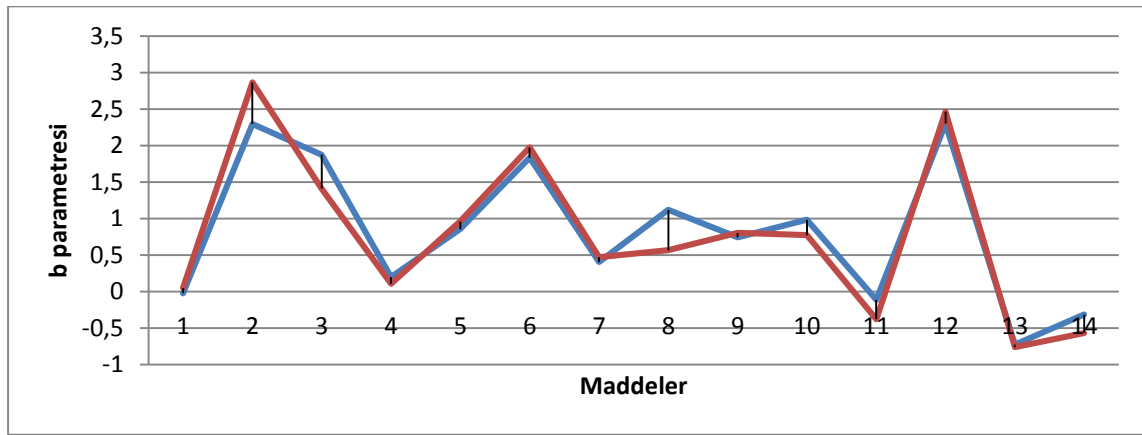
Tablo 4: Formlarda Bulunan Ortak Madde Setinin Ayırtıcılığa (a parametresine) göre İlişkisiz t-testi ile Karşılaştırılması

Ortak Madde Seti	N	Ortalama	ss	t	sd	p
Form X _{ankor}	14	1,99	0,47	0,11	26	0.915
Form Y _{ankor}	14	1,97	0,56			

Analiz sonuçlarına göre, her iki formdaki a parametrelerinin ortalaması istatistiksel olarak manidar bir farklılık göstermemektedir ($t(26)= 0.11, p>0.05$). Form X’teki ve Form Y’deki ancor madde setine (ortak madde setine) ait ayırtıcılık parametreleri (a parametreleri) ortalamalarına bakıldığında Form X_{ankor} ayırtıcılık parametreleri ortalaması 1.99 iken Form

Y_{ankor} ayırıcılık parametreleri ortalaması 1.97'dir. Ortak madde setlerine ait standart sapmaya bakıldığında ise değerlerin dağılımlarının birbirine benzer olduğu söylenebilir.

Ortak maddelerin b parametrelerine Şekil 12'deki gibi grafiksel olarak bakıldığında neredeyse örtüşmüştür yorumu yapılabilir. Şekil 12 'da mavi ile gösterilen dize Form X 'in 12. maddeden 26. Maddeye kadar olan 14 ortak maddenin güçlüğü; kırmızı ile gösterilen dize Form Y'nin 1. maddeden 14. maddeye kadar olan 14 ortak maddenin güçlüğüdür.



Şekil 12. Ortak maddelerin b parametrelerinin grafiği

Şekil 12'da görüldüğü üzere ortak maddelerin güçlükları arasındaki farkların en fazla olanı ortak maddelerden ikinci maddeye (Form X'de 14. Madde ve Form Y'de 2. madde) ait (0,569); en az fark ise on üçüncü maddeye (Form X'de 25. madde ve Form Y'de 13. madde) aittir (0,034). Ortak maddelerin güçlük parametrelerinin arasında manidar bir fark olup olmadığını belirlemek için İlişkisiz Örneklem t testi yapılmıştır. Bu analizden önce t testi sayıltıları test edilmiştir. Form X ve Form Y'den elde edilen b parametrelerinin normal dağılım gösterip göstermediğine ilişkin, örneklem büyüklüğü 30'dan küçük olduğundan, Shapiro-Wilk istatistiği kullanılmıştır. Buna göre her iki dağılımın da 0,05 manidarlık düzeyinde normal dağıldığı görülmüştür ($p=0,549>0,05$ ve $p=0,615>0,05$).

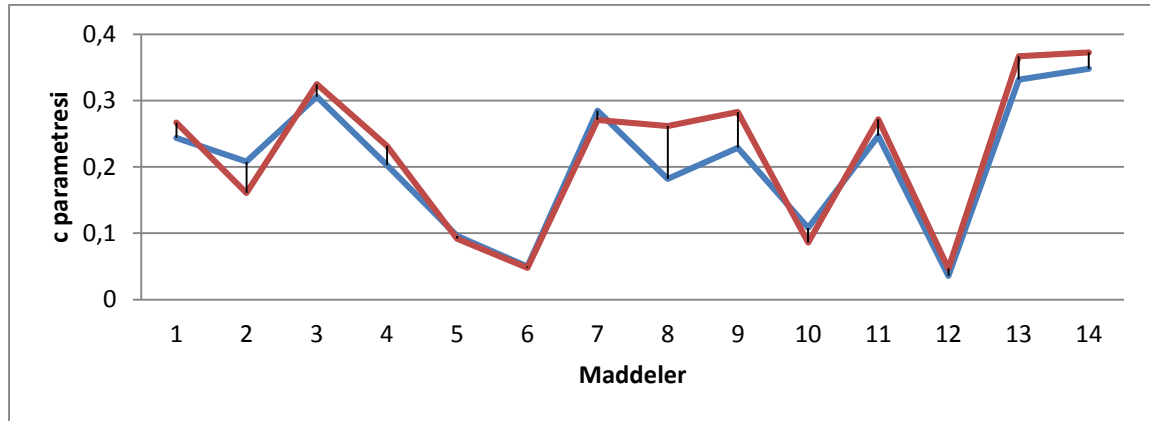
Tablo 5'te formlarda bulunan ortak madde setinin güçlüğüne (b parametresine) göre ilişkisiz t-testi ile yapılan analizin tablosu vardır.

Tablo 5. Formlarda Bulunan Ortak Madde Setinin Güçlüğüne (b parametresine) göre İlişkisiz t-testi İle Karşılaştırılması

Ortak Madde Seti	N	Ortalama	ss	t	sd	p
Form X _{ankor}	14	0,82	0,98	0,13	26	0.901
Form Y _{ankor}	14	0,77	1,09			

Analiz sonuçlarına göre, her iki formdaki b parametrelerinin ortalaması istatistiksel olarak manidar bir farklılık göstermemektedir ($t(26) = 0.13, p > 0.05$). Form X deki ve Form Y deki ankor madde setine (ortak madde setine) ait güçlük parametreleri (b parametreleri) ortalamalarına bakıldığında Form X_{ankor} güçlük parametreleri ortalaması 0.82 iken Form Y_{ankor} güçlük parametreleri ortalaması 0,77’dir. Ortak madde setlerine ait standart sapmaya bakıldığında ise değerlerin dağılımlarının birbirine benzer olduğu gözlenmiştir.

Ortak maddelerin c parametrelerine Şekil 13’teki gibi grafiksel olarak bakıldığında yaklaşık olarak örtüştüğü yorumu yapılabilir. Şekil 13’te mavi ile gösterilen dize Form X ‘in 12. maddeden 26. Maddeye kadar olan 14 ortak maddenin şans başarısı; kırmızı ile gösterilen dize Form Y’nin 1. maddeden 14. maddeye kadar olan 14 ortak maddenin şans başarısı parametresidir.



Şekil 13. Ortak maddelerin c parametrelerinin grafiği

Şekil 13’te görüldüğü üzere ortak maddelerin ayırt edicilikleri arasındaki farkların en fazla olanı ortak maddelerden sekizinci maddeye (Form X’de 20. madde ve Form Y’de 8. madde) ait (0,080); en az fark ise altıncı maddeye (Form X’de 18. madde ve Form Y’de 6. madde) aittir (0,002). Ortak maddelerin ayırtıcılık parametrelerinin arasında manidar bir fark olup olmadığını belirlemek için İlişkisiz Örneklem t testi yapılmıştır. Bu analizden önce t testi

sayıltıları test edilmiştir. Form X ve Form Y’den elde edilen b parametrelerinin normal dağılım gösterip göstermediğine ilişkin, örneklem büyüklüğü 30’dan küçük olduğundan, Shapiro-Wilk istatistiği kullanılmıştır. Buna göre her iki dağılımın da 0,05 manidarlık düzeyinde normal dağıldığı görülmüştür ($p=0,549>0,05$ ve $p=0,615>0,05$).

Tablo 6’ da formlarda bulunan ortak madde setinin şans parametre (c parametresine) göre ilişkisiz t-testi ile karşılaştırılması özetlenmiştir.

Tablo 6. Formlarda Bulunan Ortak Madde Setinin Şans Parametre (c parametresine) Göre İlişkisiz t-testi İle Karşılaştırılması

Ortak Madde Seti	N	Ortalama	ss	t	sd	p
Form X _{ankor}	14	0,21	0,10	-0,37	26	0.711
Form Y _{ankor}	14	0,22	0,11			

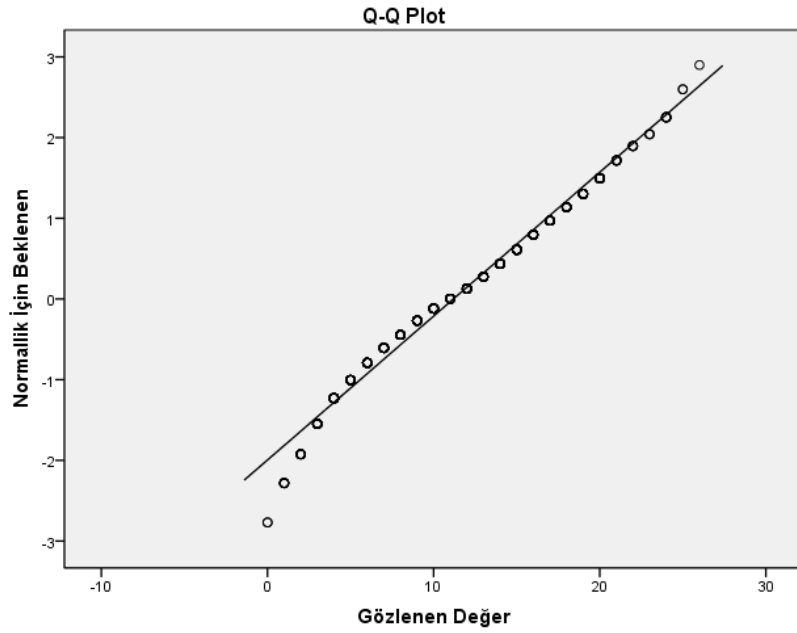
Analiz sonuçlarına göre, her iki formdaki c parametrelerinin ortalaması istatistiksel olarak manidar bir farklılık göstermemektedir ($t(26) = -0.37$, $p>0.05$). Form X deki ve Form Y deki ankor madde setine (ortak madde setine) ait şans parametreleri (c parametreleri) ortalamalarına bakıldığında Form X_{ankor} şans parametreleri ortalaması 0.21 iken Form Y_{ankor} şans parametreleri ortalaması 0,22’dir. Ortak madde setlerine ait standart sapmaya bakıldığında ise değerlerin dağılımlarının birbirine yakın olduğu gözlenmiştir.

Formlarda bulunan ortak madde setinin ayırıcılık, güçlük ve şans parametresine göre karşılaştırılmasında yalnızca ilişkisiz t-testine bakılması yanıltıcı olur. Bu yüzden her bir parametre serilerine ait ilişkisiz t testi karşılaştırması yapılan grafiksel incelemelerle birlikte değerlendirilmelidir.

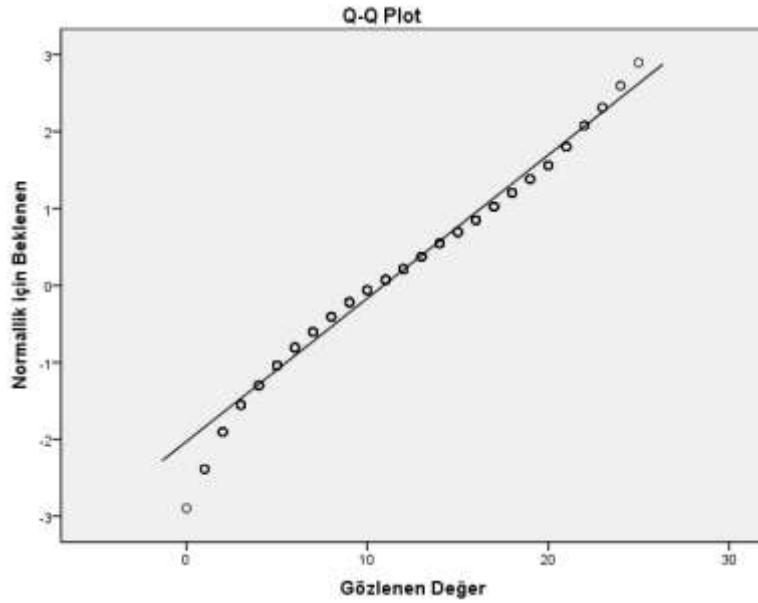
MTK Sayıltılarının Test Edilmesi

MTK’ ya ve eşitlemeye ilişkin analizlerin yapılabilmesi için tek boyutluluk, yerel bağımsızlık ve madde karakteristik eğrilerinin monotonik artışı sayıltıların sağlanıp sağlanmadığı kontrol edilmiştir. Bunların yanında yapılan son araştırmalardan elde edilen sonuçlar ışığında parametrik modellerde de karşılanması gereken normallik sayıltısı kontrol edilmiştir (Yıldırım, 2015). Çalışmanın önceki kısmındaki Tablo 1’in yorumlanmasının ışığı altında formlardan elde edilen verilerin dağılımın normal dağılımdan çok fazla sapma göstermediğine ilişkin yorum yapılmıştı. Form X ve Form Y’den elde edilen verilerin normal dağılım gösterip göstermediğine ilişkin Kolmogorov-Smirnov Uyum İyiliği istatistiği

kullanılmıştır. Buna göre her iki dağılımın 0,05 manidarlık düzeyinde normallik varsayımını karşılamadığı görülmüştür (her iki istatistikte de $p=0,000<0,05$). Bunun üzerine dağılımın normalliğini test etmek için bir başka yöntem olan grafik inceleme yapılmıştır. Bunun için, normal dağılım eğrisinin de çizdirildiği histogram, gövde-yaprak diyagramı, aşırı puanların gözlemlendiği kutu-çizgi grafiği, normal Q-Q Plot grafiği ve Dexrended normal Q-Q grafiği sıklıkla kullanılır (Büyüköztürk, 2013). Şekil 14, Form X 'e ait Q-Q Plot (quantile-quantile plot) grafiği, Şekil 15 ise Form X 'e ait Q-Q Plot grafiğidir.



Şekil 14. Form X'in Q-Q plot grafiği



Şekil 15. Form Y'nin Q-Q plot grafiği

Q-Q grafiğinde noktalar 45 derecelik doğru üzerinde veya yakın bir durumda gözüktüyorsa normal dağılıma uygunluktan söz edilebilir (Büyüköztürk, 2013). Form X ve Form Y'ye ait grafikler incelendiğinde dağılımlarının normal olduğu söylenebilir. Çalışmanın bundan sonraki kısmında tek boyutluluk, yerel bağımsızlık, madde karakteristik eğrilerinin monotonik artışı ve model veri uyumu test edilmiştir.

Tek Boyutluluk

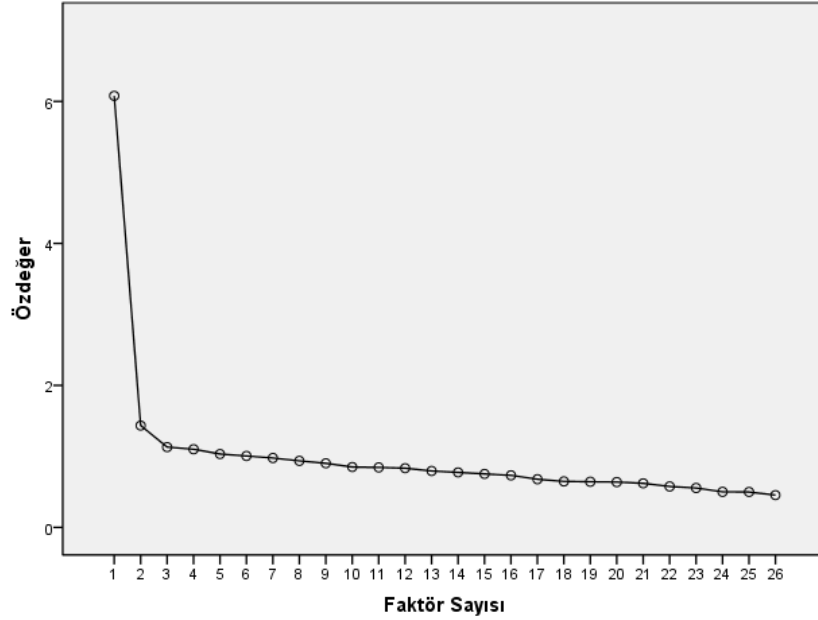
Testlerin tek boyutluluğunu kontrol edebilmek için faktör analizi yapılmıştır. Bundan önce faktör analizinin veri yapısına uygunluğunu test etmek için Kaiser- Meyer – Oklin (KMO) katsayısı ve Barlett küresellik (sphericity) testi incelenebilir (Büyüköztürk, 2013). Faktör analizinde Form X verilerine ait KMO katsayısı 0,92 bulunup 0,60 'tan büyük olduğundan ve Barlett Küresellik testi manidar bulunduğundan ($p<0,01$) bu testin faktör analizine uygunluğuna karar verilmiştir. Form Y için yapılacak faktör analizine uygunluğu içinde Form Y verilerine ait KMO katsayısı 0,91 bulunup 0,60 'tan büyük olduğundan ve Barlett Küresellik testi manidar bulunduğundan ($p<0,01$) bu testin de faktör analizine uygunluğuna karar verilmiştir. Faktör analizi tekniklerinden temel eksenler (principal axis) seçilmiştir.

Faktör analizi sonucunda Form X için birinci faktöre ait özdeğer ile ikinci faktöre ait özdeğerin 3 katından fazla olması ($\lambda_1/\lambda_2=7,25$), birinci faktörden sonra grafikte ivmeli, hızlı düşüşün yaşandığından dolayı 26 maddelik testin MTK'nın tek boyutluluk varsayımını karşıladığının kanıtı olarak kabul edilmiştir (Büyüköztürk, 2013). Ayrıca ilk özdeğer toplam varyansın en az %20'sini açıkladığından (%20,8) ötürü de testin tek boyutlu olduğu kanısına varılmıştır (Reckase, 1979).

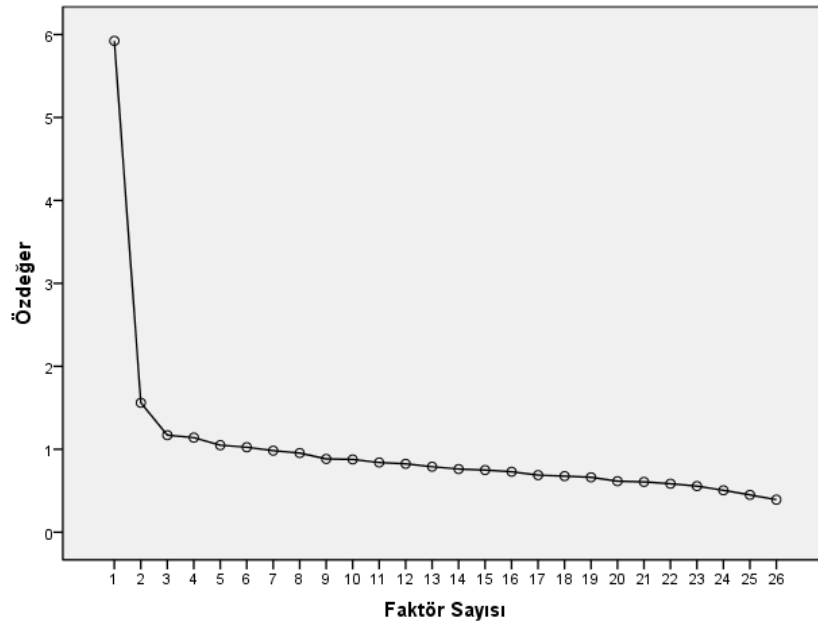
Faktör analizi sonucunda Form Y için birinci faktöre ait özdeğer ile ikinci faktöre ait özdeğerin 3 katından fazla olması ($\lambda_1/\lambda_2=5,93$), birinci faktörden sonra grafikte ivmeli, hızlı düşüşün yaşandığından dolayı 26 maddelik bu testin de MTK'nın tek boyutluluk varsayımını karşıladığının kanıtı olarak kabul edilmiştir. Ayrıca ilk özdeğer toplam varyansın en az %20'sini açıkladığından (%20,3) ötürü de testin tek boyutlu olduğu kanısına varılmıştır (Reckase, 1979).

Şekil 16'da Form X'e ait öz değerlerin saçılma grafiği Şekil 17'de Form Y'ye ait öz değerlerin saçılma grafiği vardır. Bu grafikler incelendiğinde birinci faktör altındaki maddelere ait özdeğerden sonra gelen ikinci faktör altındaki maddelere ait özdeğer ile önemli bir düşüş olduğu görülebilir. İkinci faktörden sonra meydana gelen kırılmaların

önemli düşüş göstermemesi ve diğer faktörlerle özdeğerlerinin birbirine çok yakın olmasından dolayı grafiğin yatay bir şekilde devam ettiği görülebilir. Bu yüzden de her iki formun da ölçülen özellik bakımından tek boyutlu olduğu söylenebilir.



Şekil 16. Form X'in öz değerleri saçılma grafiği



Şekil 17. Form Y'nin öz değerleri saçılma grafiği

Ayrıca testin tek boyutlu olup olmadığının belirlenmesinde Hambleton ve Swaminathan, (1985), dikotomus olarak puanlanan maddeler için maddeler arası tetrakorik korelasyon katsayısı matrisine dayalı olarak yapılan faktör analizinin Pearson korelasyon katsayısının

özel bir durumu olan phi korelasyon katsayısı matrisine dayalı olarak yapılan faktör analizine göre daha uygun olduğunu belirtmiştir. Bu bilgiye dayalı olarak da çalışmada kullanılan testlerin tek boyutluluğu varsayımının testi için R programı ‘psych’ paketi ile tetrakorik korelasyon matrisine dayalı açımlayıcı faktör analizi yapılmıştır. Analiz sonucunda her iki form için fit değeri 0,96 bulunmuştur. Analiz sonuçlarına göre her iki form için de maddelerin ölçülen özelliği baskın olarak tek boyutlu olarak ölçtüğünü göstermiştir. Form X için baskın faktöre ait özdeğer 9,73 ve açıklanan varyans oranı %37 iken Form Y için baskın faktöre ait özdeğer 10,17 ve açıkladığı varyans oranı %39 olarak bulunmuştur. Her iki form için yapılan açımlayıcı faktör analizi sonuçları Tablo 6 ve Tablo 7’de sunulmuştur. Maddelere ilişkin faktör yük değerlerine bakıldığında Form X için iki maddeye ait faktör yük değerinin 0,30’dan küçük olduğu görülmüştür ($\lambda_4 = 0,19$ ve $\lambda_{14} = 0,25$). Form Y içinse yalnızca bir maddeye ait yük değerinin 0,30’dan küçük bulunmuştur ($\lambda_2 = 0,07$). Yani bu maddeler ait oldukları formdaki baskın faktörü iyi temsil edememiş yorumu yapılabilir. Bu maddeler incelenmek istense de TIMSS değerlendirmelerinde sonraki değerlendirmelerde kullanılan “trend items” olarak adlandırılan maddeler içinde yer aldığından dolayı maddelere bakılamamıştır. Bu maddeler çıkarılınca yapılacak analizlerde bilgi kaybına uğrama olabileceğinden çalışma kapsamında tutularak analizlere devam edilmiştir.

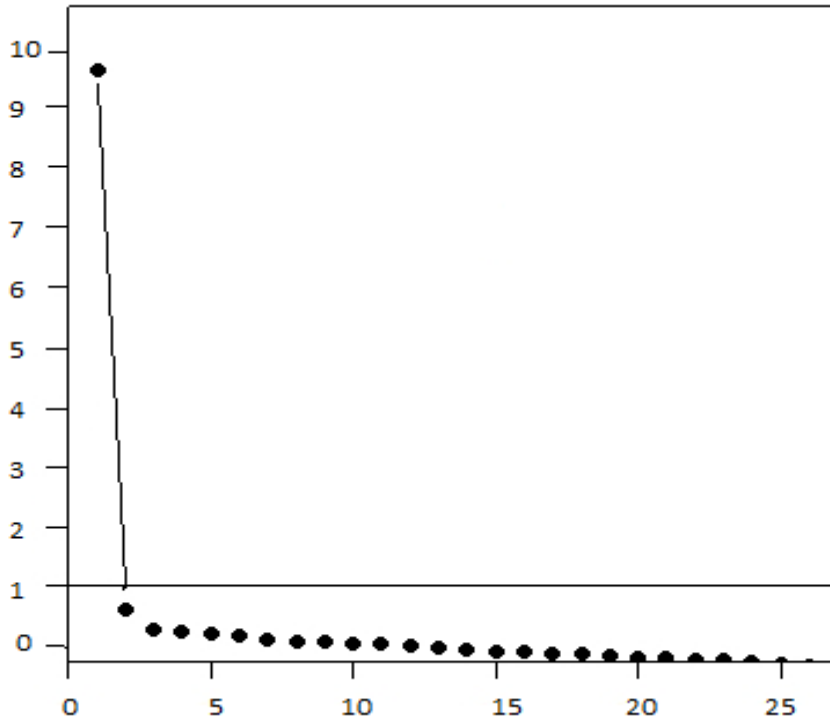
Tablo 7. Form X'e ait Açıklayıcı Faktör Analizi Sonuçları

Madde	Faktör Yüğü
1	0.47
2	0.61
3	0.65
4	0.19
5	0.65
6	0.70
7	0.56
8	0.70
9	0.64
10	0.57
11	0.84
12	0.48
13	0.78
14	0.25
15	0.37
16	0.67
17	0.74
18	0.67
19	0.64
20	0.51
21	0.55
22	0.69
23	0.73
24	0.67
25	0.58
26	0.65
Özdeğer	9,73
Açıklanan varyans	%37

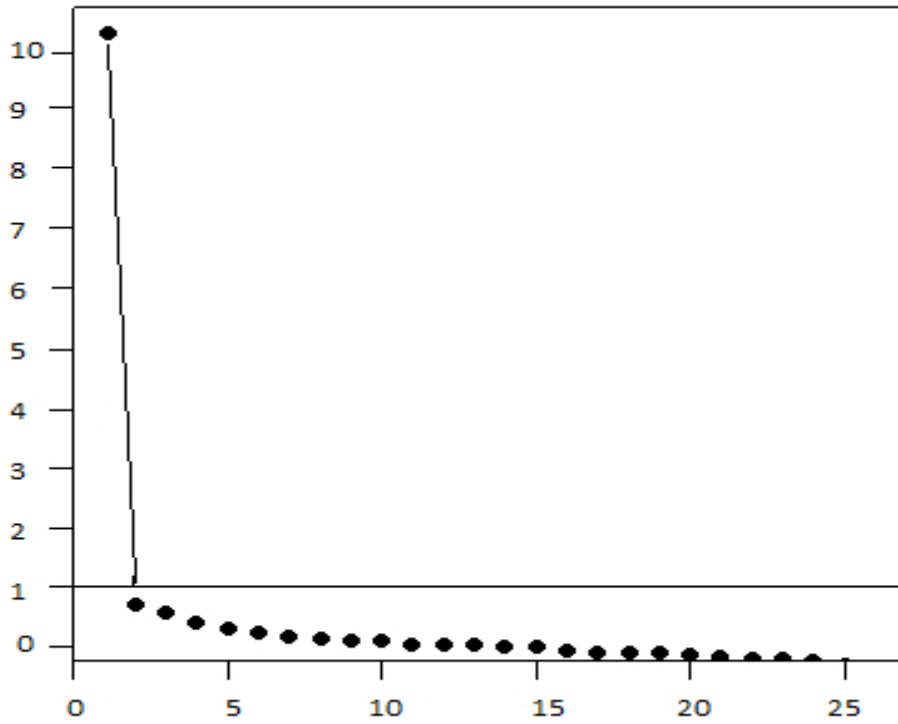
Tablo 8. Form Y'ye ait Açıklayıcı Faktör Analizi Sonuçları

Madde	Faktör Yüğü
1	0.70
2	0.07
3	0.31
4	0.59
5	0.70
6	0.65
7	0.66
8	0.50
9	0.54
10	0.77
11	0.78
12	0.66
13	0.65
14	0.62
15	0.84
16	0.84
17	0.78
18	0.54
19	0.47
20	0.52
21	0.49
22	0.59
23	0.71
24	0.36
25	0.74
26	0.58
Özdeğer	10,17
Açıklanan varyans	%39

Şekil 18'de ve Şekil 19'da sırasıyla Form X ve Form Y'ye ait öz değerler grafiğı yer almaktadır.



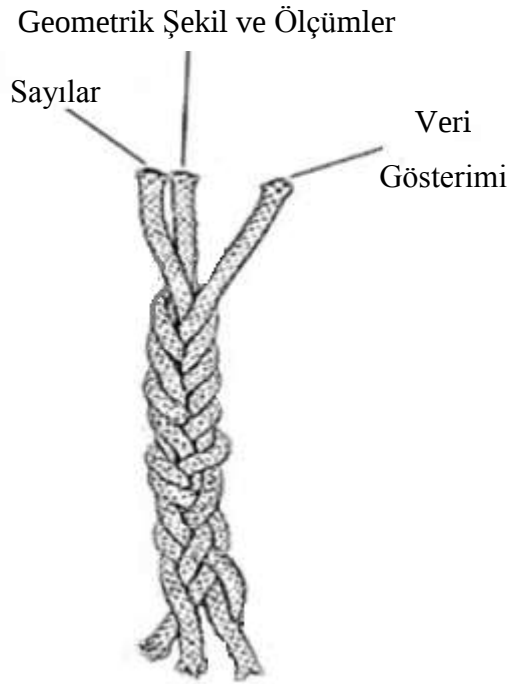
Şekil 18. Form X'e ait tetrakorik faktör analizi öz değer grafiği



Şekil 19. Form Y'ye ait tetrakorik faktör analizi öz değer grafiği

Şekil 18 ve 19'da görüldüğü gibi ilk öz değerden sonra keskin düşüş meydana gelmiştir. Buna göre her iki formda tek boyutludur yorumu yapılabilir.

Burg'e (2008) göre "boyutluluk" u yeniden "ortaya çıkarılabilir boyutluluk (detectable dimensionality)" olarak tanımlamada boyutluluğun iki önemli nitelemesi birlikte bütünleşir; psikolojik anlamı ve istatistiksel uyumu. Yalnızca, bu iki niteleme bir diğerini desteklediğinde gerçek test yapısı değerlendirilebilir, yorumlanabilir ve belki daha da önemlisi eğitimsel durumlar için çıkarımlar açıklığa kavuşturulabilir. Her bir matematik başarı testi NCTM (National Council of Teachers of Mathematics) çerçevesinde önerilen beş içerik ipliğinin (content strands) birbirleriyle ilişkili olmasını gerektirir, standart diye de atfilen bu içerik iplikleri ; (1) Sayı ve işlem, (2) Geometri, (3) Cebir/ Örüntüler ve Fonksiyon, (4) Veri analizi ve olasılık , (5) Ölçme'dir. (NCTM'den aktaran Burg, 2008). Burg'ün (2008) yaptığı çalışmaya göre bu standartlar ile boyutluluk arasında ilişki bulunamamıştır. Bu standartların iç içe ve ilişkili olmasından dolayı matematik başarı testlerinin psikolojik anlamıyla da tek boyutlu olduğu kanısına varılabilir. Aşağıda Burg'ün (2008) çalışmasından esinlenerek TIMSS 2011 4. sınıf matematik başarı testinde konu alanları; "sayılar" , "geometrik şekil ve ölçümler" ve "veri gösterimi" olarak belirlenenmiş ve bu konu alanlarının ilişkisi Şekil 20'deki gibi betimlenmek istenmiştir.



Şekil 20. TIMSS 2011 4. sınıf matematik başarı testini oluşturan konu alanları
(Burg, 2008'ten uyarlanmıştır.)

MTK'nın varsayımlarından tek boyutluluk sayılıtısının yeterince karşılanmaması sonucu yerel bağımsızlık sayılıtısı da beraberinde bozulacaktır. Yerel bağımsızlık sayılıtısının bozulması ise yanlış parametre kestirimleriyle sonuçlanacaktır (Sünbül ve Erkuş, 2013).

Yerel Bağımsızlık

Lord' a (1980) göre, yerel bağımsızlık ek bir sayılıtı olmayıp tek boyutluluğun bir getirisidir. Tek boyutluluk varsayımı karşılandığında, yerel bağımsızlık varsayımının da karşılandığı düşünülmektedir (Hambleton ve Swaminathan, 1985). Form X ve Form Y içinde bireylerin ölçülen özelliğinin (matematik başarısının) tek boyutlu olduğu sonucunun bir getirisi olarak testi alan bir bireyin, bir maddeye doğru yanıt verme olasılığının, testte bulunan diğer maddelere verdiği cevaplardan etkilenmeyeceği varsayılmıştır. Yerel bağımsızlığın testi için De Mars (2010)'ın önerdiği, maddelerin çift olarak yerel bağımsızlığını test eden Yen (1984)'in Q_3 istatistiği R programı ile hesaplanmıştır. Yen (1984) bu ikili karşılaştırmalardaki korelasyonun 0,20 den büyük çıkmasının problem olduğunu belirtmiştir (aktaran De Mars, 2010). EK 2'de Form X ve Form Y için Yen (1984)'in Q_3 istatistiğinin değerleri bulunmaktadır, bütün değerler 0,2'den küçük olduğundan yerel bağımsızlık her iki form için sağlanmıştır.

Madde Karakteristik Eğrisinin Monotonik Artışı

Madde karakteristik eğrisi, bir test tarafından ölçülen özellik ya da yetenek üzerinde madde puanının lineer olmayan regresyon fonksiyonudur (Hambleton ve Swaminathan, 1985). Bu sayılıtıya göre bireyin yeteneği arttıkça, maddeyi doğru cevaplama olasılığının da monoton olarak artış göstermesi gerekir. Çalışma kapsamındaki her iki formun maddelerinin 3PLM altındaki madde karakteristik eğrileri incelenmiş ve her bir maddeye ait karakteristik eğrilerin monoton artan fonksiyona sahip olduğu görülmüştür (EK. 4).

Model Veri Uyumu

Belirli bir durumda üç logistic modelden hangisinin kullanılacağı sorusu çok karmaşık bir durumdur. Model seçimdeki bir faktör seçilecek modelin varsayımlarının ne kadar duruma uygun olduğu ile ilgilidir denilebilir. Örneğin; bir ve iki parametrelili lojistik modeller ortada bir tahminin olmadığını varsaymaktadır. 3PLM (Birnbau, 1968) ise tahmine yer verebilir ve bu tahmin çoktan seçmeli ve doğru-yanlış maddelerde bir olasılık olarak düşünülmelidir.

Bu bağlamda üç parametre modelin diğer modellerin her ikisine de üstünlüğü vardır (Crocker ve Algina, 1986). 3PLM model, madde ayıricılıklarının farklılık gösterdiği ve öğrencilerin çoktan seçmeli maddeleri şans (tahmin) yolu ile doğru cevaplandırabileceğini varsayar (Karkee ve Wright, 2004). Tablo 2 incelendiğinde a parametresi her iki formda 1 ile 3 arasında değişmekte ve formlardaki maddelerin c parametreleri ortalamaları 0,2 çıkmaktadır. Böylece bu araştırmada model olarak 3PLM model seçilebilir.

Ayrıca formların -2loglikelihood değerlerine göre de model veri uyumu test edilebilir. Aşağıda her iki form için ve üç model için -2ln (-2loglikelihood) değerleri verilmiştir.

Tablo 9. Form X ve Form Y için -2ln Değerleri

	1PLM	2PLM	3PLM
Form X	15202,6117	14958,2219	14871,7297
Form Y	14289,9106	14070,18	13995,0991

Tablo 9'dan görüldüğü gibi Form X için modelde bulunan parametre sayısı arttıkça -2ln değerleri düşmektedir. Kline (2005)'a göre modelde bulunan parametre sayısı arttıkça -2ln değerleri azalıyorsa model uyumu parametre sayısı fazla olan modelde daha iyi uyum olduğu kabul edilebilir. Bu yüzden model veri uyumu en iyi 3PLM'de sağlanmıştır denilebilir. Form X için; 2PLM ile 1PLM arasındaki fark (15203-14958=245) χ^2 değeri olarak adlandırılabilir. 2PLM'de 1PLM'ye göre 26 madde parametre kestirimi daha yapılmıştır. Bu yüzden serbestlik derecesi 26 olarak seçildiğinde 245 değeri 0,05 manidarlık düzeyinde χ^2 kritik tablo değerinden büyük olduğundan 2PLM, 1PLM'ye göre daha uyumludur. 3PLM ile 2PLM arasındaki fark ile bulunan χ^2 değeri de tablo değerinden büyük olduğundan 3PLM daha uyumlu bulunmuştur (14958-14871= 87>38,89). Bu test Form Y için yapıldığında 1PLM ile 2PLM arasındaki χ^2 değeri 220'dir ve bu değer 26 serbestlik derecesinde 0,05 manidarlık düzeyinde χ^2 kritik tablo değerinden büyük olduğundan 2PLM, 1PLM modele göre daha uyumludur. 3PLM model ile 2PLM model arasındaki fark ile bulunan χ^2 değeri de tablo değerinden büyük olduğundan 3PLM model daha uyumlu bulunmuştur (14070-13995=75).

Her bir model altında -2ln değerlerine bakılarak yapılan değerlendirmeden sonra formlardaki maddelerin ilgili model altındaki uyumlarına da bakılmıştır. R programı ile yapılan madde

uyumu analizi sonuçlarına göre uyum sağlamayan (χ^2 değerlerinin $p=0,0001$ manidarlık düzeyi ile) madde sayısı Tablo 10’da özetlenmiştir (EK 5).

Tablo 10. Form X ve Form Y için Uyum Sağlamayan Madde Sayıları

	1PLM’deki uyumsuz madde sayısı	2PLM’deki uyumsuz madde sayısı	3PLM’deki uyumsuz madde sayısı
Form X	3	1	0
Form Y	3	1	2

Tablo 10’daki veriler ışığında her iki form için üç model arasında madde uyumu sağlamayan madde sayısı en fazla 1PLM’ye aittir. Form X için bakıldığında madde uyumu en iyi 3PLM’de sağlanmıştır. Ancak Form X için bakıldığında madde uyumu en iyi 2PLM’de sağlanmıştır. Ancak yapılan madde uyumu testi ile -2ln değerlerinin değerlendirmesinin ortak sonucu olarak 3PLM’nin model veri uyumunun sağladığı kanısına ulaşılabılır.

Araştırmada Kullanılan Ölçek Dönüştürme Yöntemleri

Bu araştırmada kullanılan ölçek dönüştürme yöntemleri giriş bölümünde anlatılan MTK’ya dayalı OO, OS, HA ve SL yöntemleridir. Yetenek ve madde parametre kestirimleri dikotomus maddeler için tek boyutlu MTK modellerinden 3PLM model ile kestirilmiştir. Böylelikle ölçek dönüştürme sonucunda elde edilen yetenek parametresi ile yatay eşitleme işlemine uygun ve gerçek puan eşitlemeye dayalı eşitlemeler yapılabilir. Bu araştırmada da ölçek dönüştürme işleminden sonra gerçek puan eşitlemeye dayalı eşitleme yapılmıştır.

Değerlendirme Kriteri

Bu araştırmada test eşitleme yöntemlerinden elde edilen eşitleme hatalarının karşılaştırılması için kriter olarak “RMSD (Root mean square difference)” seçilmiştir. Kullanılışlı bir istatistiksel indeks olan RMSD bireylerin gerçek yetenek düzeyi ile almadıkları testten kestirilen yetenek düzeyleri arası farka dayanmaktadır (Wakler ve Kim, 2010; Harris ve Crouse, 1993):

Aşağıda bu yöntemin gerçek yetenek düzeyi ve kestirilmiş yetenek düzenin kullanıldığı matematiksel gösterimi sunulmuştur.

$$RMSD = \sqrt{\frac{\sum_i f_i (\theta^* - \theta)^2}{\sum_i f_i}}$$

θ : Gerçek yetenek düzeyi

θ^* : Kestirilmiş yetenek düzeyi

f_i : Yetenek düzeyindeki frekans

Bu eşitlikte bireylerin kestirilen yetenek düzeyi ile gerçek yetenek düzeyi arasındaki fark bulunur ve kareleri alınır. Her bir birey için bu işlem yapıldıktan sonra elde edilen değerler toplanır. Bu toplam, yetenek düzeylerine ait frekansa bölünür ve karekökü alınır. Yöntemleri karşılaştırırken çıkan sonuç küçüldükçe uygulanan yöntemin daha az hatalı eşitleme sonuçları yaptığı anlamı çıkarılabilir. Bu araştırmada bireylerin kestirilen yetenek düzeyleri ve gerçek yetenek düzeyleri arasında yapılan işlemlerle daha az hatalı olan yöntem bulunmak istenmiştir.



BÖLÜM IV

BULGULAR VE YORUMLAR

Araştırmanın bu bölümünde, araştırma sorularına ait bulgular ve bu bulgulara ait yorumlara yer verilmiştir.

Araştırma Sorusu 1'e ait Bulgu ve Yorumu:

1.1. TIMSS 2011 dördüncü sınıf düzeyi ortak madde içeren iki matematik alt testinin puanlarının ortalama-ortalama yöntemi ile yetenek parametresine yönelik ölçek dönüşümü sonucunda elde edilen dönüştürme katsayısı ve sabiti nedir?

OO ile ölçek dönüştürme denklemini kurmak için elde edilen dönüştürme katsayısı ve dönüştürme sabitine ilişkin bulgu Tablo 11'de özetlenmiştir.

Tablo 11: OO ile Ölçek Dönüştürme İçin Elde Edilen Dönüştürme Katsayısı ve Sabiti

Ölçek Dönüştürme Yöntemi	A	B
OO (Ortalama-ortalama)	1,010839	-0,058427

Bulunan katsayıların yüzde bir basamağına göre yuvarlanmış hali ile OO ile ölçek dönüştürme lineer denklemi;

$$\theta_{Xi} = 1,01\theta_{Yi} - 0,06$$

şeklinde oluşturulabilir. Bu denklemden elde edilen bulgu ile Form X' i alan her bir bireyin yeteneğinin 1,01 katından 0,06 çıkarılmasıyla elde edilen yetenek değeri Form Y'yi alan bireyler ile aynı ölçek üzerinde konumlandırılabilir.

1.2. TIMSS 2011 dördüncü sınıf düzeyi ortak madde içeren iki matematik alt testinin puanlarının ortalama-standart sapma yöntemi ile yetenek parametresine yönelik ölçek dönüşümü sonucunda elde edilen dönüştürme katsayısı ve sabiti nedir?

OS ile ölçek dönüştürme denklemini kurmak için elde edilen dönüştürme katsayısı ve dönüştürme sabitine ilişkin bulgu Tablo 12’de özetlenmiştir.

Tablo 12: OS ile Ölçek Dönüştürme İçin Elde Edilen Dönüştürme Katsayısı ve Sabiti

Test eşitleme yöntemi	A	B
OS (Ortalama-standart sapma)	1,123007	-0,150068

Bulunan katsayıların yüzde bir basamağına göre yuvarlanmış hali ile OS ile ölçek dönüştürme lineer denklemi;

$$\theta_{Xi} = 1,12\theta_{Yi} - 0,15$$

şeklinde oluşturulabilir. Bu denklemden elde edilen bulgu ile Form X’ i alan her bir bireyin yeteneğinin 1,12 katından 0,15 çıkarılmasıyla elde edilen yetenek değeri Form Y’yi alan bireyler ile aynı ölçek üzerinde konumlandırılabilir.

1.3. *TIMSS 2011 dördüncü sınıf düzeyi ortak madde içeren iki matematik alt testinin puanlarının Haebara ile yetenek parametresine yönelik ölçek dönüşümü sonucunda elde edilen dönüştürme katsayısı ve sabiti nedir?*

HA ile ölçek dönüştürme denklemini kurmak için elde edilen dönüştürme katsayısı ve dönüştürme sabitine ilişkin bulgu Tablo 13’de özetlenmiştir.

Tablo 13: HA ile Ölçek Dönüştürme İçin Elde Edilen dönüştürme Katsayısı ve Sabiti

Test eşitleme yöntemi	A	B
HA (Haebara)	1,094939	-0,130086

Bulunan katsayıların yüzde bir basamağına göre yuvarlanmış hali ile HA ile ölçek dönüştürme lineer denklemi;

$$\theta_{Xi} = 1,09\theta_{Yi} - 0,13$$

şeklinde oluşturulabilir. Bu denklemden elde edilen bulgu ile Form X’ i alan her bir bireyin yeteneğinin 1,09 katından 0,13 çıkarılmasıyla elde edilen yetenek değeri Form Y’yi alan bireyler ile aynı ölçek üzerinde konumlandırılabilir.

1.4. *TIMSS 2011 dördüncü sınıf düzeyi ortak madde içeren iki matematik alt testinin puanlarının Stocking-Lord ile yetenek parametresine yönelik ölçek dönüşümü sonucunda elde edilen dönüştürme katsayısı ve sabiti nedir?*

SL ile ölçek dönüştürme denklemini kurmak için elde edilen dönüştürme katsayısı ve dönüştürme sabitine ilişkin bulgu Tablo 14’ de özetlenmiştir.

Tablo 14: SL ile Ölçek Dönüştürme İçin Elde Edilen dönüştürme Katsayısı ve Sabiti

Test eşitleme yöntemi	A	B
SL (Stocking&Lord)	1,075369	-0,123209

Bulunan katsayıların yüzde bir basamağına göre yuvarlanmış hali ile SL ile ölçek dönüştürme lineer denklemi;

$$\theta_{Xi} = 1,08\theta_{Yi} - 0,12$$

şeklinde oluşturulabilir. Bu denklemden elde edilen bulgu ile Form X’ i alan her bir bireyin yeteneğinin 1,08 katından 0,12 çıkarılmasıyla elde edilen yetenek değeri Form Y’yi alan bireyler ile aynı ölçek üzerinde konumlandırılabilir.

Araştırma Sorusu 2’ye ait Bulgu ve Yorum:

2. *TIMSS 2011 4. sınıf düzeyi ortak madde içeren iki matematik alt testinin puanlarının OO, OS, HA ve SL ile bulunan eşitleme denklemlerinden yararlanarak; yetenek bakımından eşitleme hataları karşılaştırıldığında RMSD ortalamaları nedir?*

Farklı yöntemler ile eşitleme sonucunda elde edilen RMSD ortalamaları Tablo 15’te özetlenmiştir. Bu tabloya göre en az eşitleme hatası (RMSD) veren yöntemden en fazla eşitleme hatası veren yöntemine göre şöyle sıralanabilir; ortalama-ortalama yöntemi, Stocking-Lord yöntemi, Haebara yöntemi ve ortalama-standart sapma yöntemi.

Tablo 15: Farklı Yöntemler ile Ölçek Dönüştürme Sonucunda Yetenek Parametresine Yönelik Olarak Elde Edilen RMSD Ortalamaları

Test Eşitleme Yöntemleri	RMSD Ortalaması
OO (Ortalama-ortalama)	0,058557
OS (Ortalama-standart sapma)	0,159290
HA (Haebara)	0,135347
SL (Stocking&Lord)	0,125809

RMSD kriterine göre, bu araştırma sorularının bulguları ışığında çalışmada kullanılan veriler için yetenek parametrelerine yönelik olarak en uygun ölçek dönüştürme yöntemi ortalama-ortalama yöntemi olarak kabul edilebilir. Uygunluk sırasına devam edilirse Stocking-Lord, Haebara ve ortalama-standart sapma şeklinde devam eder.

Araştırma Sorusu 3'e ait Bulgu ve Yorum :

3. *TIMSS 2011 dördüncü sınıf düzeyi ortak madde içeren iki matematik alt testinin puanlarının yetenek parametresine yönelik bulunan en uygun ölçek dönüştürme yönteminden yararlanarak MTK gerçek puan eşitleme yöntemi kullanılması sonucu elde edilen eşitlenmiş puanlar nasıldır?*

Tablo 16'da OO ile ölçek dönüşümü yapılarak aynı ölçek üzerine getirilen tetalar ile MTK gerçek puan eşitleme yöntemi kullanılması sonucu elde edilen eşitlenmiş puanlar hesaplanmıştır.

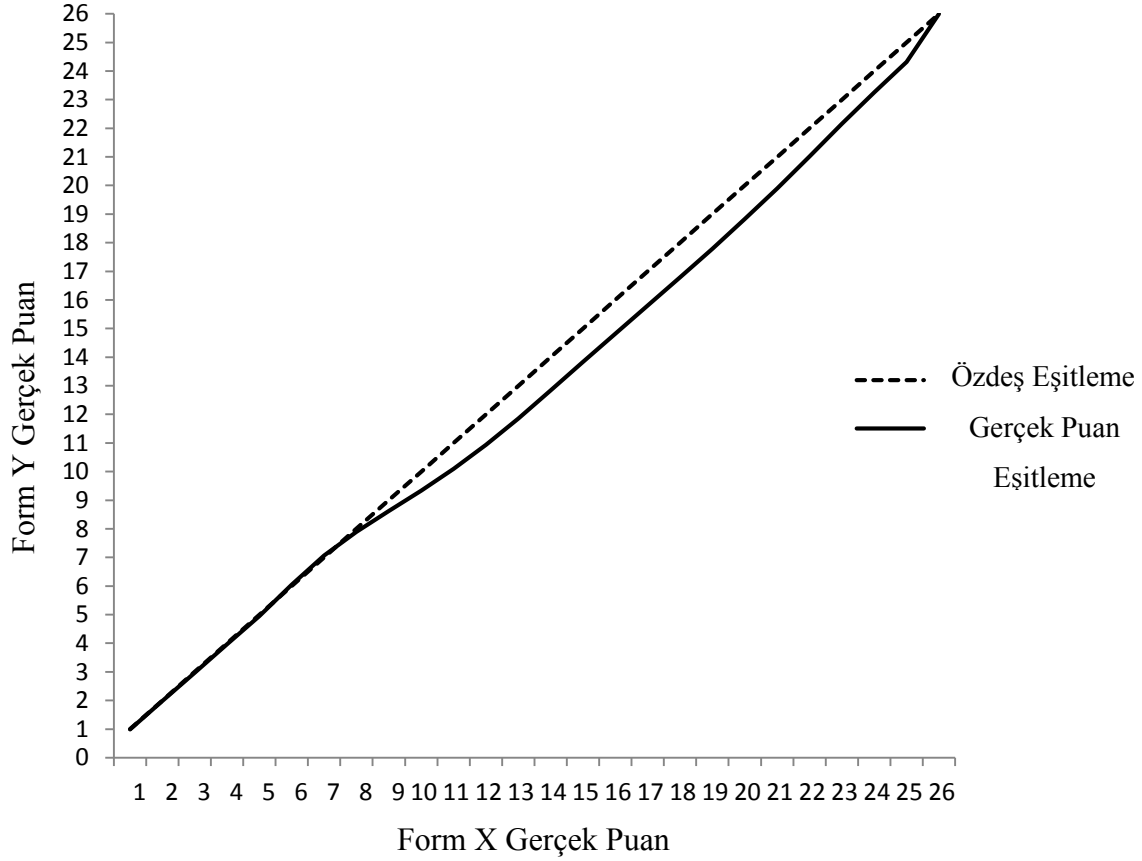
Tablo 16: MTK Gerçek Puan Eşitleme Yöntemi Kullanılması Sonucu Elde Edilen Eşitlenmiş Puanlar

Form X Puanı	Teta Eşdeğeri	Form Y'deki Puan Eşdeğeri
0	-----	0
1	-----	0,991
2	-----	1,982
3	-----	2,972
4	-----	3,963
5	-----	4,954
6	-0,828	6,050
7	-0,459	7,066
8	-0,245	7,893
9	-0,083	8,621
10	0,057	9,337
11	0,187	10,096
12	0,311	10,934
13	0,433	11,856
14	0,552	12,835
15	0,669	13,830
16	0,784	14,819
17	0,901	15,799
18	1,022	16,784
19	1,154	17,787
20	1,303	18,822
21	1,477	19,900
22	1,682	21,021
23	1,915	22,161
24	2,178	23,259
25	2,513	24,316
26	-----	26

Bu tabloda çok düşük Form X puanları (beş ve daha düşük puanları) ve Form X'in 26 puanı için teta eşdeğerleri verilmemiştir. Elde edilen c parametreleri kestirimleri toplamı 5,185 çıkmıştır (EK 3). Bu yüzden Form X'in tam sayı puanlarının gerçek puan eşdeğerleri 5 puan ve altındaki puanlar için tanımlanmamıştır.

Tablo 16'ya bakılarak MTK gerçek puan eşitleme yöntemine göre Form X puanının Form Y'deki değerleri bulunabilir. Örneğin Form X'den alınan 18 puanın Form Y'deki eşdeğerini bulmaya odaklanılsın. İlk olarak Form X'te 18 gerçek puanının ilişkili olduğu teta bulunmalı, bu teta değerinin de Form Y de hangi puana denk geldiği bulunmalıdır. Bu tablo ile sürecin sonunda Form X'e ait her bir tam sayı puanı için PIE programı çıktısı ile

özetlenmiştir. Form X için 18 puanı 1,022 teta düzeyine denk gelmekte ve Form Y’de bu teta düzeyi de 16,784 puanına eşdeğer olmaktadır. Tablodaki puanlar ve eşdeğerleri incelendiğinde Form Y’nin Form X’e göre daha zor olduğu yorumu yapılabilir. Ayrıca Şekil 21’deki özdeş eşitleme ile MTK gerçek puan eşitleme kullanarak Form X gerçek puanlarının kestirilen Form Y gerçek puan eşdeğerleri grafiği de yorum yapılmasını sağlayabilir.



Şekil 21. MTK gerçek puan eşitleme kullanarak Form X gerçek puanlarının kestirilen Form Y gerçek puan eşdeğerleri

Bu grafiğe göre gerçek puan eşitleme eğrisi özdeş eşitleme çizgisinden yukarıda olduğundan Form Y’nin Form X’e göre daha zor olduğu yorumu yapılabilir.

BÖLÜM V

SONUÇ VE TARTIŞMA

Araştırmanın bu kısmında elde edilen sonuçlara ve bu sonuçlara ilişkin önerilere yer verilmiştir.

Araştırma Sorusu 1'e ait Sonuç ve Tartışma:

OO ile ölçek dönüştürme sonucu dönüştürme katsayısı bire, dönüştürme sabiti ise sıfıra oldukça yakın çıkmıştır. Yani kestirilen tetalar, gerçek tetalara oldukça yakın çıkacaktır. Böylelikle kestirilen teta ile gerçek tetanın arasındaki fark az çıkacak ve eşitleme hatası oldukça düşük çıkacaktır. Bu durum literatürdeki bazı araştırmacıların bulgularıyla paralellik göstermiştir. Baker ve Al-Karni (1991) yaptığı simülasyon çalışmasında (orta yetenekli bireyler ile ayırıcılık ve güçlük seviyesi düşük maddelerde) OO ile ölçek dönüşümü yaptığı simülasyon çalışmasında eşitleme katsayısını 1'e yakın ($A=1,0909$), eşitleme sabitini 0'a oldukça yakın bulmuştur ($B=0,0156$). Kilmen (2010) 500 kişilik ve benzer yetenek dağılımı koşulunda OO ile eşitleme yaptığı simülasyon çalışmasında da eşitleme katsayısını 1'e oldukça yakın ($A=1,002$), eşitleme sabitini ise 0'a oldukça yakın bulmuştur ($B=-0,007$). Gök (2012) iki parametrelili lojistik model altında 1000 kişilik 10 maddesi ortak 30 maddeli iki formu eşitleme sürecinde OO ile eşitleme katsayısını 1'e çok yakın ($A=1,008$) bulmuştur. Ancak Gök (2012)'ün bulduğu eşitleme sabiti kestirilen teta ile gerçek tetanın arasındaki farkı daha belirgin kılmıştır ($B=0,471$).

OS ile eşitleme sonucu eşitleme katsayısının bire olan uzaklığı OO ile yapılan eşitleme sonucu bulunan eşitleme katsayısından daha büyük çıkmıştır. Eşitleme sabitinin sıfıra uzaklığı OO ile yapılan eşitleme sonucu bulunan eşitleme sabitinden daha fazla çıkmıştır. Bu demek oluyor ki kestirilen tetalar, gerçek tetalara OO'ya göre daha uzak çıkacak; yani eşitleme hatası daha büyük çıkacaktır. Kilmen (2010), 500 kişilik ve benzer yetenek dağılımı koşulunda OS ile eşitleme yaptığı simülasyon çalışmasında eşitleme katsayısını OO' ile

yapılan eşitlemeye göre 1'den daha uzak eşitleme katsayısına ($A=0,979$), eşitleme sabitini ise 0'a daha uzak bulmuştur ($-0,013$).

HA ile yapılan eşitleme sonucu bulunan eşitleme katsayısı; OS ile yapılan eşitleme sonucunda bulunan eşitleme katsayısına göre 1'e daha yakın, OO ile yapılan eşitleme sonucunda bulunan eşitleme katsayısına göre ise 1'den daha uzaktır. HA ile yapılan eşitleme sonucu bulunan eşitleme sabiti ise; OS ile yapılan eşitleme sonucunda bulunan eşitleme sabitine göre 0'a daha yakın, OO ile yapılan eşitleme sonucunda bulunan eşitleme asbitine göre 0'dan daha uzaktır. Bu durumda kestirilen tetalar, gerçek tetalara OO'ya göre daha uzak; OS'ye göre daha yakın çıkacaktır. Yani HA için bulunacak eşitleme hatası daha OO'ya göre daha büyük çıkması, OS'ye göre daha küçük çıkması olası bir durumdur.

SL ile yapılan eşitleme sonucu bulunan eşitleme katsayısı; OS ve HA ile yapılan eşitleme sonucunda bulunan eşitleme katsayısına göre 1'e daha yakın, OO ile yapılan eşitleme sonucunda bulunan eşitleme katsayısına göre ise 1'den daha uzaktır. SL ile yapılan eşitleme sonucu bulunan eşitleme sabiti ise; OS ve HA ile yapılan eşitleme sonucunda bulunan eşitleme sabitine göre 0'a daha yakın, OO ile yapılan eşitleme sonucunda bulunan eşitleme sabitine göre 0'dan daha uzaktır. Bu durumda kestirilen tetalar, gerçek tetalara OO'ya göre daha uzak; OS'ye ve HA'ya göre daha yakın çıkacaktır. Yani SL için bulunacak eşitleme hatası OO'ya göre daha büyük çıkması, OS'ye ve HA'ya göre daha küçük çıkması olası bir durumdur.

Araştırma Sorusu 2'ye ait Sonuç ve Tartışma:

TIMSS 2011 dördüncü sınıf düzeyi ortak madde içeren iki matematik alt testini alan bireylerin tetalarını, dört farklı yöntemle ölçek dönüştürmeye tâbi tutularak aynı ölçek üzerine yerleştirilmesini sağlayan lineer denklemler bulunduktan sonra, bu yöntemler arasında yetenek düzeyleri bakımından ve RMSD kriteri ile karşılaştırma yapılmıştır. Buna göre ortalama RMSD'si en düşük olan yöntem OO iken en yüksek olan OS olarak bulunduğu, araştırmada kullanılan veriler için en uygun ölçek dönüştürme denklemi OO iken en az uygun olan yöntem OS olarak bulunmuştur. Baker ve Al-Karni (1991) yaptıkları araştırma sonucunda OO'nun en iyi sonuçlar verdiğini bulmuştur. Bu araştırmanın bulgusu Baker ve Al-Karni (1991)'nin yaptığı çalışmanın bulgularının paralelinde bulunsa da literatür genelinin paralelinde bulunmamıştır. Literatür genelinde yöntemler en uygun olandan başlamak üzere sıralandığında SL, HA, OO ve OS olarak dizilir. Ancak bu

çalışmanın verileri ile yapılan eşitleme de OO sıralamayı bozarak en iyi yöntem olarak başa geçmiştir. Literatür ile bu bulgunun örtüşmemesinin bir sebebi olarak formlarda bulunan ortak madde sayısının oranının fazla olması olabilir. Ortak maddeli eşdeğer olmayan gruplar deseni altında yürütülen bu araştırmada formların ortak madde oranı %50'den fazladır (14/26). Ölçek dönüştürme işleminin hesaplamalarında ortak maddelerin testin tüm maddelerine oranı ile ilgili çalışmalara bakıldığında %50 den fazla ortak madde içeren testlerin eşitlenmesi ya da ölçek dönüşümü ile ilgili herhangi bir çalışma ile karşılaşılmamıştır. Örneğin; Tian (2011) MTK altında SL ve eş zamanlı kalibrasyon yöntemlerini karşılaştırırken ortak madde oranının etkisini araştırmak için ortak madde oranını %10, %20 ve %30 olan koşullar altında çalışmasını sürdürmüştür. Benzer olarak Uysal (2014), Meng (2012), Hanson ve Béguin (1999), Kilmen (2010) yaptıkları çalışmalarda ortak madde oranını en fazla %33,3 olma durumunu araştırmışlardır. Ortak madde oranının fazla olması, bu çalışmadaki ölçek dönüştürme yöntemlerinin uygunluğunun sırasını etkileyen bir faktör olabilir.

Önceki araştırmalar (Hanson ve Béguin, 1999; 2002; Kolen ve Brennan , 2004) karakteristik eğri yöntemlerinin moment yöntemlerine göre tercih edilmesi gerektiğini göstermektedir. Ancak bu çalışmanın sonucunda en iyi iki yöntemden birisi moment yöntemi diğeri karakteristik eğri yöntemi bulunmuştur. Bu yüzden bu çalışmanın sonucunda moment ve karakteristik eğrisi yöntemleri arasında karşılaştırılmaya gidilmemiştir.

Karakteristik eğri yöntemlerinin kendi içinde SL yönteminin HA yöntemine göre daha uygun bulunması Hanson ve Béguin (1999), Kilmen (2010), Kim ve Kolen (2006), Speron (2009), Gök (2012), Uysal (2014)' ın yaptıkları çalışmaların paralelinde bulunmuştur.

Moment yöntemleri ile kendi aralarında karşılaştırma yapıldığında ise OO ile yapılan ölçek dönüştürme OS ile yapılan ölçek dönüştürmeye göre daha az hata içerdiğinden OO daha uygun bulunmuştur. Bu sonuç da Hanson ve Béguin (1999), Kilmen (2010), Kim ve Kolen (2006), Speron (2009), Gök (2012), Uysal (2014)'ın yaptıkları çalışmalarla örtüşmüştür.

Araştırma Sorusu 3'e ait Sonuç ve Tartışma:

Araştırmada kullanılan veriler için ölçek dönüştürme yöntemlerinden en uygun yöntem ortalama-ortalama yöntemi bulunduğundan MTK gerçek puan eşitleme işlemini yapmak için OO'nun dönüştürme katsayısı ve dönüştürme sabitini kullanarak aynı ölçek üzerine koyulan

madde parametreleri kullanılmış ve eşitleme süreci tamamlanmıştır. Bu eşitleme işleminin sonucunda ise Form Y'nin Form X'e göre daha zor olduğu sonucuna ulaşılmaktadır. Bu sonuca göre eğer çalışmada kullanılan veriler üzerinde bir karşılaştırılmaya gidilecekse eşitlenmiş puanlar kullanılarak karşılaştırma yapılmalıdır.

Özetle, bu çalışmada TIMSS 2011 dördüncü sınıf düzeyinden seçilen ortak maddeli iki matematik formuna ait iki gruptaki bireylerin verilerinin toplama biçimine göre ortak maddeli eşdeğer olmayan gruplar deseni uygun bulunmuş, eşitleme çalışmalarından önce MTK varsayımları test edilmiştir. Ardından bu formları alan bireylerin puanlarının MTK'ya dayalı test eşitleme yapılması için; parametrelerin aynı ölçek üzerinde konumlandırılması gereği ortaya çıkmıştır. Madde Tepki Kuramı'na dayalı ölçek dönüştürme yöntemlerinden ortalama-ortalama, ortalama-standart sapma, Stocking-Lord ve Haebara ile ölçek dönüştürme denklemleri bulunmuş, seçilen RMSD kriteri ile, yetenek kestirimine yönelik olarak, hesaplanan hatalara göre en az hata veren yöntem olan ortalama-ortalama yöntemi diğer üç yöntemle göre daha uygun bulunmuştur. Bu yöntemden yararlanarak MTK gerçek puan eşitleme sonucunda da Form Y'nin Form X'e göre daha zor olduğu bulunmuştur. Böylece araştırma kapsamındaki formları alan bireyler arasında bu forma dayalı olarak karşılaştırma yapılmak istenirse eşitleme yapılması gerekmektedir.

BÖLÜM VI

ÖNERİLER

Araştırmanın bu kısmında çalışmanın sonuçlarına dayanarak, araştırmacılara ve uygulayıcılara bazı önerilerde yer verilmiştir.

Bu araştırmadan elde edilen sonuçlara göre,

1. TIMSS değerlendirmelerinde eş zamanlı kalibrasyon yöntemi kullanılmaktadır (Mullis vd, 2009). Araştırmada kullanılan iki formdan gelen parametreleri aynı ölçek üzerine koymak için eş zamanlı ölçek dönüştürmeden başka, ayrı ölçek dönüştürme yöntemlerinden araştırma kapsamında çalışılan dört yöntemden en uygun bulunan OO'nun kullanılması önerilebilir.
2. İki formun MTK gerçek puan eşitlemesinde Form Y'nin Form X'e göre daha zor olduğu bulunmuştur. Çalışma kapsamındaki veriler için bireyler arasında doğru bir karşılaştırma yapılabilmesi için MTK gerçek puan test eşitleme işleminin uygulanması önerilebilir.
3. Bu çalışmanın koşullarına benzer koşullar içinde (grupların yetenek dağılımları benzer, formlarda bulunan ortak madde oranı %50'den fazla) olan eşitleme çalışmaları için OO ile ölçek dönüştürme yapılması önerilebilir.

Ayrıca bundan sonra yapılacak araştırmalar içinse aşağıdaki öneriler değerlendirilebilir.

1. Bu çalışma MTK modelleri ile sınırlandırmıştır. Benzer kapsamda bir çalışma KTK modelleri ile de yapıp elde edilen bulgular bu araştırmada elde edilen bulgularla karşılaştırılabilir.
2. Test eşitleme sürecinde parametre kestirimi için BILOG-MG 3.00, ölçek dönüştürme denklemleri için ST programları kullanılmıştır. Farklı parametre kestirimi ve farklı eşitleme programları kullanılarak bu araştırmanın sonuçları ile karşılaştırılabilir.

3. TIMSS verilerinden aynı numaralı kitapçıklar üzerinden ancak örneklem büyüklüğü farklı olan veri setleri ile eşitleme yapılarak bu çalışma ile karşılaştırıldığında örneklem büyüklüğünün MTK'ya dayalı eşitleme yöntemleri üzerindeki etkisi araştırılabilir.
4. TIMSS verilerinden aynı numaralı kitapçıklar üzerinden ancak farklı ülkelerden olan veri setleri ile eşitleme yapıp bu çalışma ile karşılaştırıldığında farklı güçlük düzeyindeki grupların MTK'ya dayalı eşitleme yöntemleri üzerindeki etkisi araştırılabilir.
5. Bu araştırmada eşitleme yöntemlerini karşılaştırmada kriter olarak RMSD kullanılmıştır. Karşılaştırmada tek bir değerlendirme ölçütü yerine birden fazla kriter kullanılarak karşılaştırma yapılabilir.
6. TIMSS verilerinden aynı numaralı kitapçıklar üzerinden ancak ortak madde sayısı azaltıldıktan sonra eşitleme yapıp bu çalışma ile karşılaştırıldığında ortak madde sayısının madde sayısına oranının MTK'ya dayalı eşitleme yöntemleri üzerindeki etkisi araştırılabilir.

KAYNAKLAR

- Aiken, L. R. (2000). *Psychological testing and assesment*. Boston: Allyn and Bacon.
- Angoff, W. H. (1984). *Scales, norms and equivalent scores*. Princeton, NJ: Educational Testing Service
- Baker, F. B. (2001). *The basics of item response theory*. USA: ERIC.
- Baker, F.B. and Al-Karni, A. (1991). A Comparison of Two Procedures for Computing IRT Equating Coefficients. *Journal of Educational Measurement*, 28 (2), 147-162.
- Braun, H. I. & Holland, P. W. (1982). *Observed score test equating: a mathematical analysis of some ETS equating procedures*. New York: Academic Press
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. F. M. Lord & M. R. Novick (Ed.), *Statistical theories of mental test scores içinde* (pp. 397–479). MA: Addison-Wesley.
- Bränberg, K. (2010). Observed score equating with covariates. *Print & Media*, 41(1), 1-24.
- Burg, S. (2008, Mart). *An investigation of dimensionality across grade levels and effects on vertical linking for elementary grade mathematics achievement tests*. National Council on Measurement in Education'da sunulmuş bildiri, New York City.
- Büyüköztürk, Ş., Çokluk, Ö. & Köklü, N. (2012). *Sosyal bilimler için istatistik* (11. baskı). Ankara: Pegem.
- Büyüköztürk, Ş. (2013). *Sosyal bilimler için veri analizi el kitabı* (18. baskı). Ankara: Pegem.
- Cohen, A. S. & Kim, S. (1998). An investigation of linking methods under the graded response model. *Applied Measurement in Education*, 22(2), 116-130.

- Cook, L. L. (2007). Practical problems in equating test scores: A practitioner's perspective. *Linking and aligning scores and scales* içinde. (pp. 73-88). Springer New York. N. J. Dorans, M. Pommerich, & P. W. Holland (Ed.), *Linking and aligning scores and scales* içinde (pp.73-87). New York: Springer.
- Cook, L. L. & Eignor, D. R. (1985). An investigation of the feasibility of applying item response theory to equate achievement tests 1, 2, 3. New Jersey: ETS, Princeton.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Harcourt Brace Jovanovich College.
- Çetin, E. (2009). *Dikey ölçeklemede klasik test ve madde tepki kuramına dayalı yöntemlerin karşılaştırılması*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- De Gruijter, D.N.M., Van der Kamp, L. J. T. (2008). *Statistical test theory for the behavioral sciences*. London: Chapman & Hall/CRC, Taylor & Francis Group.
- DeMars, C. (2010). *Item response theory*. USA: Oxford University
- Doğan, N., & TEZBAŞARAN, A. A. (2003). Klasik test kuramı ve örtük özellikler kuramının örneklem bağlamında karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 25(25).
- Dorans, N. J., Moses, T. P., & Eignor, D. R. (2010). Principles and practices of test score equating. *ETS Research Report Series*, 2010(2), 41.
- Dorans, N. J., Pommerich, M., & Holland, P. W. (Eds.). (2007). *Linking and aligning scores and scales*. New York: Springer.
- Dvorak, R. L. (2009). *A comparison of Kernel equating to the test characteristic curve method*. Doktora Tezi, The University Of Nebraska, Lincoln.
- Embretson, S. E. & Reise, S. P. (2013). *Item response theory for psychologists*. New Jersey: Psychology.
- Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme*. Ankara: Pegem Akademi.
- French, D. J. (1996). *The utility of Stocking & Lord's equating procedure for equating norm-referenced and criterion-referenced tests with both dichotomous and polytomous components*. Doktora Tezi, University of Texas, Texas.

- Gök, B. (2012). *Denk olmayan gruplarda ortak madde deseni kullanılarak madde tepki kuramına dayalı eşitleme yöntemlerinin karşılaştırılması*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Gök, B.& Kelecioğlu, H. (2014). Denk olmayan gruplarda ortak madde deseni kullanılarak madde tepki kuramına dayalı eşitleme yöntemlerinin karşılaştırılması. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 10(1).
- Haebara, T. (1980). Equating logistic ability scales by a weighted least squares methods. *Japanese Psychological Research*, 22, 144-149.
- Haertel, E. (2004, Mayıs). *The behavior of linking items in test equating*. National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education Information Studies 'te sunulmuş bildiri, Los Angeles.
- Hagge, S. L. (2010). The impact of equating method and format representation of common items on the adequacy of mixed-format test equating using nonequivalent groups. Doktora tezi, University of Iowa, Iowa.
- Hagge, S. L., Liu, C., He, Y., Powers, S. J., Wang, W., & Kolen, M. J. (2011). A comparison of IRT and traditional equipercentile methods in mixed-format equating. *M. J. Kolen & W. Lee (Ed.), Mixed-format tests: Psychometric properties with a primary focus on equating (volume 1) (CASMA monograph number 2.1)* içinde (pp. 19 50). Iowa: The University of Iowa.
- Hambleton, R. K. (1989) Principles and selected applications of item response theory. Linn, R. L. (ed.) *Educational Measurement* içinde. (p 147–200). New York: American Council on Education
- 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.
- 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*, 99, 8.

- Hanson, B. A. & Béguin, 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., & Crouse, J. D. (1993). A study of criteria used in equating. *Applied Measurement in Education*, 6(3), 195-240.
- He, Y. (2011). *Evaluating equating properties for mixed-format tests* (Order No. 3461151). Available from ProQuest Dissertations & Theses Global. (879634637). 10/03/2015 tarihinde <http://search.proquest.com/docview/879634637?accountid=11054> sayfasından erişilmiştir.
- Jurich, D. P., DeMars, C. E., & Goodman, J. T. (2012). Investigating the impact of compromised anchor items on IRT equating under the nonequivalent anchor test design. *Applied Psychological Measurement*, 36(4), 291-308.
- Karasar, N. (2009). *Bilimsel araştırma yöntemi* (19. Baskı). Ankara: Nobel
- Karkee, T. B. & Wright, K. R. (2004, Nisan). *Evaluation of linking methods for placing threeparameter logistic item parameter estimates onto a one-parameter scale*. Annual Meeting of the American Educational Research Association'da sunulmuş bildiri, San Diego, California.
- Kelecioğlu, H. (1994). *Öğrenci Seçme Sınavı Puanlarının Eşitlenmesi Üzerine Bir Çalışma*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Kilmen, S. (2010). *Madde Tepki Kuramı'na dayalı test eşitleme yöntemlerinden kestirilen eşitleme hatalarının örneklem büyüklüğü ve yetenek dağılımına göre karşılaştırılması*. Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Kim, H. Y. (2014). A comparison of smoothing methods for the common item nonequivalent groups design. Doktora Tezi, The University of Iowa, Iowa.
- 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. (2006). An extension of four IRT linking methods for mixed-format tests. *Journal of Educational Measurement*, 43(1), 53-76.
- Kim, S. & Cohen, A. S. (2002). A comparison of linking and concurrent calibration under the graded response model. *Applied Psychological Measurement*, 26(1), 25-41.

- Kline, T. J. (2005). *Psychological testing: A practical approach to design and evaluation*. London: Sage.
- Kolen, M. J. & Brennan, R. L. (2014). *Test equating, scalling, and linking*. (3. baskı). USA: Springer.
- Lee, W. C.& Ban, J. C. (2010). A comparison of IRT linking procedures. *Applied Measurement in Education*, 23: 23–48.
- Livingston, S. A. (2004). *Equating test scores (without IRT)*. Princeton, NJ: Educational Testing Service.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. N. J.: Lawrence Erlbaum.
- Loyd, B.H. & Hoover, H.D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, 17, 179-193.
- Lord, F.M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, 14, 139-160.
- Marco, G. L., Petersen, N. S., & Stewart, E. E. (1983). A test of the adequacy of curvilinear score equating models. D. White (Ed.), *New Horizons in Testing*. New York: Academic.
- Meng, H (2007). *A comparison study of IRT calibration methods for mixed-format tests in vertical scaling*. Doktora Tezi, University of Iowa, Iowa.
- Meng, Y. (2012). *Comparison of Kernel equating and item response theory equating methods* Doktora Tezi, Massachusetts University, Massachusetts.
- Millî Eğitim Bakanlığı Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü (2014). *TIMSS 2011 ulusal matematik ve fen raporu:4. sınıflar*.
- Muraki, E., Hombo, C. M., & Lee, Y. W. (2000). Equating and linking of performance assessments. *Applied Psychological Measurement*, 24(4), 325-337.
- Mullis, I. V., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 Assessment Frameworks*. International Association for the Evaluation of Educational Achievement. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

- Speron, E. (2009). *A comparison of metric linking procedures in item response theory*. Doktora Tezi, Illinois Institute of Technology, Chicago.
- Stocking, M. L. & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201–210.
- Sünbül, Ö, & Erkuş, A. (2013). Madde parametrelerinin değişmezliğinin çeşitli boyutluluk özelliği gösteren yapılarda madde tepki kuramına göre incelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 9(2), 378-398
- ÖSYM. YDS Kılavuzu. (2014). 15/01/2015 tarihinde <http://www.osym.gov.tr/dosya/1-71452/h/2014-ydskilavuz.pdf> sayfasından erişilmiştir.
- Petersen, N. S., Cook, L. L. & Stocking, M. L. (1983). IRT versus conventional equating methods: A comparative study of scale stability. *Journal of Educational Statistics*, 8, 137-156.
- Peterson, J. L. (2014). *Multidimensional item response theory observed score equating methods for mixed-format tests*. Doktora Tezi, The University of Iowa, Iowa.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: results and implications. *Journal of Educational Statistics*, 4, 207-230.
- Tan, Ş. (2015). Küçük örneklemelerde beta 4 ve polynomial loglineer öndüzgünleştirme ve kübik eğri öndüzgünleştirme metotlarının uygunluğu. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 35(1), 123-151.
- Tian, F. (2011). *A comparison of equating/linking using the stocking-lord method and concurrent calibration with mixed-format tests in the non-equivalent groups common-item design under IRT*. Doktora Tezi, Boston Collage, Boston.
- Tong, Y. & Kolen, M. J. (2005). Assessing equating results on different equating criteria. *Applied Psychological Measurement*, 29(6), 418-432.
- Uysal, İ. (2014). *Madde tepki kuramına dayalı test eşitleme yöntemlerinin karma modeller üzerinde karşılaştırılması*. Yüksek Lisans Tezi, Abant İzzet Baysal Üniversitesi Eğitim Bilimleri Enstitüsü, Bolu.
- Walker, M. E., & Kim, S. (2010). Examining Two Strategies to Link Mixed-Format Tests Using Multiple-Choice Anchors. Research Report. ETS RR-10-18. *Educational Testing Service*.

- Wang, W. (2013). *Mixed-format test score equating: Effect of item-type multidimensionality, length and composition of common-item set, and group ability*. Doktora Tezi, University of Iowa, Iowa.
- Wolf, R. (2013). *Assessing the Impact of Characteristics of the Test, Common-items, and Examinees on the Preservation of Equity Properties in Mixed-format Test Equating*. Doktora Tezi. University of Pittsburgh, Pittsburgh.
- Yıldırım, Y. (2015). *Derecelendirilmiş tepki modeli temelli parametre kestiriminde normalliğin ihlalinin ölçme kesinliğine etkisi*. Yüksek Lisans Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.







EK 1- ST 2.0 Output Dosyası

ST 2.0

Input file: C:\Users\Tuba\Desktop\st_for_pc_gui\st_wg_v20\ST_TIMSS.txt

Number of common items: 14

Item Parameter Means

	a	b	c
New Form	1.9918	0.8170	0.2053
Old Form	1.9704	0.7674	0.2204

Item Parameter Standard Deviations

	a	b	c
New Form	0.4569	0.9428	0.0969
Old Form	0.5469	1.0588	0.1091

Transformation Functions

	Stocking-Lord	Haebara	Mean/Mean	Mean/Sigma
Intercept	-0.123209	-0.130086	-0.058427	-0.150068
Slope	1.075369	1.094939	1.010839	1.123007

EK 2 - Form X ve Form Y için Yen'in Q₃ Testi

Form X'in yerel bağımsız madde çiftleri	Q3	Form X'in yerel bağımsız madde çiftleri	Q3	Form X'in yerel bağımsız madde çiftleri	Q3
1	0,0166847	39	0,0037769	77	0,0501483
2	-0,0333182	40	-0,0911222	78	0,0470148
3	-0,0687876	41	-0,0142356	79	0,0469765
4	0,0036786	42	-0,0580577	80	-0,0061416
5	0,0535868	43	-0,0363219	81	-0,0037269
6	-0,0551374	44	0,0607531	82	-0,0259351
7	-0,0113420	45	-0,0540263	83	-0,0220876
8	0,0059333	46	0,0175376	84	-0,0634322
9	0,0131147	47	-0,0639825	85	-0,0184396
10	-0,1019736	48	-0,0225106	86	-0,0861899
11	-0,0198507	49	-0,0324183	87	-0,0070262
12	0,0035845	50	0,0396385	88	-0,0141138
13	0,0288137	51	-0,0550831	89	0,0730188
14	-0,0602045	52	-0,0374581	90	0,0119835
15	-0,0520674	53	0,0316682	91	-0,0224790
16	-0,0974072	54	-0,0709696	92	0,0043777
17	0,0378113	55	-0,0372057	93	0,0014961
18	-0,0223223	56	0,0684684	94	-0,0374086
19	-0,0534487	57	0,0068503	95	-0,0571832
20	0,0175220	58	-0,0092724	96	0,0013923
21	-0,0711504	59	-0,1062855	97	-0,0159312
22	0,0513276	60	0,0024170	98	-0,0673583
23	0,0277868	61	-0,0298412	99	-0,0398786
24	-0,0788037	62	-0,0478935	100	-0,0448167
25	-0,0570961	63	-0,0352559	101	-0,0143713
26	0,0612422	64	-0,0630177	102	-0,1190574
27	0,0047566	65	-0,0612717	103	0,0025908
28	-0,0513858	66	-0,1169081	104	-0,0354789
29	-0,0582607	67	-0,0292886	105	-0,0323998
30	-0,0442519	68	-0,0988381	106	-0,0685901
31	-0,0189478	69	-0,0584505	107	-0,0706037
32	-0,0477780	70	0,0189125	108	-0,0458932
33	-0,0281196	71	-0,0256593	109	-0,0116419
34	-0,0651755	72	-0,0802879	110	-0,0610790
35	-0,0476495	73	-0,0072119	111	0,0006899
36	-0,0982578	74	-0,0485625	112	-0,0312425
37	0,0315158	75	0,0367485	113	-0,1222789
38	-0,0112053	76	-0,0632011		

Form X'in yerel bağımsız madde çiftleri	Q3	Form X'in yerel bağımsız madde çiftleri	Q3	Form X'in yerel bağımsız madde çiftleri	Q3
114	0,0889884	155	-0,0595780	197	-0,0010802
115	-0,0429596	157	0,0219484	198	-0,0785102
116	-0,0353995	158	-0,1044307	199	0,0072565
117	-0,0038332	159	-0,0475146	200	0,0242996
118	-0,0428887	160	-0,0533300	201	-0,0939743
119	-0,0124812	161	-0,0026629	202	0,0426658
120	-0,0924354	162	-0,0910177	203	-0,0807890
121	-0,0774937	163	-0,0705433	204	-0,0387564
122	-0,0700584	164	-0,0097312	205	-0,0901870
123	-0,0978905	165	-0,0683675	206	-0,0090241
124	-0,0165458	166	-0,0219599	207	-0,1244804
125	0,0103423	167	-0,0380543	208	-0,0320111
126	-0,0523874	168	-0,0647341	209	-0,0704198
127	-0,1159976	169	-0,0892222	210	-0,1179596
128	-0,0449068	170	-0,0862310	211	-0,0355553
129	-0,0079660	171	-0,0834168	212	-0,0850740
130	-0,0539071	172	-0,1015556	213	-0,0701955
131	-0,1014687	173	-0,0246721	214	-0,0322710
132	-0,0101542	174	-0,1036147	215	-0,1394701
133	-0,0283244	175	-0,0677795	216	-0,1072578
134	-0,0190073	176	-0,0622902	217	-0,1063597
135	-0,0266383	177	-0,0173213	218	-0,0064112
136	0,0254955	178	-0,0205235	219	-0,0263658
137	-0,0405490	179	-0,0169579	220	-0,0589542
138	-0,0825698	180	0,0646601	221	0,0126490
139	-0,0019519	181	-0,0550052	222	-0,0368598
140	-0,0466641	182	0,0589379	223	-0,0315828
141	-0,0632925	183	0,0646438	224	-0,0322019
142	-0,0276779	184	0,0063160	225	0,0016243
143	0,0428051	185	0,0275994	226	0,0091132
144	-0,0688603	186	-0,0421482	227	-0,0303984
145	-0,0864291	187	-0,0319281	228	-0,0382504
146	-0,1006297	188	-0,0192462	229	-0,0961508
147	0,0190382	189	-0,0988856	230	0,0466574
148	-0,0624572	190	-0,0392568	231	-0,0445977
149	-0,0096387	191	-0,0637038	232	0,0008798
150	-0,0107743	192	-0,0462231	233	0,0564334
151	-0,0617263	193	-0,0077502	234	0,0050272
152	0,0617859	194	0,0186129	235	0,0358883
153	-0,0307547	195	-0,1204112	236	-0,0593910
154	-0,0415731	196	-0,0946476	237	0,0042966

Form X'in yerel bağımsız madde çiftleri	Q3	Form X'in yerel bağımsız madde çiftleri	Q3	Form X'in yerel bağımsız madde çiftleri	Q3
238	-0,0325398	279	-0,0115431	320	-0,0082913
239	-0,0067898	280	-0,0558837	321	-0,0279756
240	-0,0029206	281	-0,0025101	322	-0,0618593
241	0,0000176	282	-0,0067911	323	-0,0406244
242	-0,0116551	283	0,0175522	324	0,0103545
243	-0,0652734	284	-0,0797908	325	0,0917714
244	0,0059055	285	-0,0895304		
245	0,0099941	286	-0,0769694		
246	-0,0651028	287	-0,1128392		
247	-0,0902140	288	-0,0866745		
248	0,0160045	289	-0,0288379		
249	0,0566551	290	-0,0228905		
250	0,0354292	291	-0,0132035		
251	0,0465670	292	-0,0687347		
252	0,0355255	293	-0,0386343		
253	-0,0141877	294	-0,0518694		
254	-0,0405666	295	-0,0363781		
255	-0,0568141	296	-0,0688495		
256	-0,0011205	297	0,0304831		
257	-0,0361770	298	0,0168491		
258	-0,0301061	299	0,0074935		
259	-0,0476211	300	-0,0534573		
260	0,0040875	301	-0,1335458		
261	0,0667492	302	-0,0216314		
262	-0,0000762	303	-0,0168419		
263	0,0129522	304	-0,0181390		
264	-0,0411802	305	0,0004101		
265	-0,0433232	306	0,0355521		
266	0,0254299	307	-0,1307976		
267	-0,0496577	308	-0,0157360		
268	0,0228092	309	-0,1481019		
269	0,0124639	310	-0,0222119		
270	-0,0175182	311	0,0159419		
271	-0,0624701	312	-0,0029424		
272	-0,0127453	313	-0,0189350		
273	-0,0440616	314	-0,0519723		
274	-0,0426215	315	-0,0142566		
275	-0,0584513	316	-0,0738145		
276	-0,0158379	317	0,0808478		
277	0,0762329	318	-0,0052957		
278	0,0426088	319	0,0054459		

Form Y'nin yerel bağımsız madde çiftleri	Q3	Form Y'nin yerel bağımsız madde çiftleri	Q3	Form Y'nin yerel bağımsız madde çiftleri	Q3
1	0.0143219185	42	0.0563615871	84	-0.1222902676
2	-0.0544093945	44	0.0347062922	85	-0.0489637567
3	0.1037042659	45	-0.0512151416	86	0,0219484987
4	0.1002696659	46	-0.0740355172	87	0.0594494144
5	0.0187230407	47	-0.0563675162	88	-0.0194916267
6	0.0223053313	48	-0.0441054258	89	0.0410404490
7	0.0284154953	49	0.0207693613	90	-0.0027468526
8	0.0169837535	50	0.1067287616	91	0.0774776658
9	0.0036238653	51	0.0764513312	92	0.0093573583
10	0.0135423341	52	-0.0007422818	93	0.0909985963
11	0.0114611922	53	-0.0662312318	94	-0.0347950130
12	-0.0186049585	54	-0.0457996036	95	-0.0506725174
13	0.0552335680	55	0.0663476441	96	0.0766115627
14	-0.1175497720	56	0.0082267960	97	-0.0031552660
15	-0.1175497720	57	-0.0087361533	98	0.0090063534
16	0.0318233079	58	-0.0123045358	99	0.0300682615
17	-0.0461714210	59	0.0469628941	100	-0.0234668696
18	0.0143622755	60	0.0488085673	101	-0.0455092432
19	0.0325462163	61	-0.0262078556	102	0.0622128576
20	0.0073112058	62	-0.0262078556	103	0.0183270324
21	0.0294649981	63	-0.0030978168	104	-0.0838118210
22	0.0421370361	64	-0.0044208077	105	-0.0838118210
23	0.0285712677	65	-0.0100255066	106	-0.0497933607
24	0.0099212984	66	0.0102240531	107	-0.0036616058
25	0.0146847644	67	-0.0158254444	108	-0.0149216465
26	0.0624064157	68	-0.0306237866	109	0.0154273460
27	-0.0332852454	69	-0.0111609046	110	0.1215162202
28	0.0963859725	70	0.0084911755	111	0.0228384062
29	-0.0156001467	71	-0.0408883235	112	-0.0226283570
30	-0.0365561401	72	-0.0481031384	113	0.0289313792
31	-0.0280864372	73	-0.0061996183	114	-0.0700666990
32	0.0084032427	74	0.0719556561	115	0.0745196969
33	-0.0891144960	75	-0.0434976756	116	-0.0396862539
34	0.0092607062	76	0.0956468646	117	0.0431552807
35	-0.0533907576	77	-0.0147612623	118	0.0151843450
36	0.0362967656	78	0.1249332967	119	-0.0044756164
37	-0.0742550842	79	-0.0002031656	120	-0.0112807884
38	-0.0215408781	80	-0.0452221705	121	0.0388277582
39	-0.0215408781	81	0.0231427884	122	0.0120068612
40	0.0173744692	82	-0.0237332600	123	-0.0209270414
41	-0.0127256214	83	-0.1222902676	124	-0.0795765994

Form Y'nin yerel bağımsız madde çiftleri	Q3	Form Y'nin yerel bağımsız madde çiftleri	Q3	Form Y'nin yerel bağımsız madde çiftleri	Q3
125	-0.0795765994	166	0.0184233729	207	-0.0994138146
126	0.0278823853	167	0.0251364223	208	0.0043412024
127	0.0819027060	168	0.1327316308	209	-0.1320277142
128	0.0483294229	169	0.0407848867	210	-0.1320277142
129	0.0061996183	170	-0.0437512787	211	0.0225912548
130	-0.0145084464	171	0.0597842329	212	-0.0462786076
131	0.0130935376	172	0.0561976183	213	0.0486727045
132	0.0572245626	173	0.0576865351	214	-0.0065875286
133	-0.0180186824	174	0.0830527822	215	0.045913218
134	-0.0357998601	175	0.0663995595	216	-0.0056647972
135	0.0191641291	176	0.0176282877	217	0.0142284927
136	-0.0108383648	177	-0.0724307618	218	-0.0670346349
137	-0.0151319280	178	-0.1050659709	219	0.0235057889
138	0.1254052936	179	-0.1050659709	220	-0.0091494201
139	0.0295902257	180	-0.0128314154	221	0.0610926285
140	0.0487111707	181	0.0535875780	222	0.0271010494
141	-0.0454379472	182	-0.0271601699	223	-0.1685109945
142	0.0172517485	183	-0.0344994801	224	-0.1685109945
143	-0.1370890900	184	-0.0414826931	225	0.0346732787
144	-0.1370890900	185	0.0113620509	226	0.0421845656
145	0.0441766152	186	-0.0242758480	227	0.0079915218
146	-0.0154991413	187	-0.0340389937	228	0.0407298525
147	0.0143226986	188	-0.0568258138	229	0.0358031123
148	-0.0151266716	189	-0.0061491358	230	0.0682981385
149	0.0988199775	190	-0.0034259347	231	0.0905771535
150	0.0559525387	191	0.0985095848	232	0.0000517387
151	0.1806943326	192	0.0158975828	233	0.0765215333
152	0.0593223148	193	0.0919056056	234	0.0008462882
153	0.0317072390	194	-0.1665388384	235	-0.0399655869
154	0.1020553116	195	-0.1665388384	236	-0.1692609859
155	-0.0044115162	196	-0.0133712711	237	-0.1692609859
156	0.0163658875	197	-0.0634721332	238	0.0329015781
157	-0.0242431669	198	0.0552560233	239	0.0589900378
158	0.0875960145	199	-0.0307441319	240	0.0233177190
159	0.0713091168	200	0.0143816924	241	0.0310831112
160	0.0143494685	201	0.0582691323	242	0.0220325967
161	-0.1414679456	202	0.3623743567	243	0.0369515117
162	-0.1414679456	203	-0.0191061604	244	0.0921413798
163	0.0530942656	204	0.0379666904	245	0.0284661521
164	-0.0235033015	205	0.1269153871	246	0.0230692239
165	-0.0677524768	206	0.0101525350	247	0.0149190829

Form Y'nin yerel bağımsız madde çiftleri	Q3	Form Y'nin yerel bağımsız madde çiftleri	Q3
248	-0.1551991111	289	0.0066050997
249	-0.1551991111	290	-0.0365792257
250	0.0405141034	291	0.0142255935
251	0.0334646368	292	-0.0867534893
252	-0.0180854884	293	-0.0112996671
253	-0.0286654054	294	0.0331180728
254	0.0596346757	295	0.0151122060
255	0.0037239446	297	0.0104153437
256	0.1325117980	298	0.0179629838
257	0.0503229427	299	-0.0060269017
258	0.0295773613	300	-0.0758905691
259	0.0016156365	301	0.0402317321
260	0.0574994324	302	0.0357693613
261	-0.0651570976	303	0.0033534732
262	-0.0511077521	304	0.0435627399
263	-0.0504439094	305	0.0094440272
264	-0.0465622516	306	0.0701574073
265	-0.0878578586	307	-0.0004093077
266	-0.1316179716	308	0.0584986408
267	-0.2305218147	309	0.0292586650
268	-0.0490902696	310	0.0266461224
269	-0.0912051344	311	-0.0029383252
270	-0.0266159448	312	0.0962835404
271	-0.0651570976	313	-0.0066161290
272	-0.0511077521	314	-0.0272217221
273	-0.0504439094	315	0.0346308705
274	-0.0465622516	316	0.0980819096
275	-0.0878578586	317	-0.0343647359
276	-0.1316179716	318	0.1024951935
277	-0.2305218147	319	0.0228418162
278	-0.0490902696	320	-0.0248672564
279	-0.0912051344	321	0.0253491134
280	-0.0266159448	322	0.0779377549
281	0.1165928870	323	0.0512538490
282	0.0959666004	324	-0.0139923368
283	0.0171858505	325	0.0866298590
284	-0.0145268493		
285	0.0653325955		
286	0.0229909785		
287	-0.0261737763		
288	0.0107471827		

EK 3- PIE 2.0 Input ve Output Dosyası

PIE 2.0 INPUT DOSYASI

IRT gerçek puan esitleme

Parametreler 00 ölçek donusturme ile aynı ölcege alınmis

\$ Item_Parameters

1	2.914	-0.081	0.244
2	1.844	2.262	0.208
3	1.410	1.838	0.306
4	1.676	0.143	0.203
5	2.451	0.808	0.096
6	2.208	1.794	0.050
7	2.355	0.352	0.285
8	1.345	1.073	0.182
9	1.558	0.692	0.229
10	1.899	0.936	0.108
11	2.370	-0.173	0.247
12	2.268	2.267	0.036
13	1.434	-0.792	0.332
14	1.851	-0.375	0.348
15	1.290	0.525	0.318
16	2.642	0.740	0.220
17	1.945	1.189	0.096
18	1.497	2.158	0.281
19	1.621	-0.021	0.178
20	2.019	0.859	0.095
21	1.441	1.080	0.150
22	1.861	0.103	0.169
23	2.250	1.146	0.116
24	1.569	0.227	0.261
25	2.928	0.192	0.102
26	2.052	0.807	0.325

\$ Link_Item_Parameters

1	2.512	0.050	0.267
2	1.381	2.865	0.161
3	1.416	1.412	0.325
4	1.480	0.107	0.232
5	1.981	0.962	0.092
6	2.363	1.974	0.048
7	2.903	0.470	0.271
8	1.114	0.567	0.262
9	1.879	0.804	0.283
10	2.591	0.774	0.086
11	2.764	-0.378	0.272
12	1.640	2.464	0.047
13	1.644	-0.760	0.367

14	1.918	-0.567	0.373
15	1.503	-0.277	0.218
16	2.935	0.450	0.105
17	2.352	1.713	0.038
18	1.713	2.166	0.060
19	2.505	1.276	0.217
20	1.781	1.187	0.200
21	1.392	0.406	0.339
22	2.054	0.638	0.253
23	2.045	0.465	0.139
24	1.016	1.754	0.225
25	2.120	0.991	0.080
26	2.221	1.019	0.177

\$ Theta_Distribution

-4.000	0.00007648
-3.429	0.0006387
-2.857	0.003848
-2.286	0.01673
-1.714	0.05245
-1.143	0.1186
-0.5714	0.1936
-0.0001	0.2280
0.5714	0.1936
1.143	0.1186
1.714	0.05245
2.286	0.01673
2.857	0.003848
3.429	0.0006387
4.000	0.00007648

\$ Link_Conversion

Form Y

27	0	26	1	1
0	0	0	0	
1	1	1	1	
2	2	2	2	
3	3	3	3	
4	4	4	4	
5	5	5	5	
6	6	6	6	
7	7	7	7	
8	8	8	8	
9	9	9	9	
10	10	10	10	
11	11	11	11	
12	12	12	12	
13	13	13	13	

14	14	14	14
15	15	15	15
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21	21	21	21
22	22	22	22
23	23	23	23
24	24	24	24
25	25	25	25
26	26	26	26

(NOT:Form X e ait ortak maddeler sırası 1-14 olarak yer değiştirilmiştir, OO' ya göre a ve b parametreleri dönüştürülmüştür, c parametresi dönüşümden bağımsızdır.)

PIE 2.0 OUTPUT DOSYASI

Results for True Score Equating

Sum of c parameters: 5.1850

Raw Score	Theta Equivalent	True Score Equivalent	Unrounded Scale Score	Rounded Scale Score
0	-----	0.0000	0.0000	0.00
1	-----	0.9907	0.9907	1.00
2	-----	1.9815	1.9815	2.00
3	-----	2.9722	2.9722	3.00
4	-----	3.9630	3.9630	4.00
5	-----	4.9537	4.9537	5.00
6	-0.8279	6.0497	6.0497	6.00
7	-0.4590	7.0657	7.0657	7.00
8	-0.2451	7.8925	7.8925	8.00
9	-0.0827	8.6213	8.6213	9.00
10	0.0573	9.3369	9.3369	9.00
11	0.1867	10.0964	10.0964	10.00
12	0.3111	10.9340	10.9340	11.00
13	0.4327	11.8560	11.8560	12.00
14	0.5519	12.8346	12.8346	13.00
15	0.6686	13.8300	13.8300	14.00
16	0.7839	14.8185	14.8185	15.00
17	0.9006	15.7994	15.7994	16.00
18	1.0224	16.7840	16.7840	17.00
19	1.1544	17.7870	17.7870	18.00
20	1.3033	18.8221	18.8221	19.00
21	1.4772	19.8998	19.8998	20.00
22	1.6817	21.0211	21.0211	21.00
23	1.9152	22.1612	22.1612	22.00

24	2.1780	23.2590	23.2590	23.00
25	2.5126	24.3156	24.3156	24.00
26	-----	26.0000	26.0000	26.00

Results for Observed Score Equating

Raw Score	Equated Raw Score	Unrounded Scale Score	Rounded Scale Score
0	0.0284	0.0284	0.00
1	1.0341	1.0341	1.00
2	2.0395	2.0395	2.00
3	3.0427	3.0427	3.00
4	4.0399	4.0399	4.00
5	5.0241	5.0241	5.00
6	5.9847	5.9847	6.00
7	6.9119	6.9119	7.00
8	7.8024	7.8024	8.00
9	8.6590	8.6590	9.00
10	9.4875	9.4875	9.00
11	10.3220	10.3220	10.00
12	11.1447	11.1447	11.00
13	11.9936	11.9936	12.00
14	12.8974	12.8974	13.00
15	13.8436	13.8436	14.00
16	14.8139	14.8139	15.00
17	15.8014	15.8014	16.00
18	16.7956	16.7956	17.00
19	17.7896	17.7896	18.00
20	18.8000	18.8000	19.00
21	19.8613	19.8613	20.00
22	20.9581	20.9581	21.00
23	22.0509	22.0509	22.00
24	23.1534	23.1534	23.00
25	24.2550	24.2550	24.00
26	25.4310	25.4310	25.00

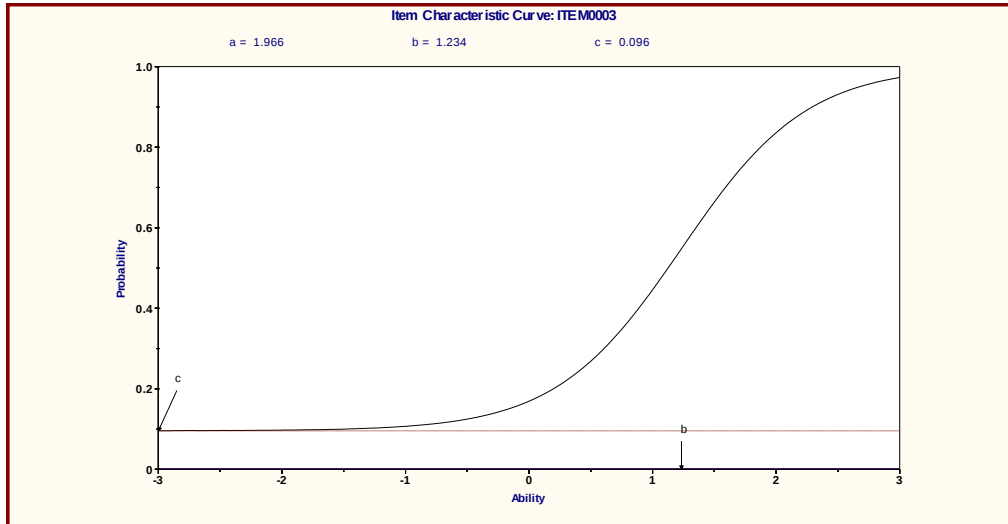
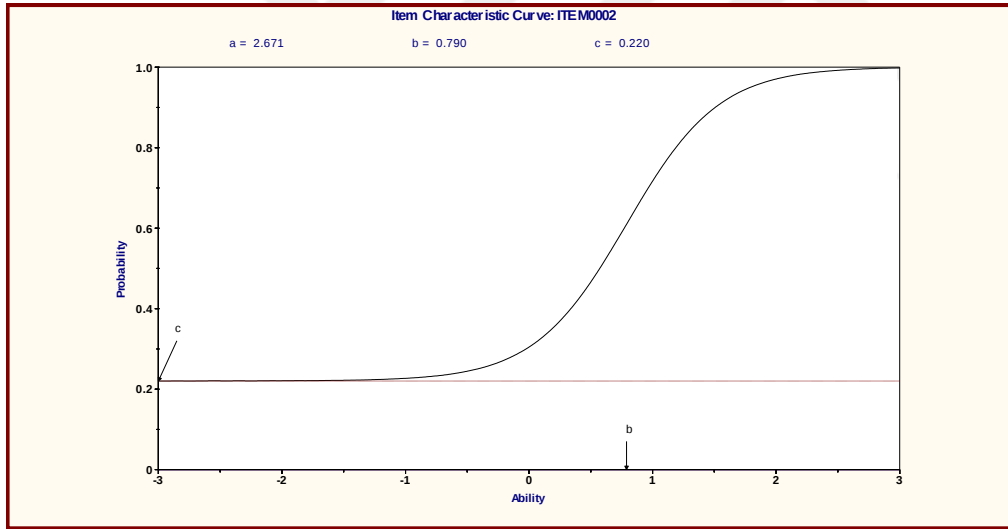
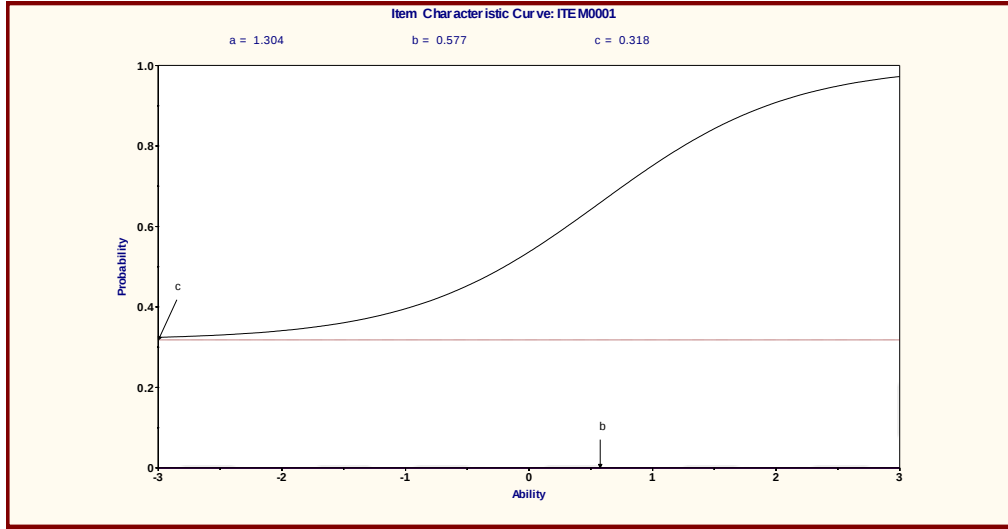
Fitted observed score distributions

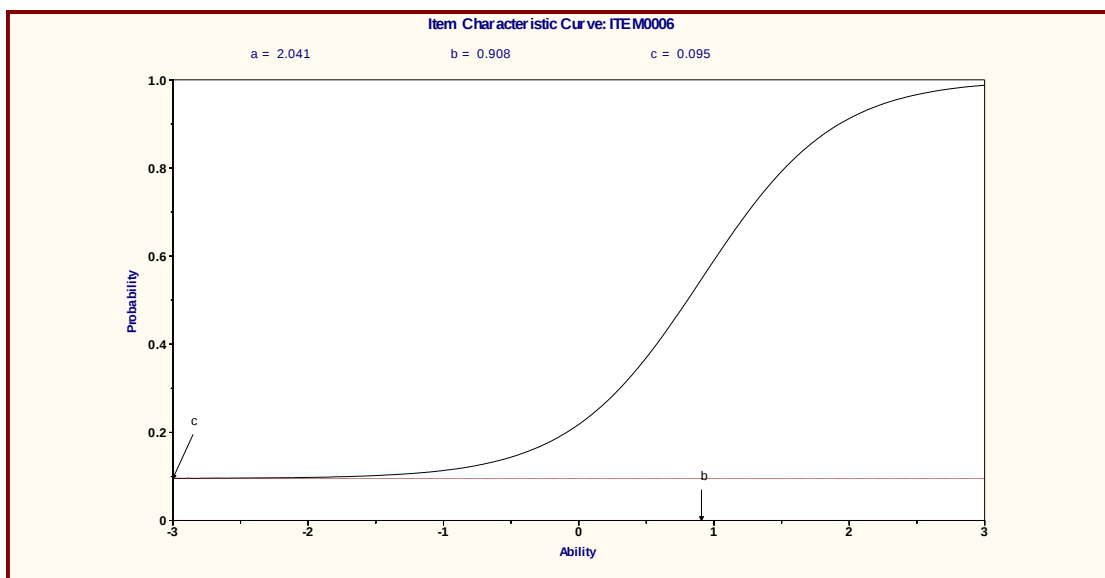
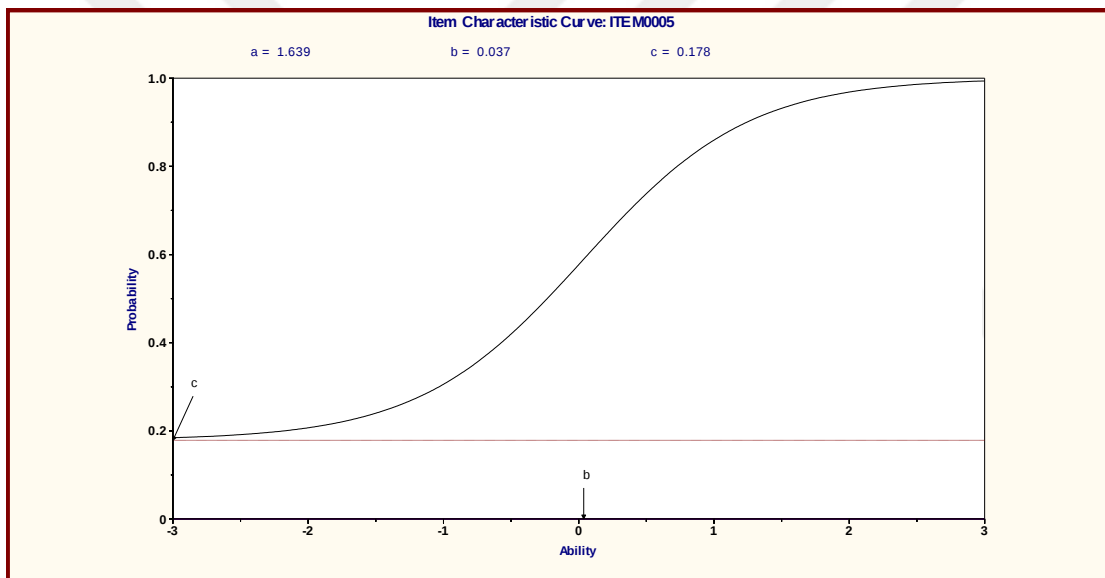
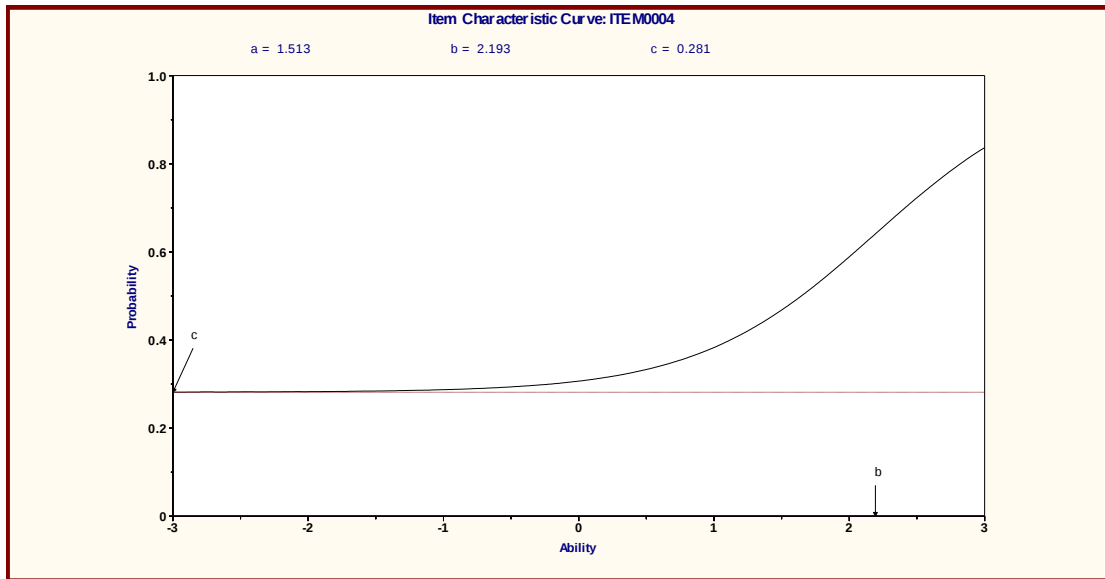
Raw Score	New Form	Old Form
0	0.00037671	0.00035649
1	0.00300160	0.00284759
2	0.01151871	0.01099806
3	0.02840855	0.02745559
4	0.05082565	0.05012370
5	0.07099961	0.07222056
6	0.08218912	0.08710874
7	0.08346336	0.09212093
8	0.07810486	0.08797362

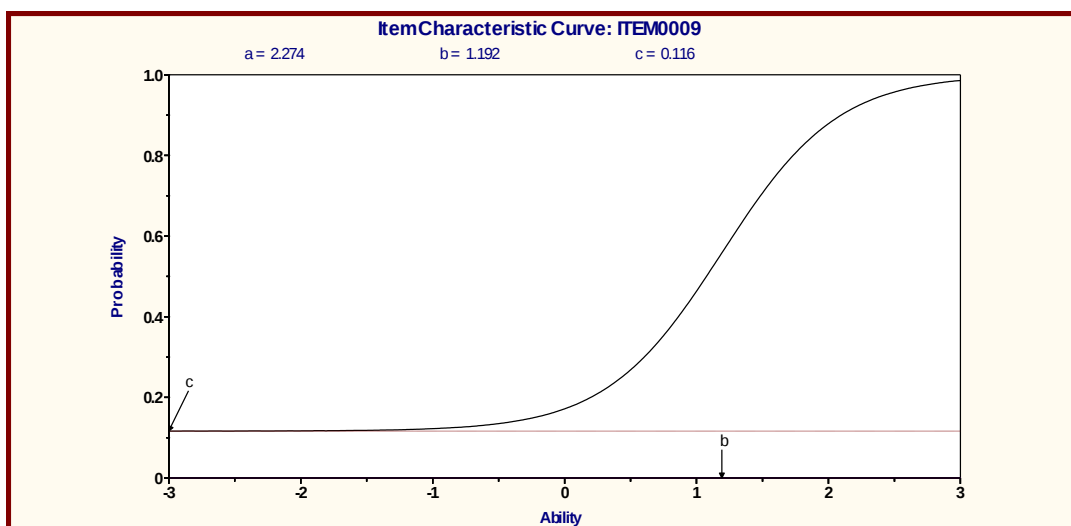
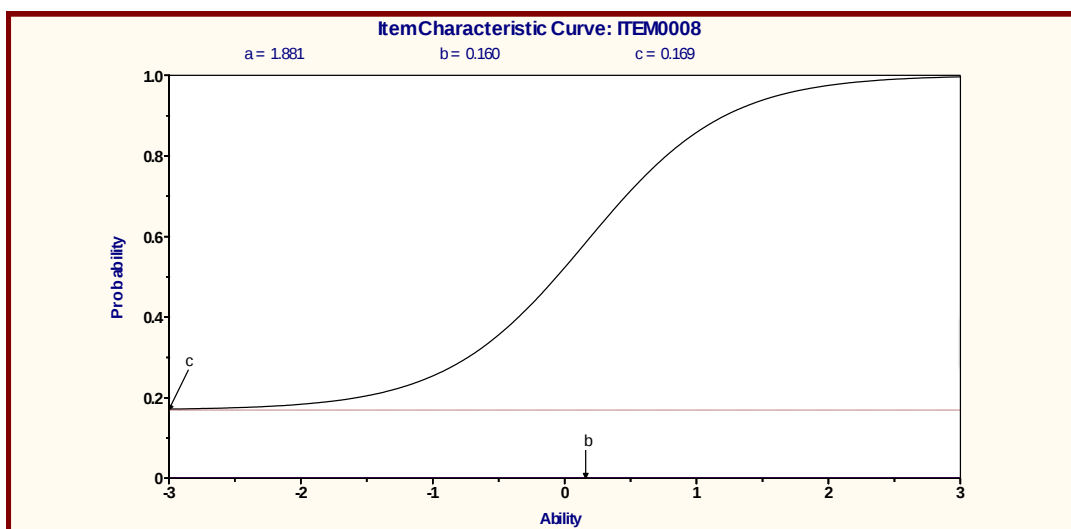
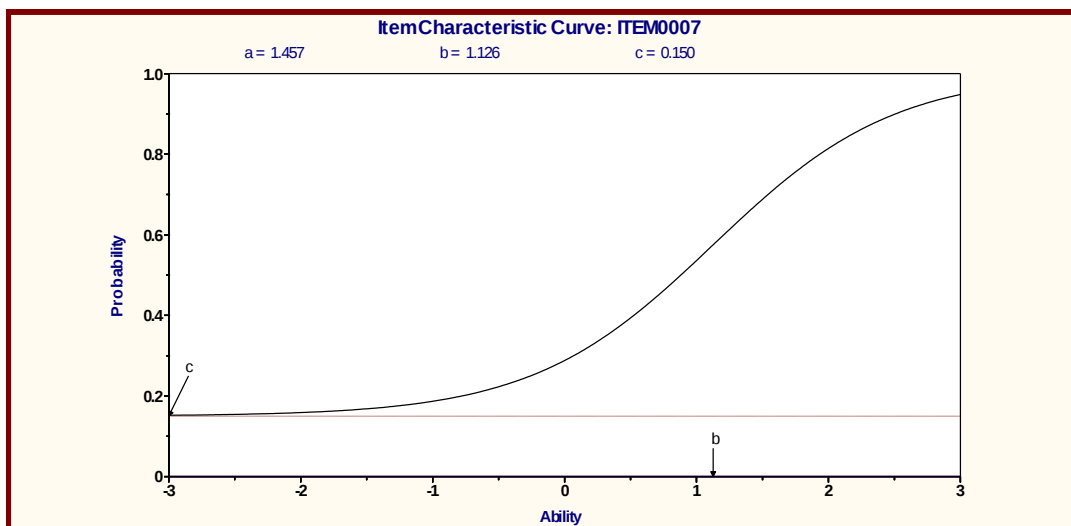
9	0.06925165	0.07739170
10	0.05897306	0.06517148
11	0.05010909	0.05623544
12	0.04560756	0.05134452
13	0.04503875	0.04665944
14	0.04404898	0.03934662
15	0.03921981	0.03194492
16	0.03249091	0.02899426
17	0.02881822	0.03018017
18	0.02953479	0.03008836
19	0.03041488	0.02610194
20	0.02799264	0.02193978
21	0.02440573	0.01956785
22	0.02154578	0.01601742
23	0.01731093	0.01160929
24	0.01224703	0.00834202
25	0.00839697	0.00531731
26	0.00559141	0.00242859

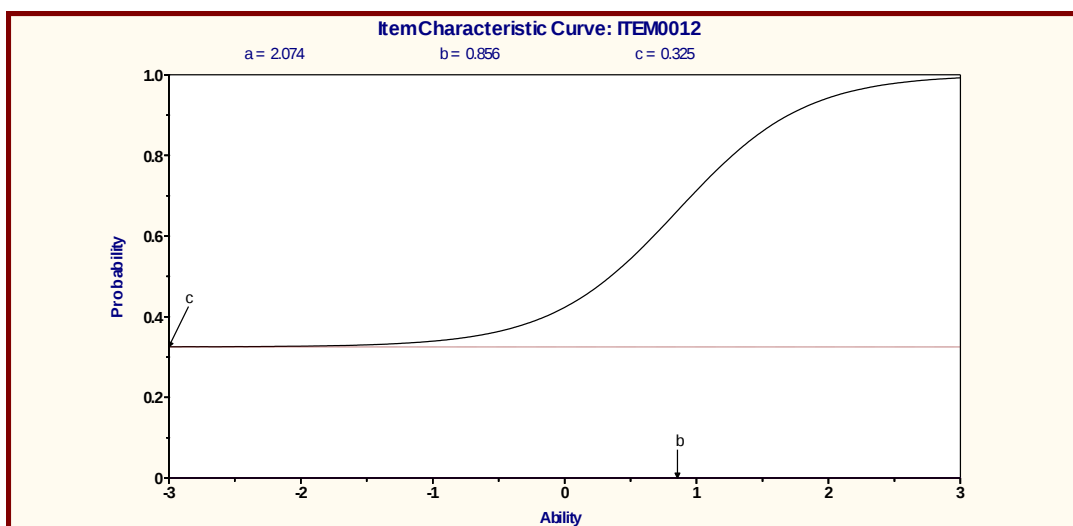
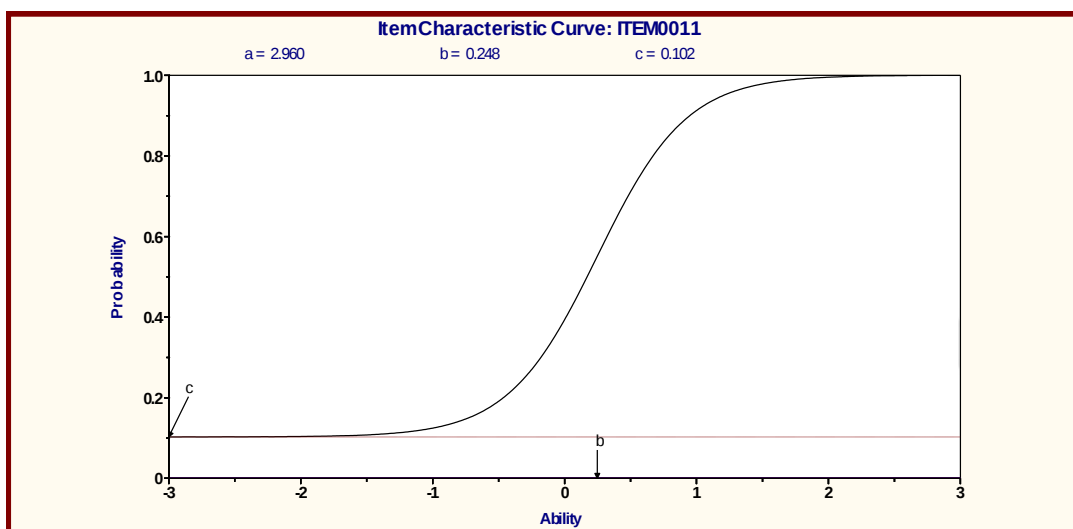
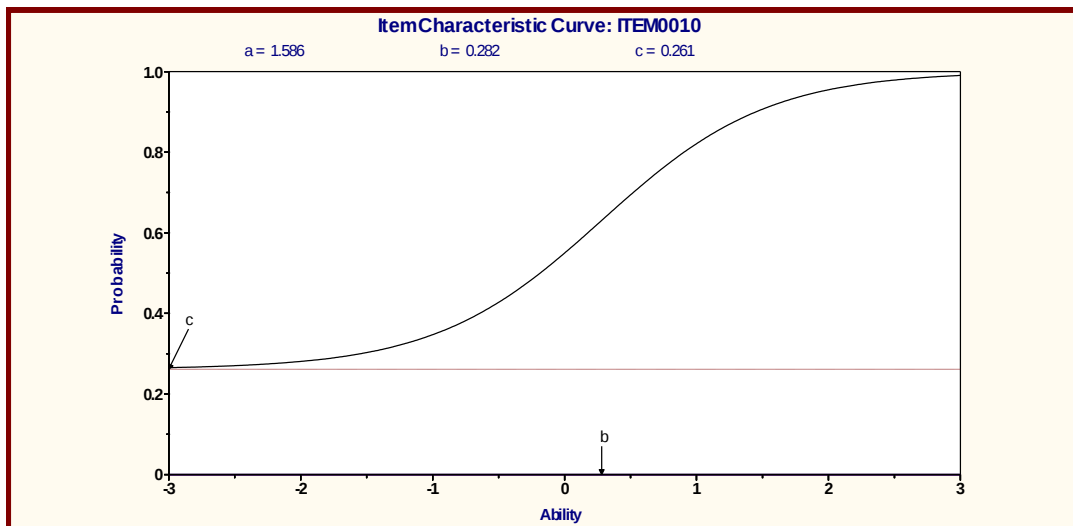
EK 4- Form X ve Form Y'ye Ait Madde Karakteristik Eğrileri

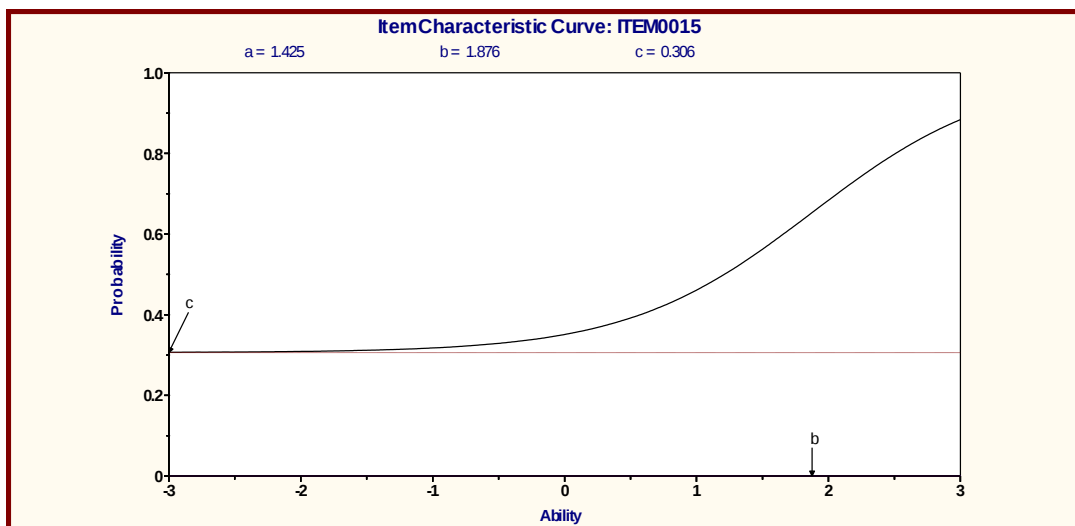
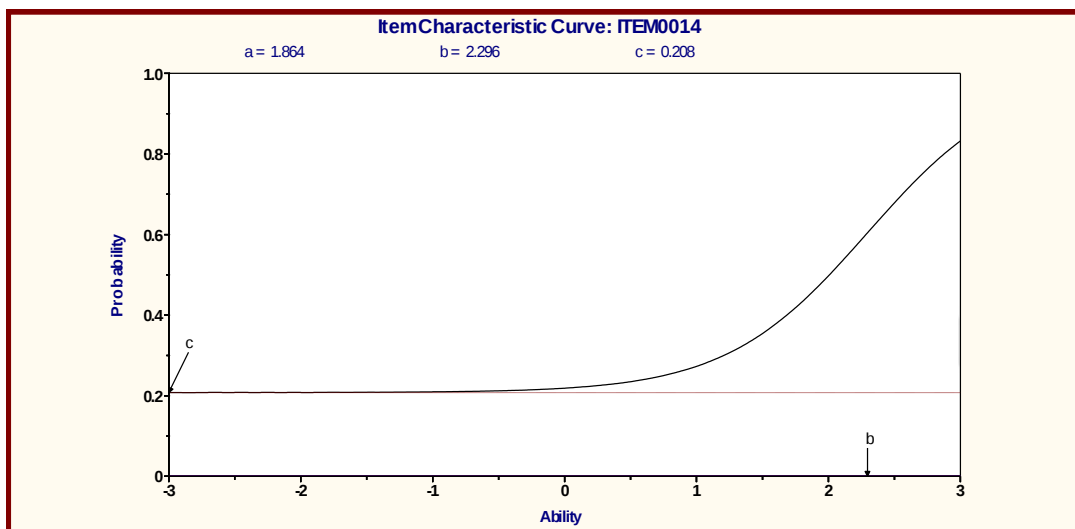
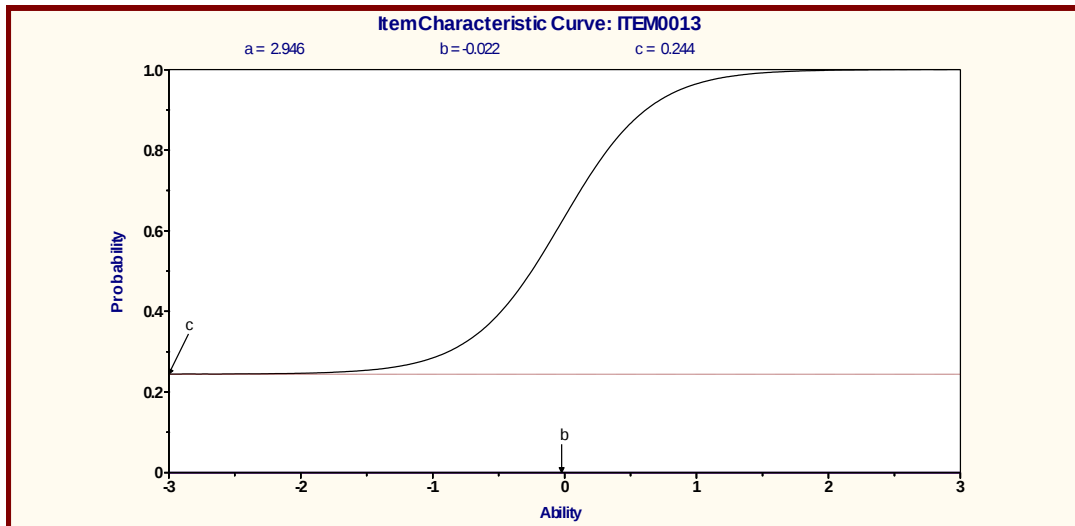
Form X'e ait MKE'ler

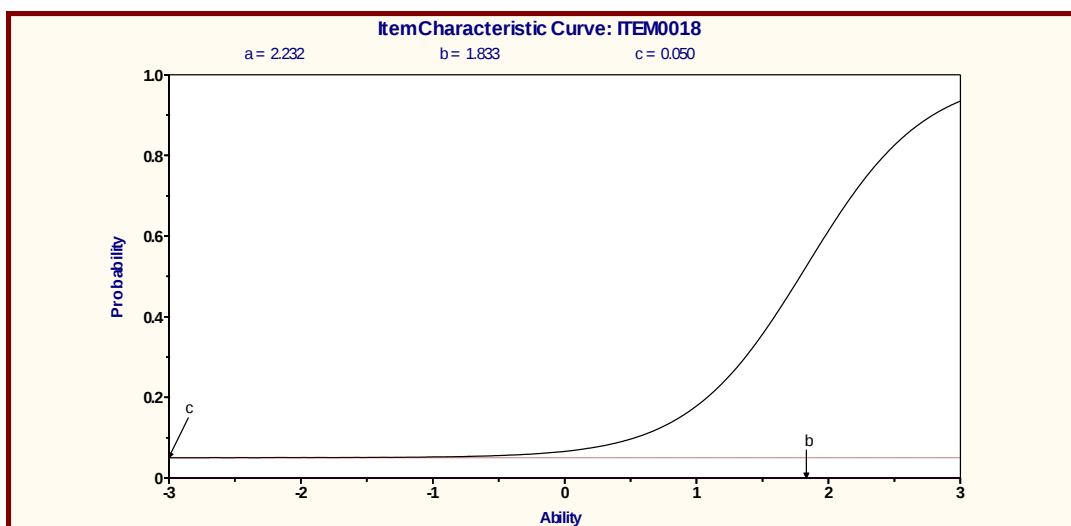
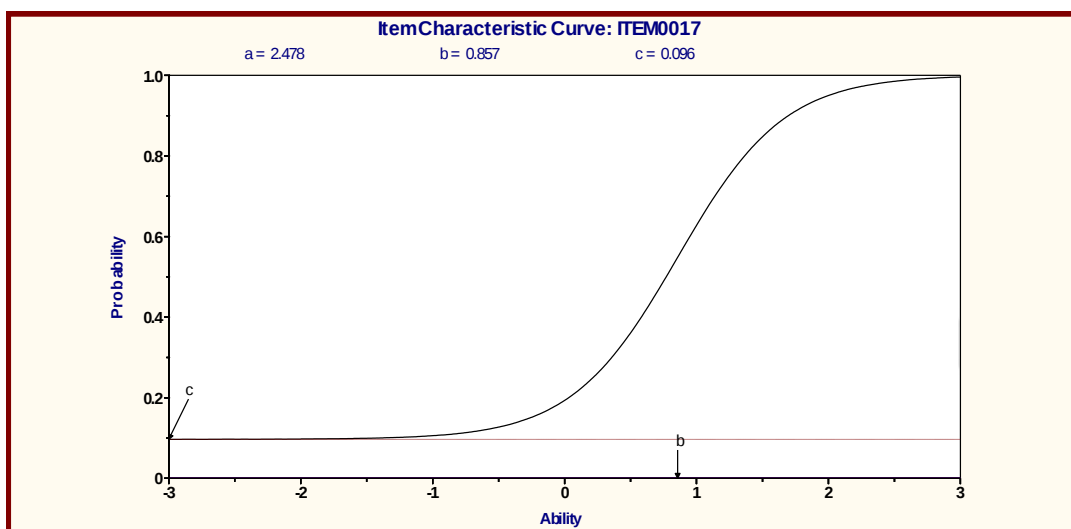
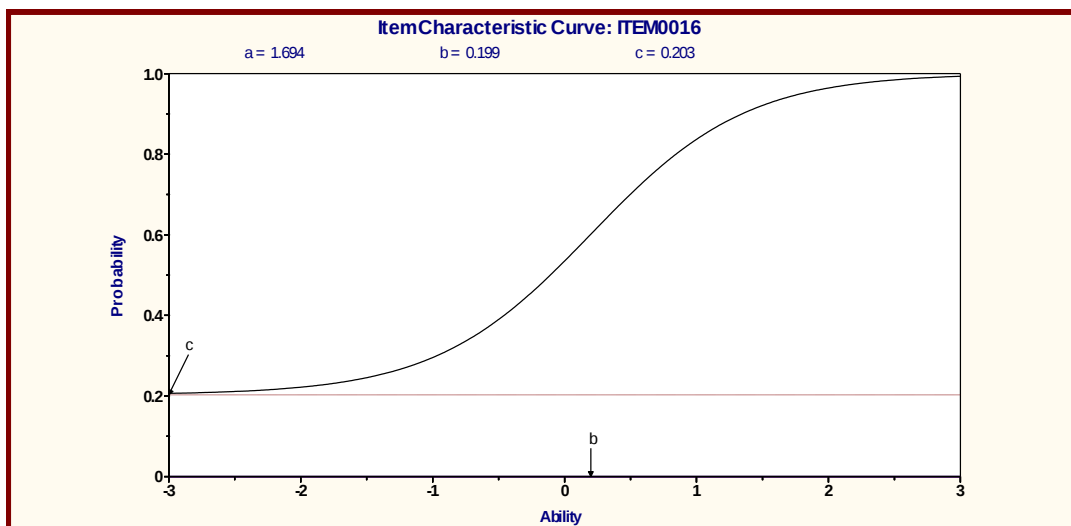


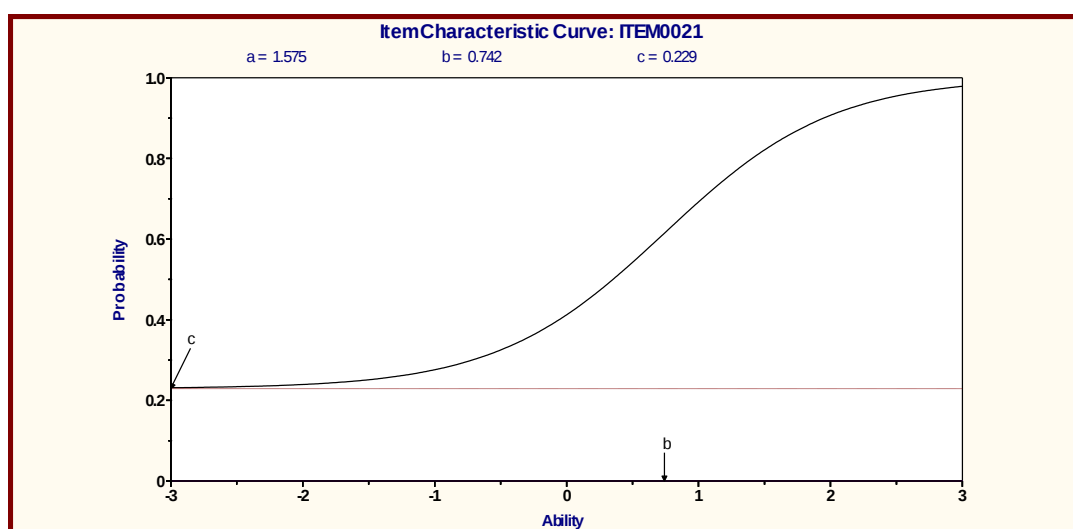
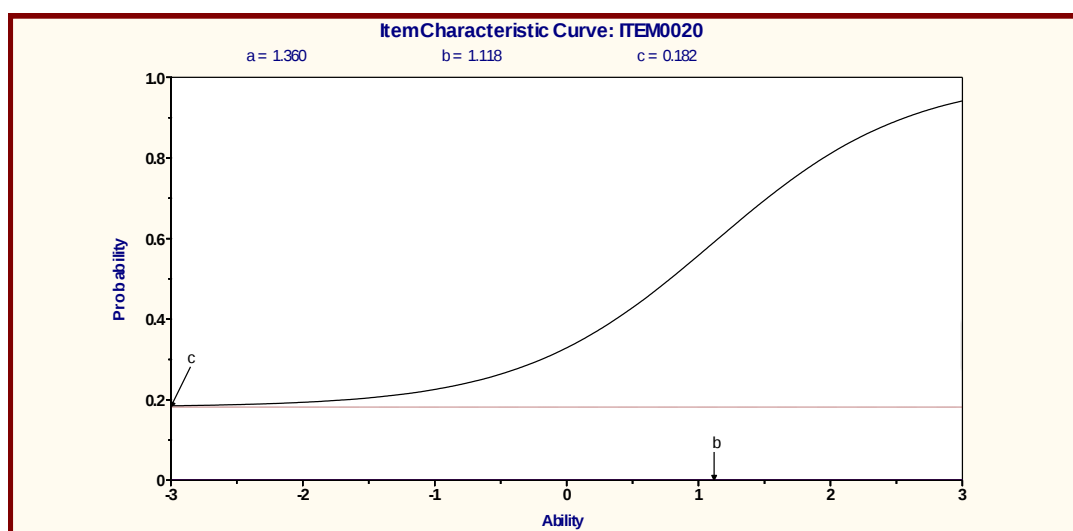
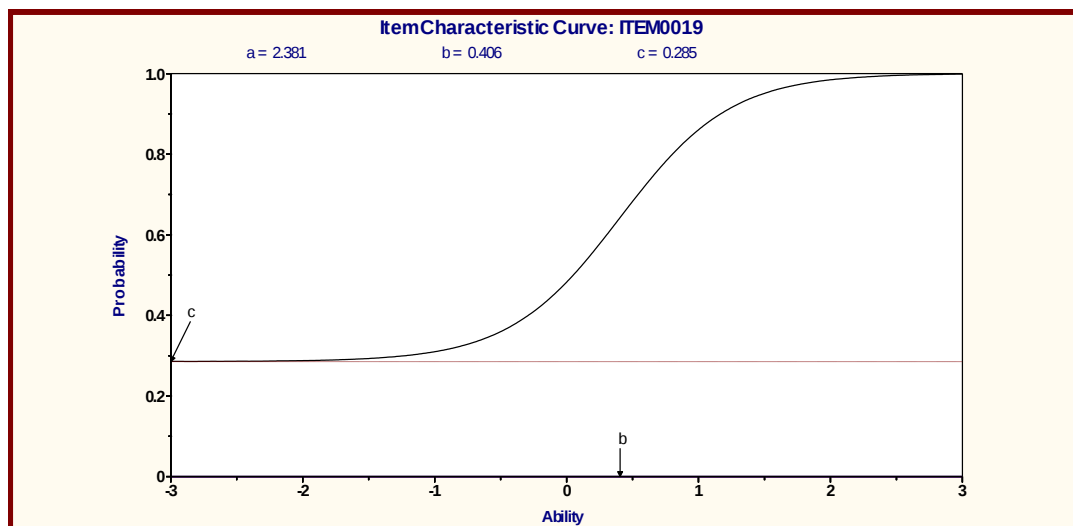


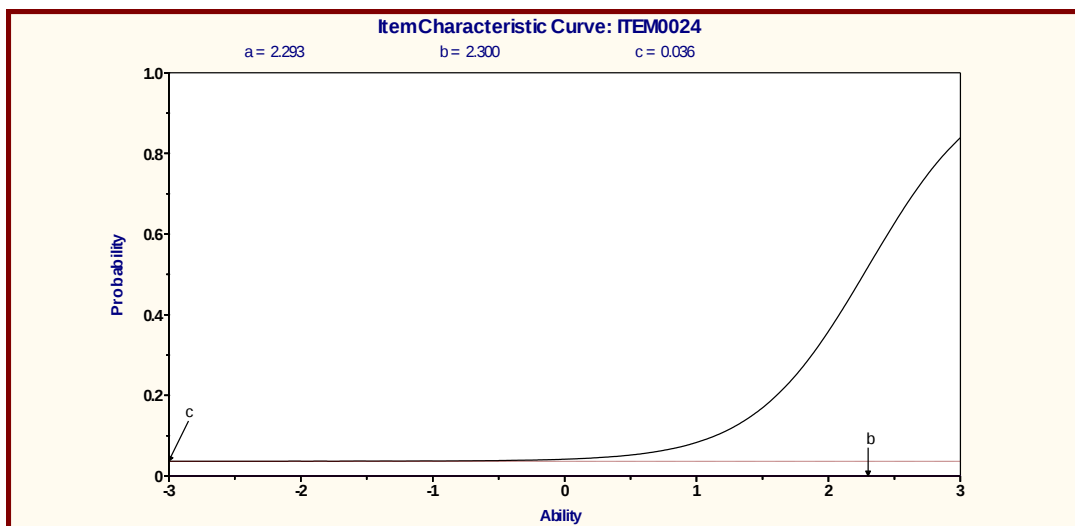
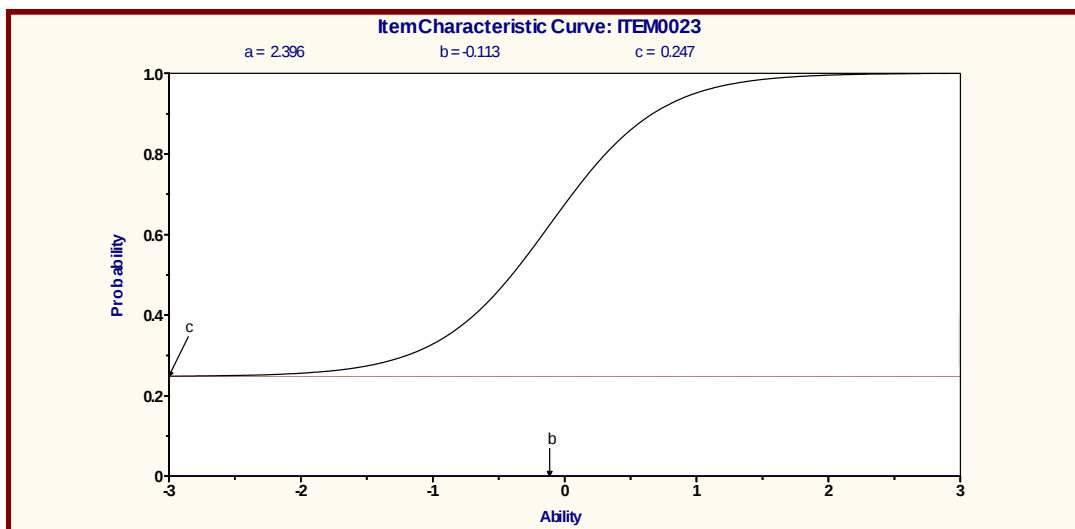
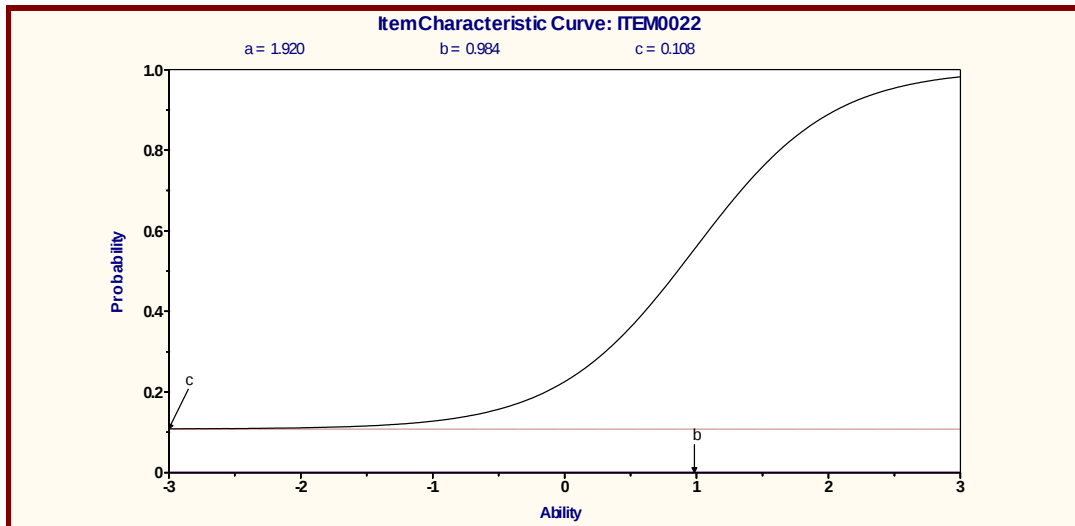


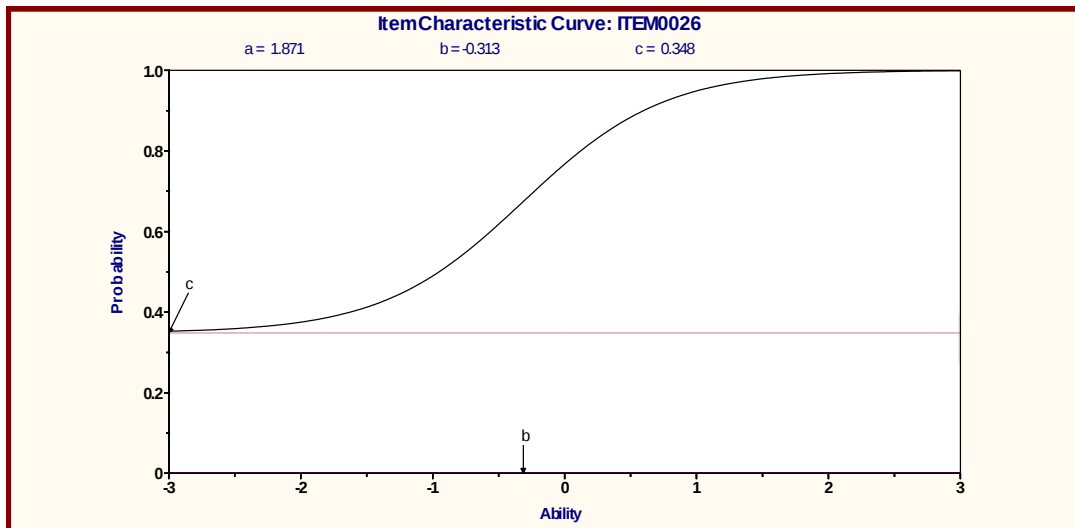
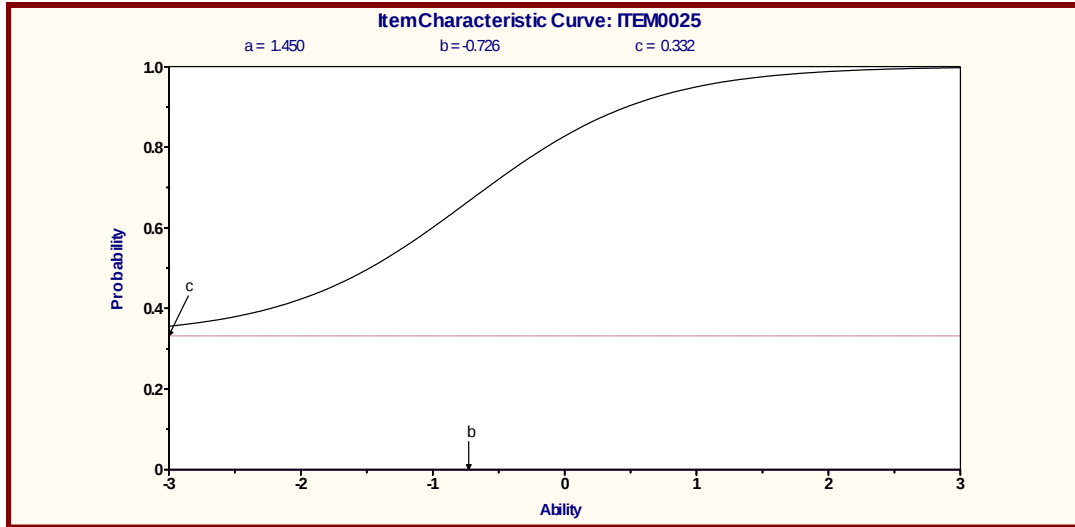




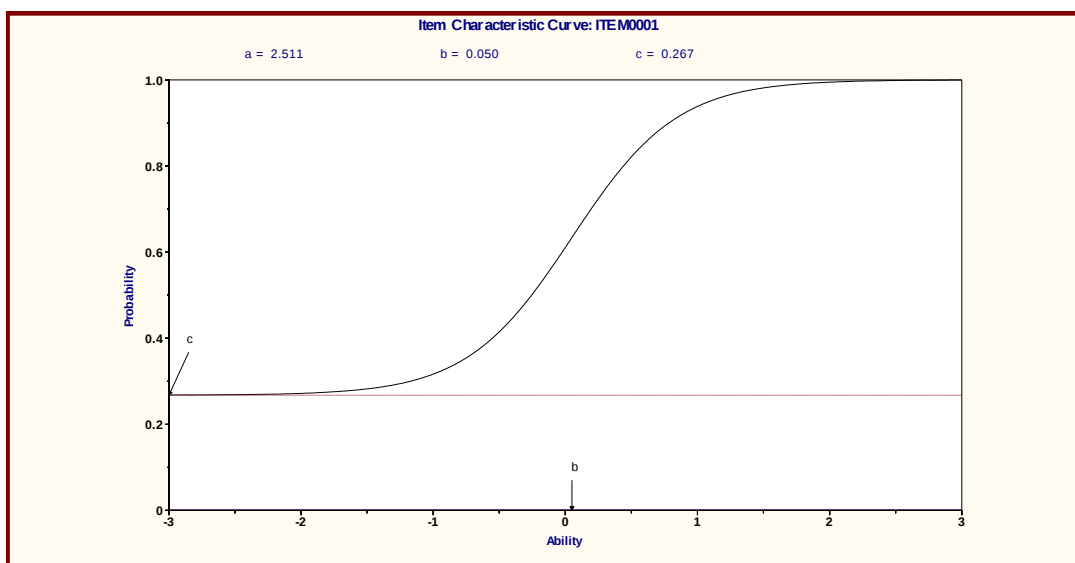


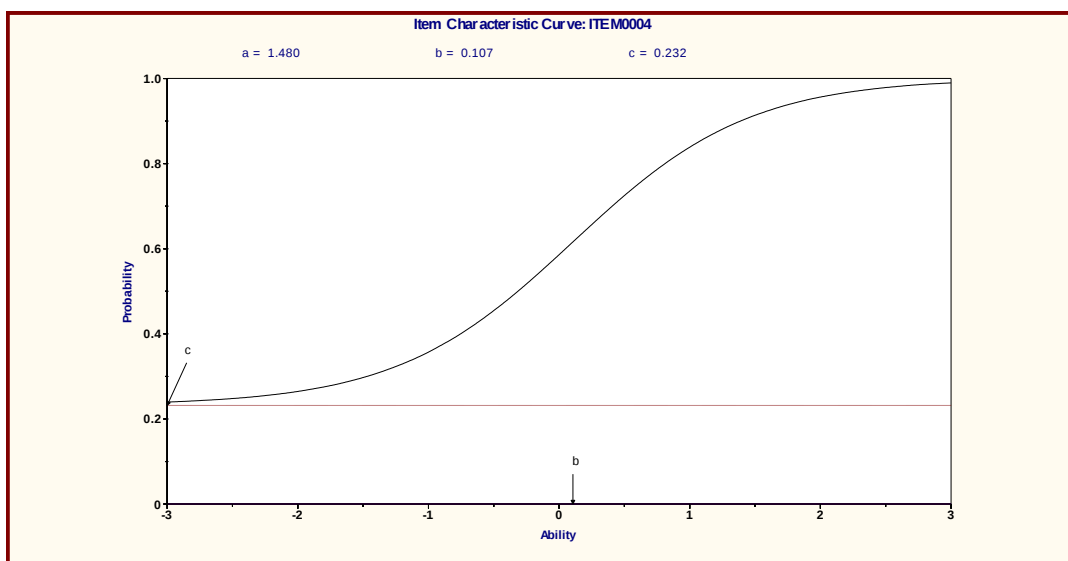
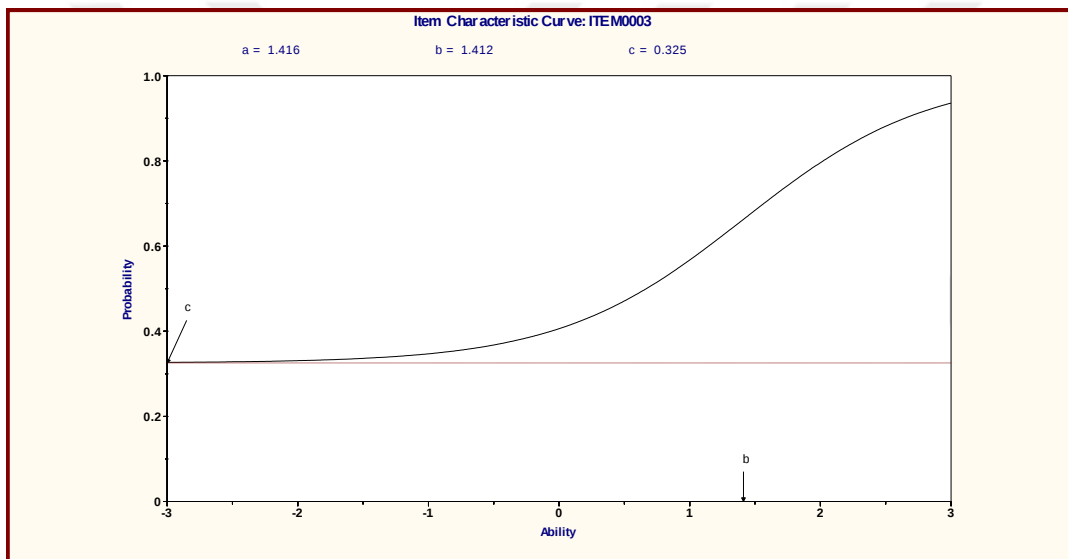
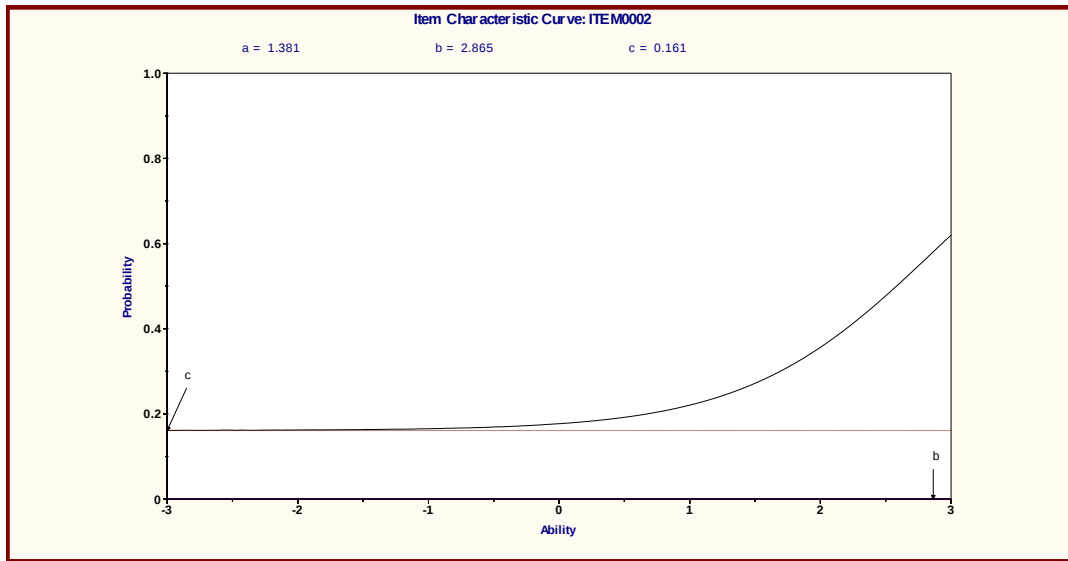


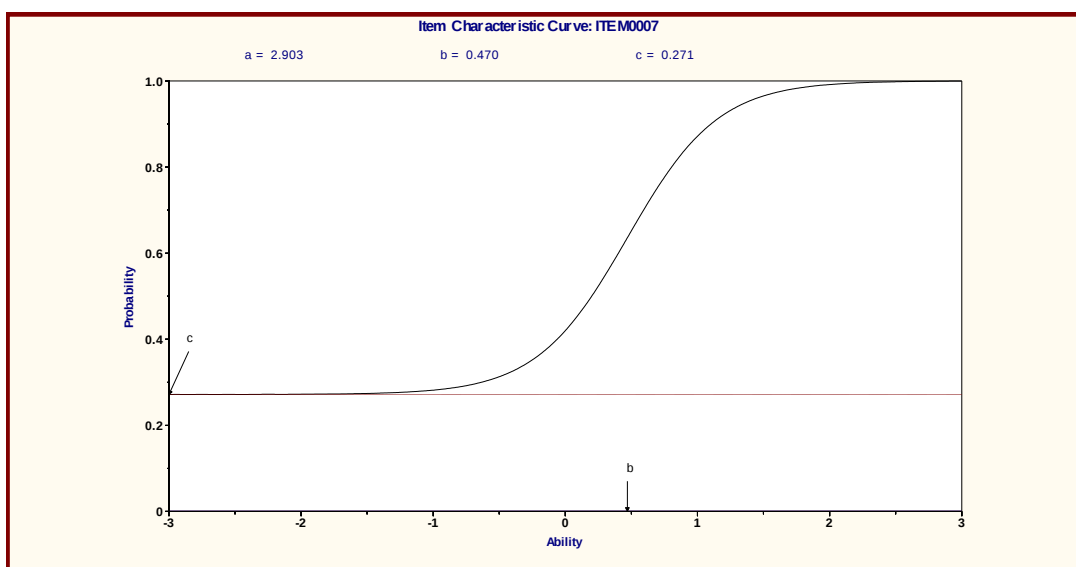
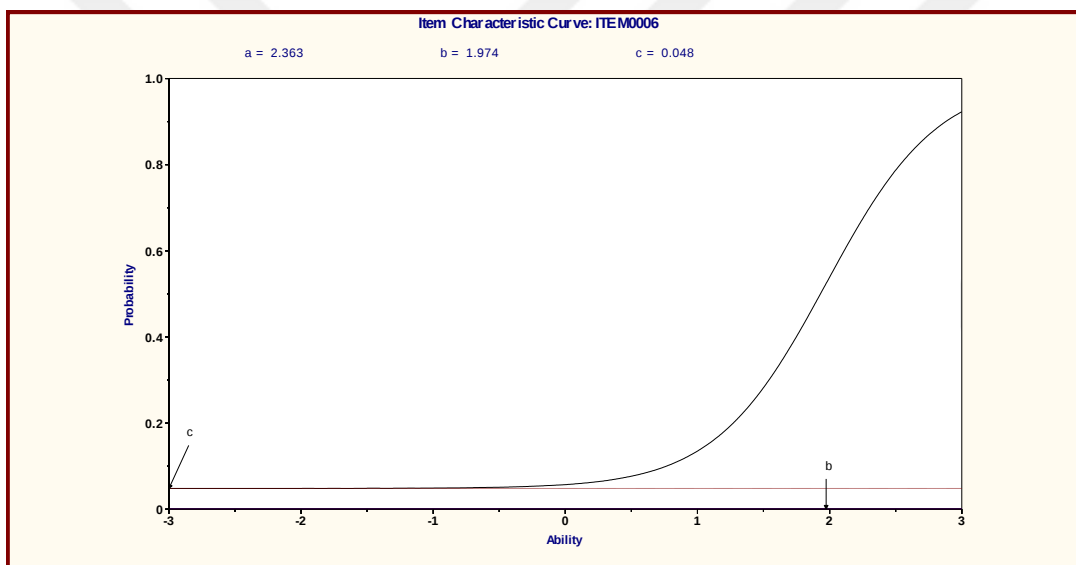
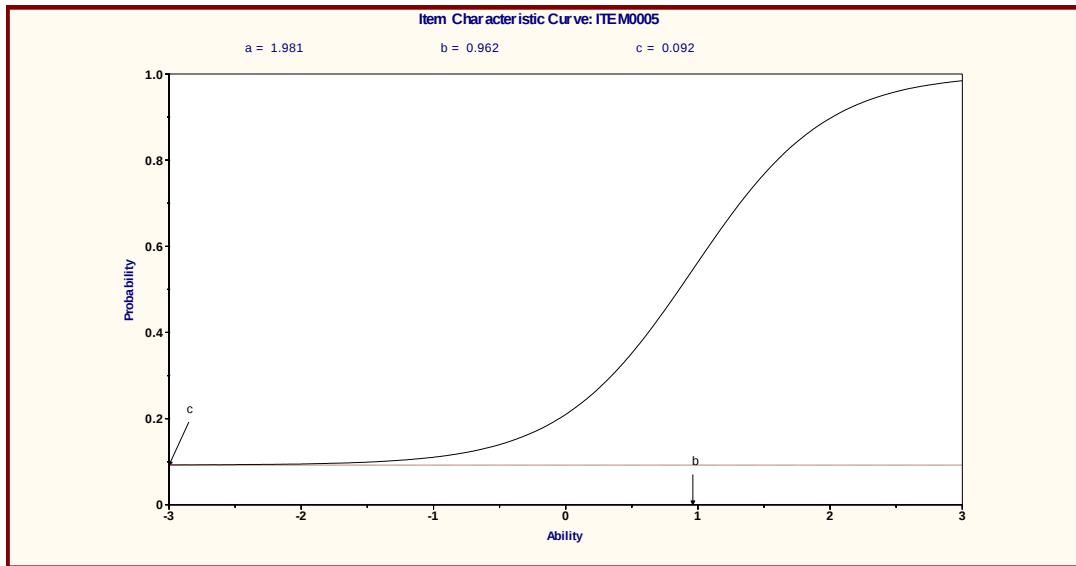


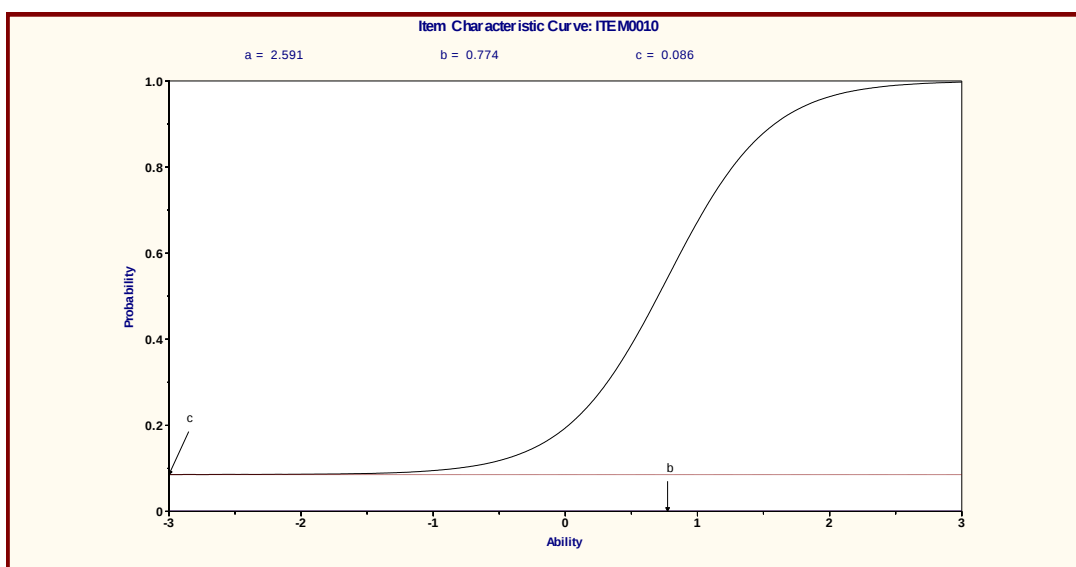
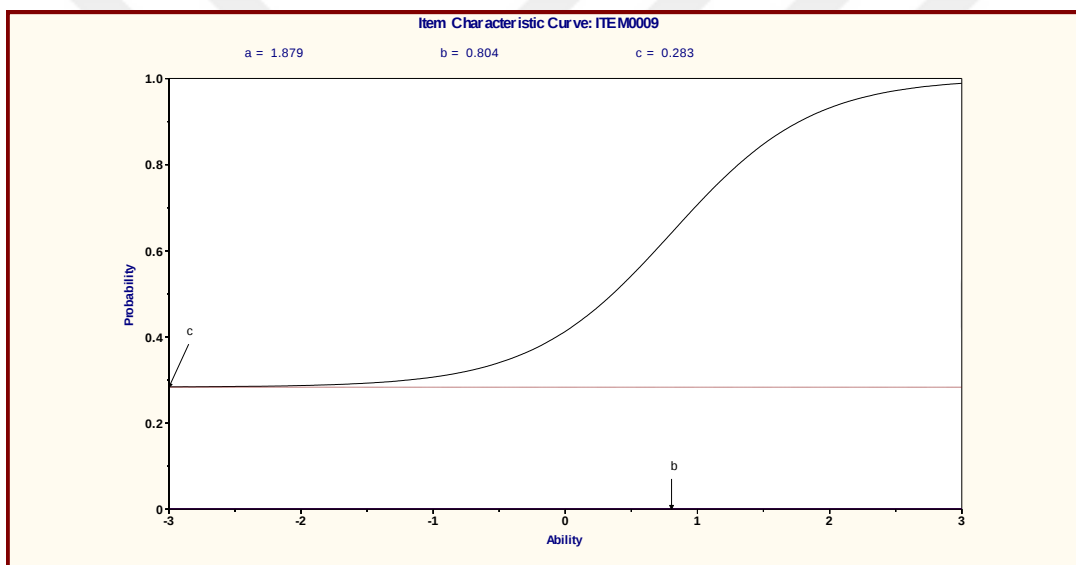
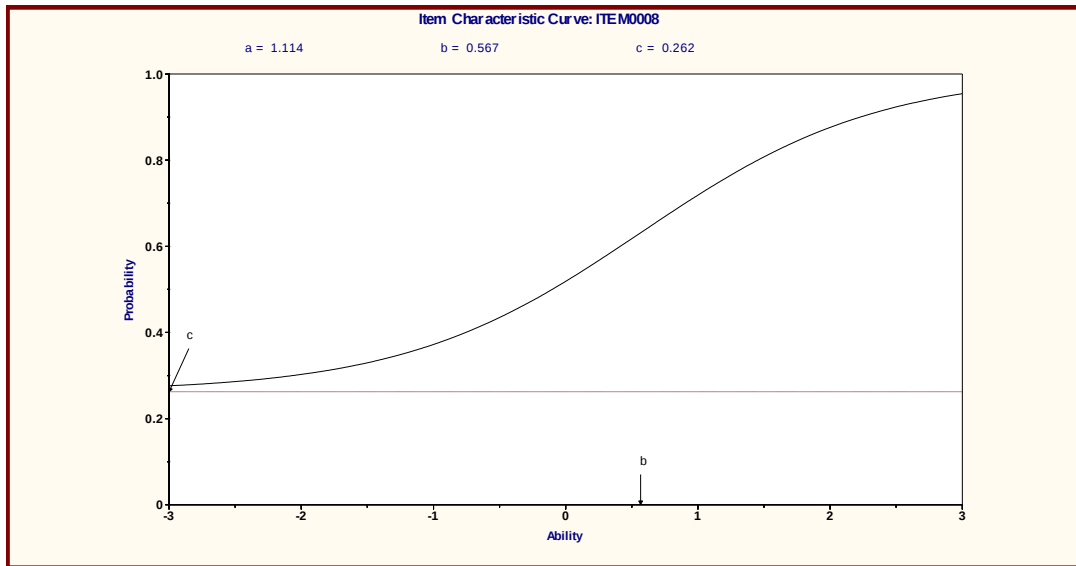


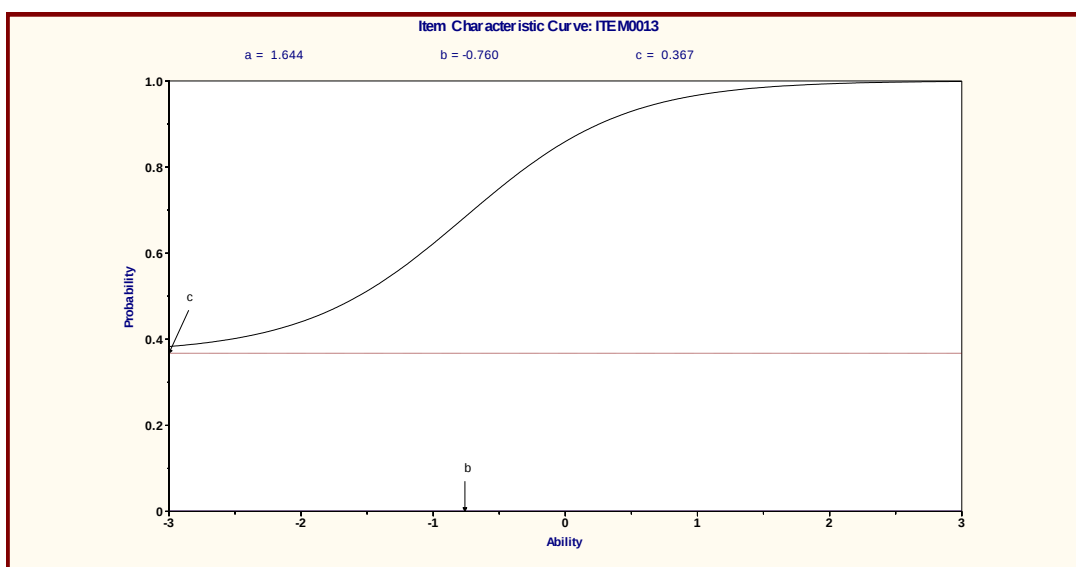
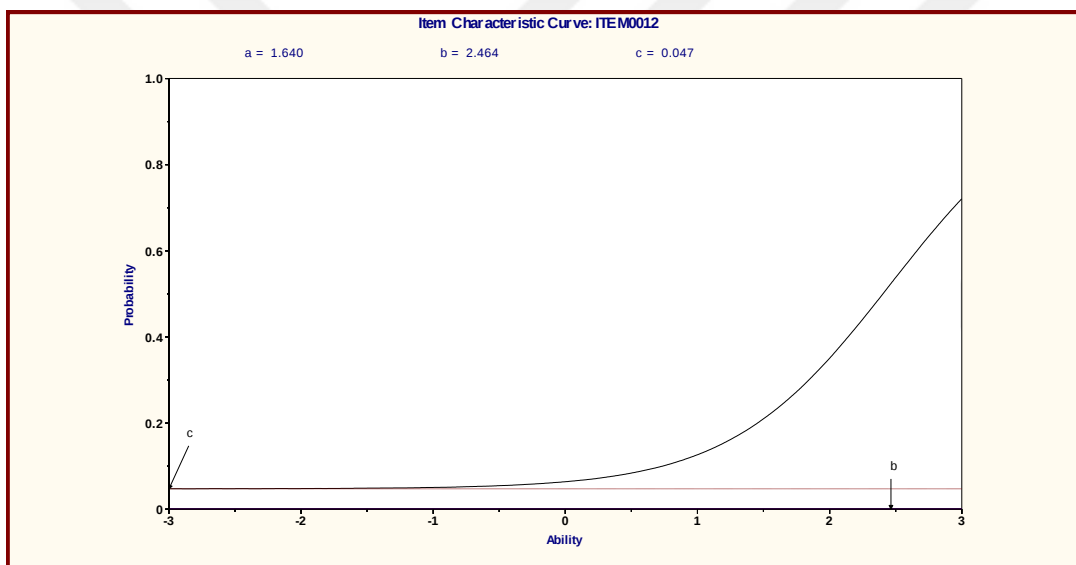
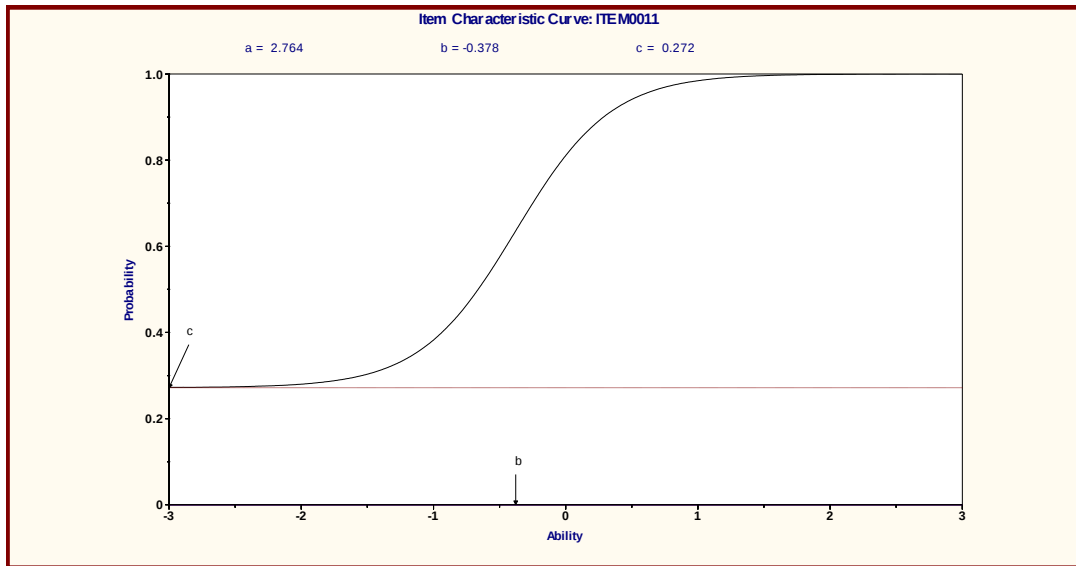
Form Y'ye ait MKE'ler

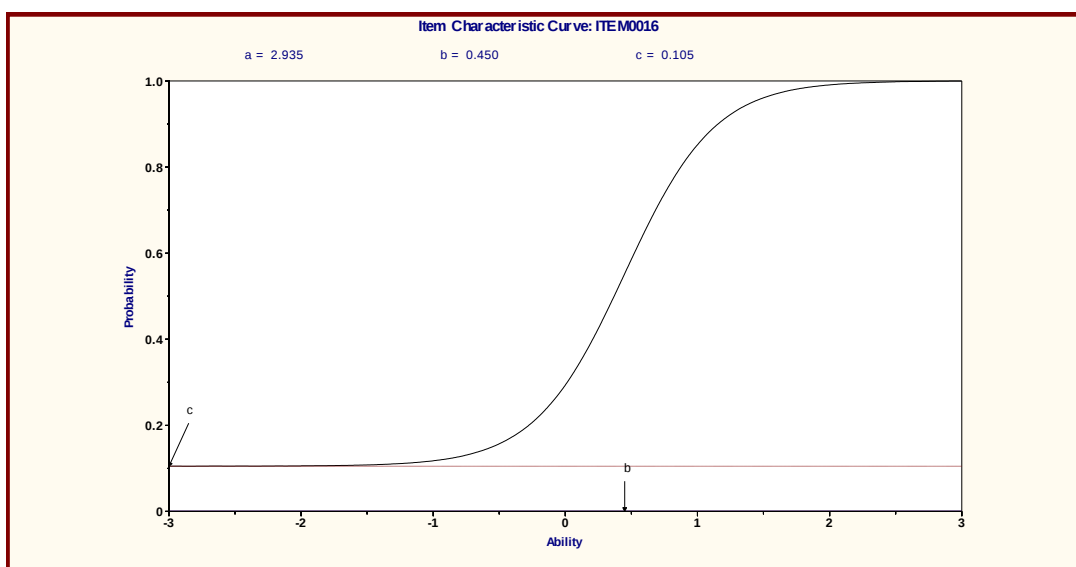
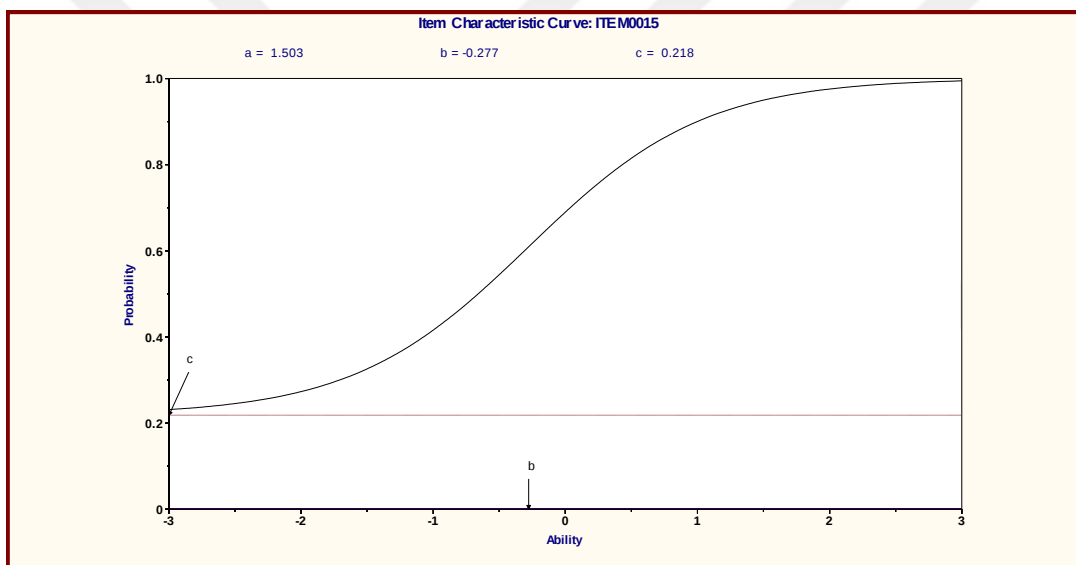
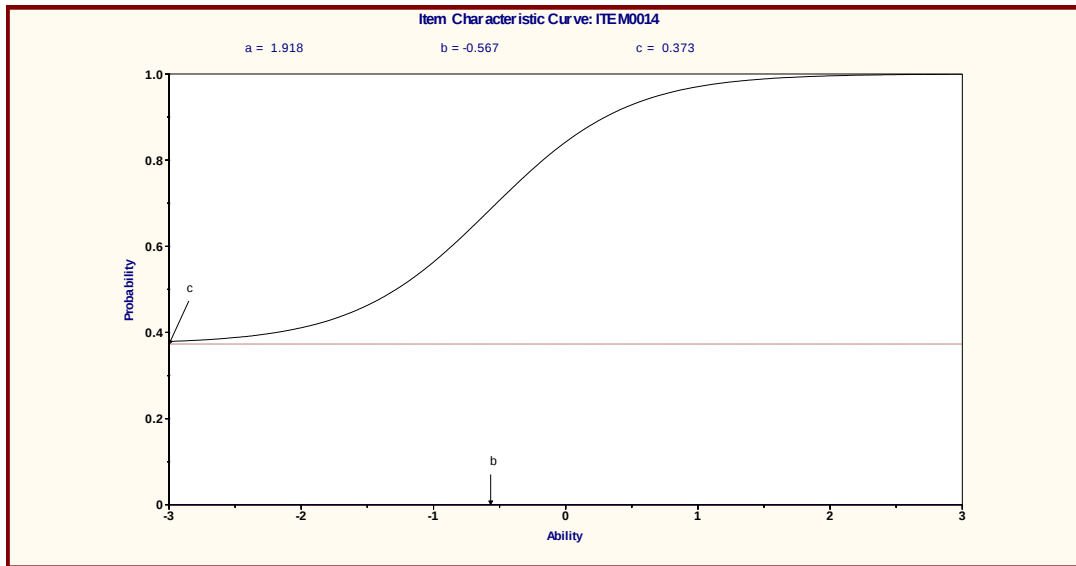


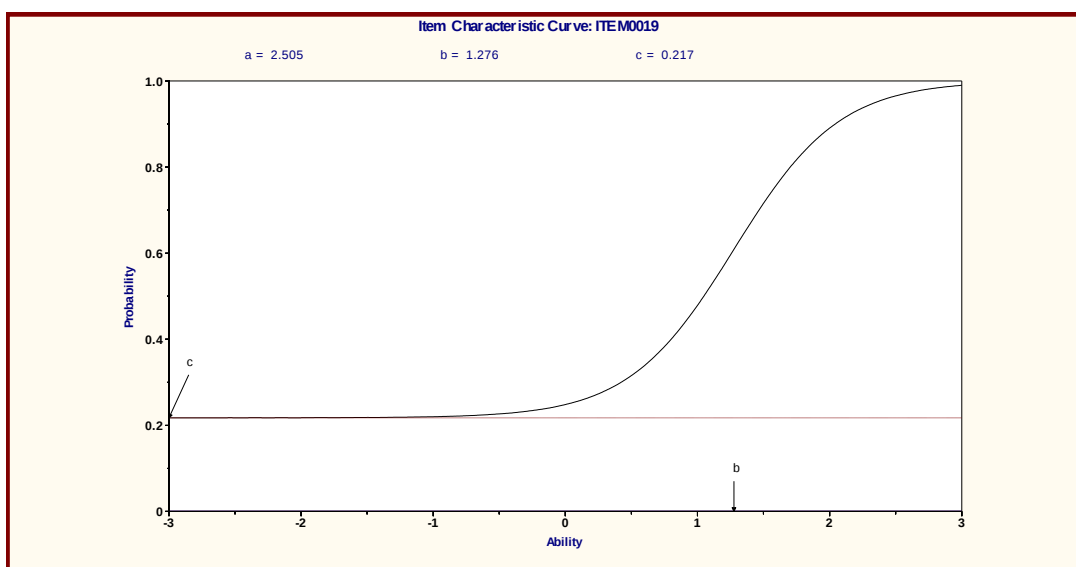
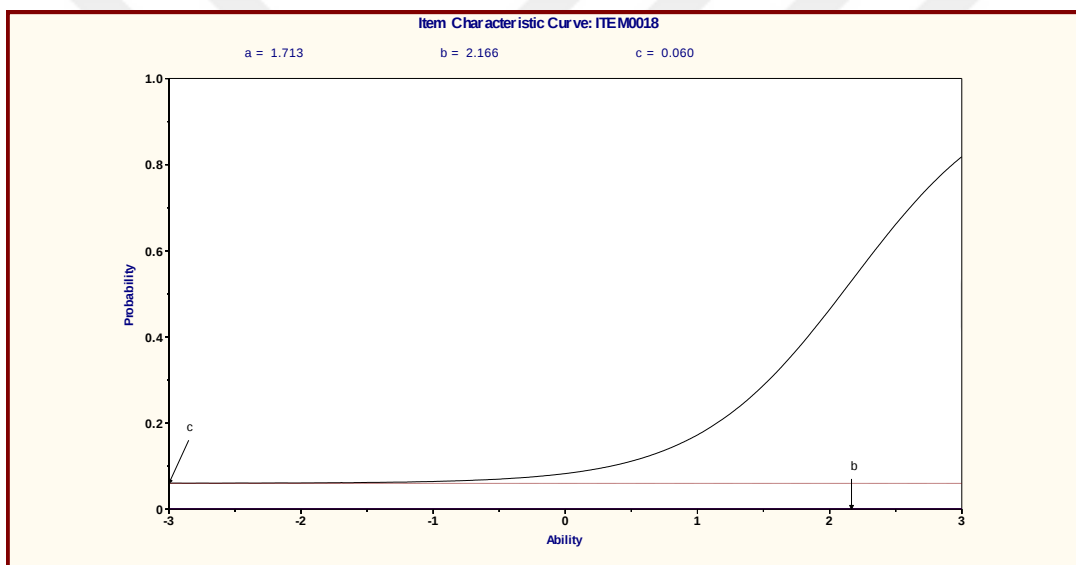
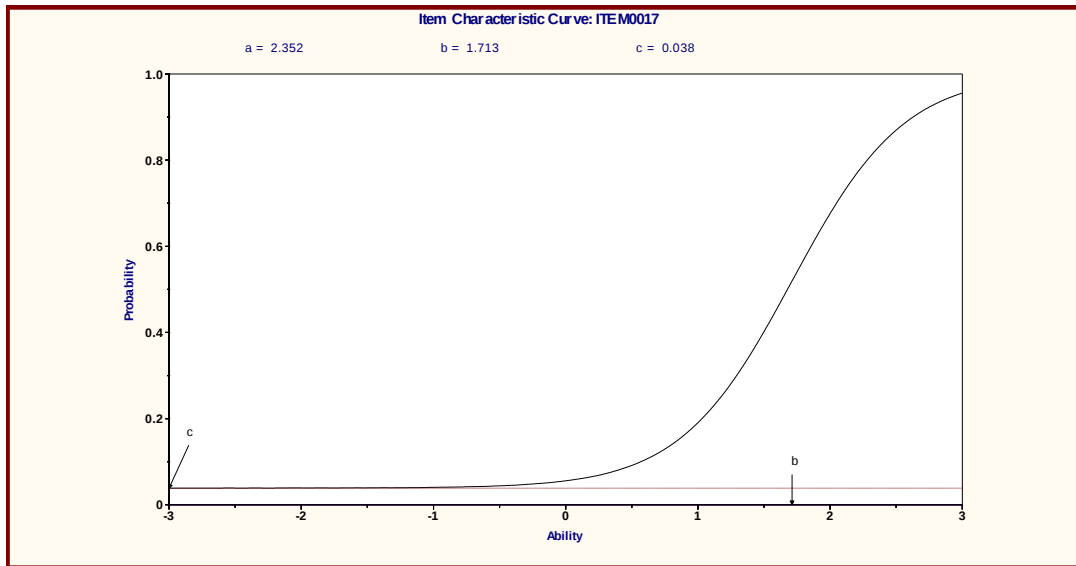


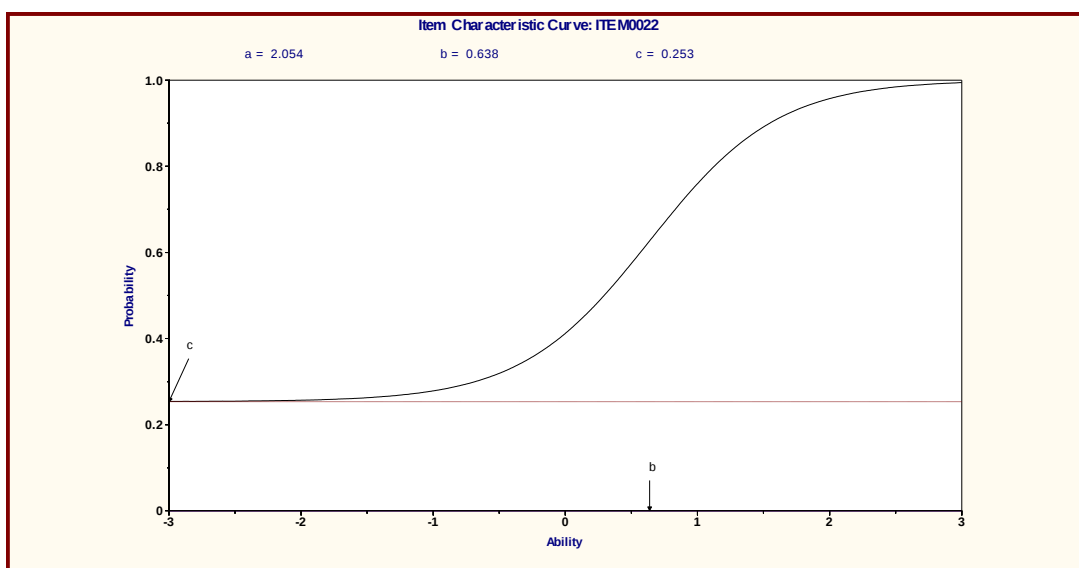
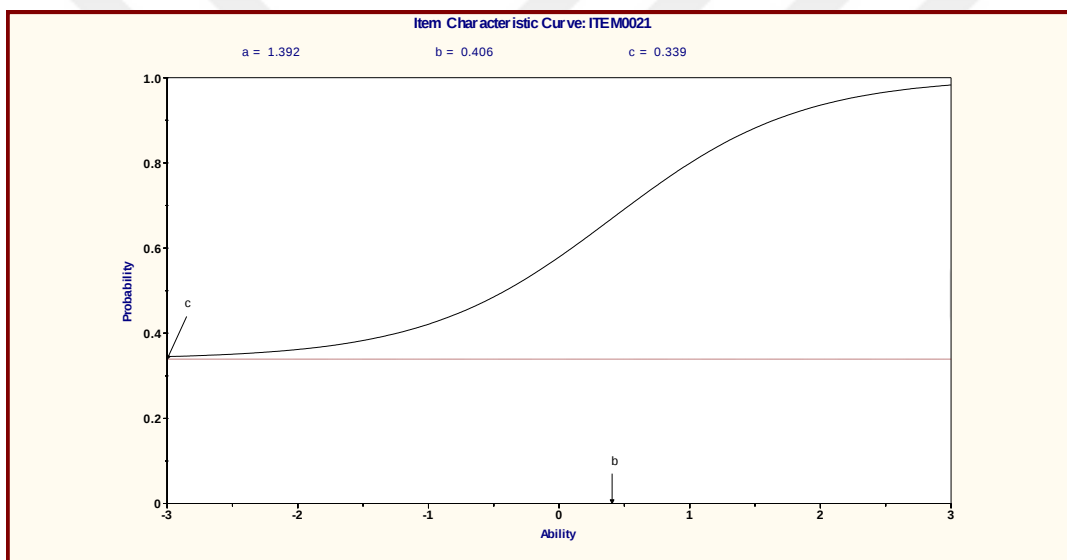
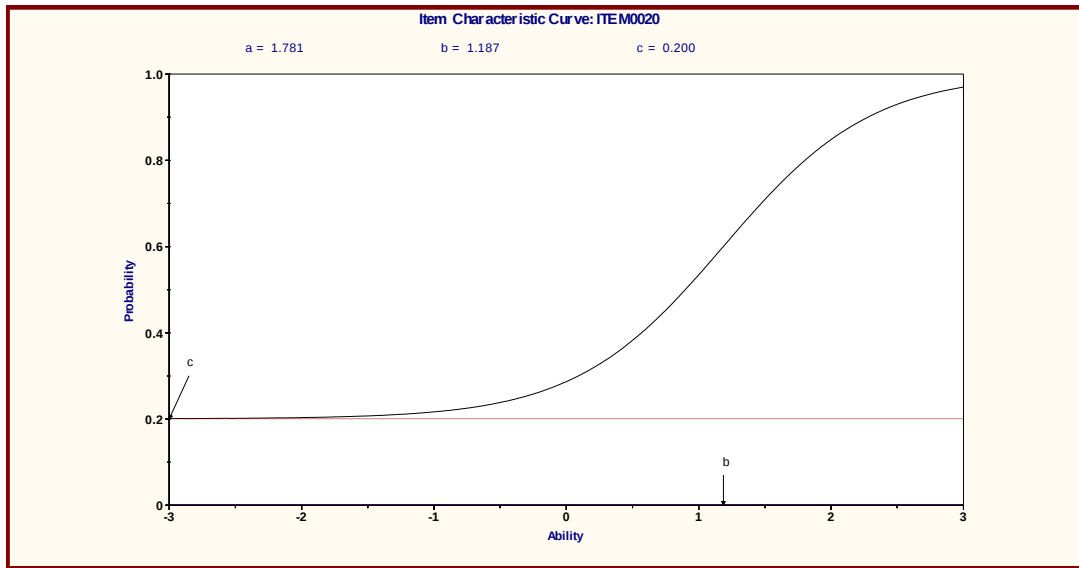


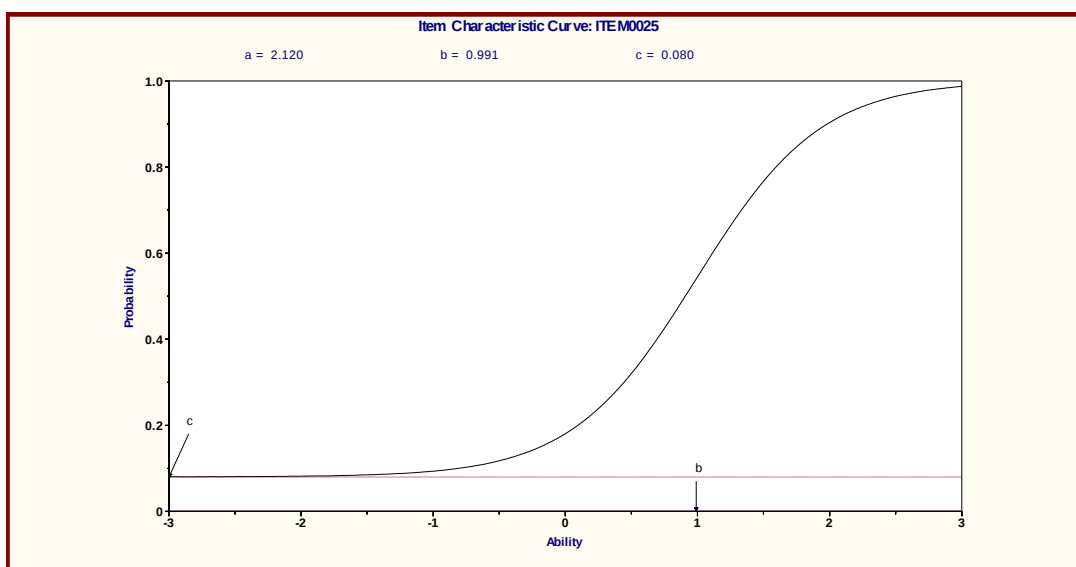
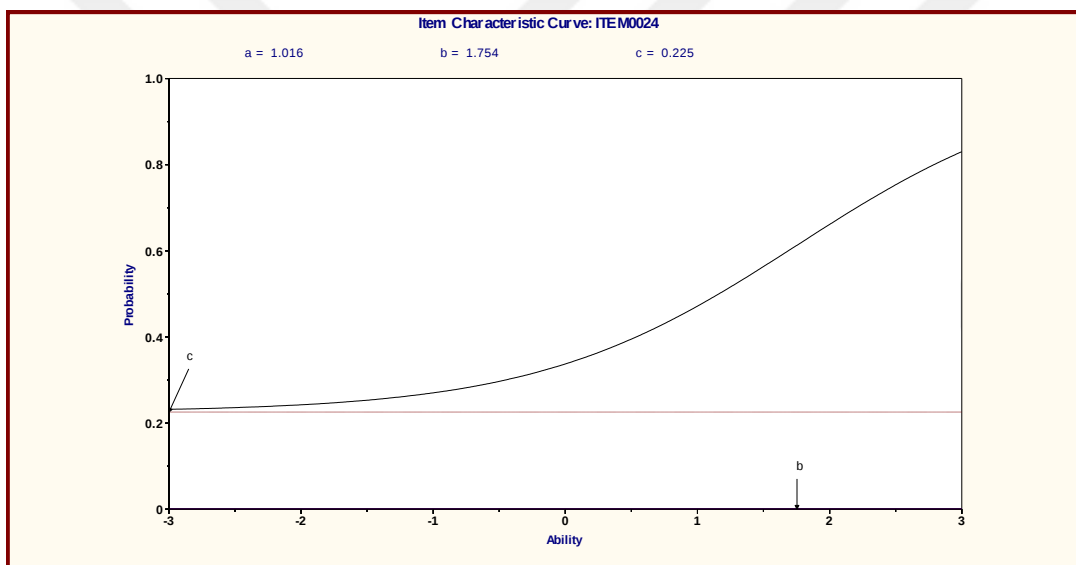
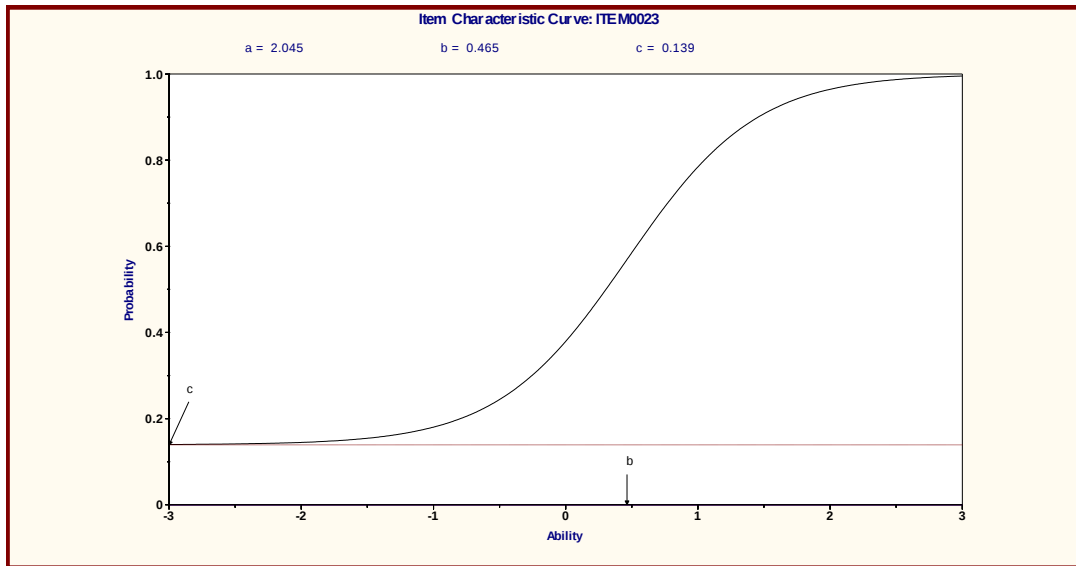


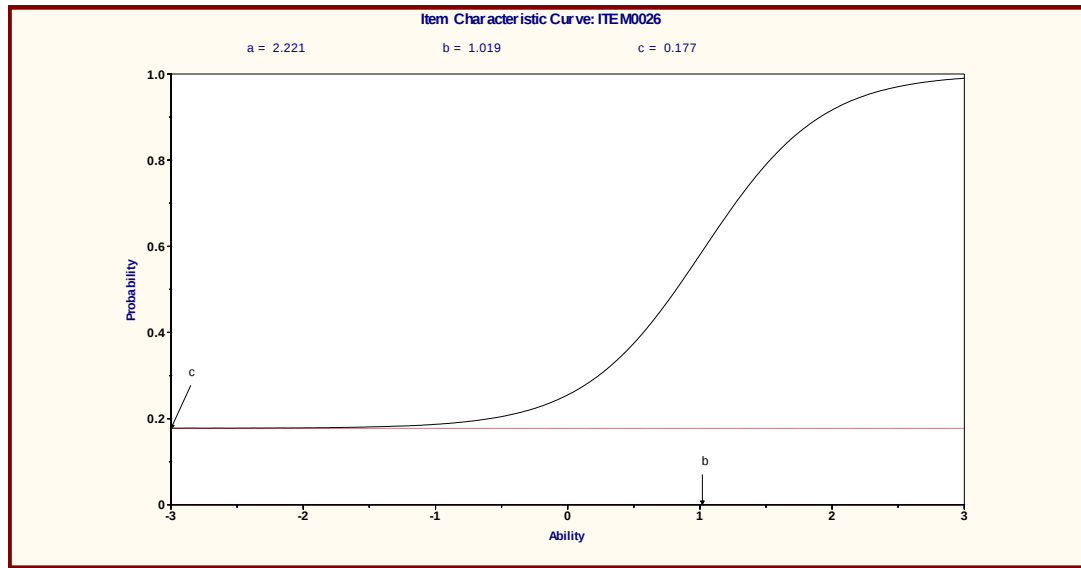












EK 5 - Form X ve Form Y'ye ait Her Bir Model Altındaki Madde Uyumu

FORM X

1PLM için

	X ²	Pr(>X ²)
1	14,595	0,748
2	30,2957	0,0482
3	23,114	0,2324
4	79,5078	<0,0001
5	36,3261	0,0096
6	23,7029	0,2078
7	15,7441	0,6743
8	32,6556	0,0263
9	30,8463	0,042
10	20,9603	0,339
11	65,9415	<0,0001
12	15,3125	0,7025
13	40,3168	0,003
14	82,8479	<0,0001
15	52,497	0,0001
16	37,222	0,0074
17	40,8296	0,0025
18	30,186	0,0495
19	26,1074	0,1272
20	17,6783	0,544
21	16,6596	0,6129
22	33,007	0,024
23	46,3987	0,0004
24	13,4423	0,8152
25	29,3575	0,0606
26	21,0913	0,3318

2PLM için

	X ²	Pr(>X ²)
1	25,9261	0,5771
2	59,965	0,0004
3	17,3675	0,941
4	31,0633	0,3143
5	44,8709	0,0227
6	23,2229	0,7218
7	38,6676	0,0864
8	39,2795	0,0765
9	36,1633	0,1385
10	27,5965	0,486
11	16,8767	0,951
12	66,2505	<0,0001
13	27,9988	0,4645
14	33,1347	0,2308
15	25,364	0,608
16	30,7869	0,3266
17	31,9346	0,2772
18	36,0264	0,1419
19	27,1134	0,5121
20	20,6507	0,8396
21	26,8443	0,5267
22	37,916	0,1
23	28,1822	0,4548
24	25,967	0,5749
25	37,5884	0,1064
26	27,4491	0,4939

3PLM için

	X ²	Pr(>X ²)
1	27,2008	0,453
2	27,3839	0,4432
3	27,4005	0,4424
4	20,0067	0,8305
5	59,0952	0,0003
6	27,1261	0,457
7	23,1842	0,6751
8	36,0371	0,1145
9	26,2566	0,5044
10	30,1465	0,3076
11	15,5116	0,9616
12	30,0296	0,3128
13	31,0779	0,268
14	32,6842	0,2077
15	34,2221	0,1597
16	24,6624	0,5934
17	30,3793	0,2974
18	23,0572	0,6819
19	17,1984	0,9261
20	16,4308	0,9442
21	37,4557	0,0869
22	36,7325	0,1002
23	29,2932	0,3468
24	21,1785	0,7779
25	46,2442	0,012
26	40,6562	0,0444

FORM Y

1PLM için

	X ² Pr	(>X ²)
1	29,6586	0,0563
2	93,4118	<0,0001
3	43,0572	0,0013
4	29,9507	0,0524
5	27,109	0,1021
6	18,6324	0,4806
7	23,7412	0,2063
8	37,7386	0,0064
9	21,0754	0,3326
10	33,7878	0,0195
11	49,2805	0,0002
12	31,4752	0,0358
13	44,9409	0,0007
14	26,1077	0,1272
15	79,6024	<0,0001
16	79,6024	<0,0001
17	31,7034	0,0337
18	15,5785	0,6852
19	24,1562	0,1902
20	10,7418	0,9322
21	14,9909	0,7232
22	29,1604	0,0635
23	25,8494	0,1344
24	39,2872	0,0041
25	38,6158	0,005
26	23,1312	0,2316

2PLM için

	X ² Pr	(>X ²)
1	47,738	0,0114
2	43,7655	0,0293
3	26,5364	0,5436
4	43,87	0,0286
5	23,7993	0,6921
6	31,8543	0,2805
7	44,0795	0,0273
8	56,9933	0,001
9	35,6646	0,1513
10	46,7213	0,0147
11	39,4793	0,0735
12	52,7626	0,0031
13	52,9971	0,0029
14	37,3323	0,1117
15	36,097	0,1401
16	36,097	0,1401
17	35,2608	0,1624
18	34,8193	0,1752
19	50,4935	0,0057
20	40,0351	0,0657
21	27,6836	0,4813
22	36,639	0,127
23	71,7596	<0,0001
24	18,3858	0,916
25	33,3047	0,2247
26	41,5044	0,0482

3PLM için

	X ²	Pr(>X ²)
1	31,805	0,2394
2	32,2021	0,2247
3	30,6875	0,2842
4	38,6351	0,0684
5	28,428	0,3892
6	25,4281	0,5505
7	44,9137	0,0166
8	59,6961	0,0003
9	39,3119	0,0593
10	39,1788	0,061
11	27,7925	0,4217
12	45,7396	0,0136
13	45,4222	0,0147
14	30,6961	0,2838
15	350,448	<0,0001
16	64,524	<0,0001
17	36,9653	0,0957
18	28,907	0,3654
19	31,5292	0,25
20	29,3273	0,3452
21	26,7116	0,4794
22	43,3515	0,0241
23	54,7349	0,0012
24	34,8995	0,1414
25	32,7004	0,2071
26	34,9556	0,14