

174891

**ULUSLARARASI OKUMA BECERİLERİNDE GELİŞİM PROJESİ (PIRLS)
2001 TESTİNİN PSİKOMETRİK ÖZELLİKLERİNİN İNCELENMESİ:
TÜRKİYE-AMERİKA BİRLEŞİK DEVLETLERİ ÖRNEĞİ**

Tuncay ÖĞRETMEN

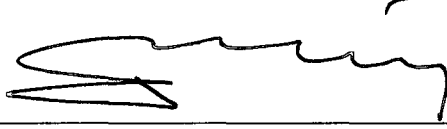
Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü
Eğitim Bilimleri Anabilim Dalı
Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Doktora Tezi

Ankara, 2006

KABUL VE ONAY

Tuncay Öğretmen tarafından hazırlanan “Uluslararası Okuma Becerilerinde Gelişim Projesi (PIRLS) 2001 Testinin Psikometrik Özelliklerinin İncelenmesi: Türkiye-Amerika Birleşik Devletleri Örneği” başlıklı bu çalışma, 31.05.2006 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Doktora Tezi olarak kabul edilmiştir.



Prof. Dr. Nizamettin KOÇ (Başkan)



Yrd. Doç. Dr. Hülya KELECİOĞLU (Danışman)



Doç. Dr. Aysun UMAY

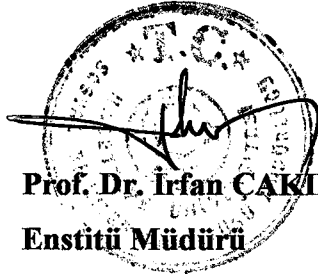


Doç. Dr. Selahattin GELBAL



Doç. Dr. Şener BÜYÜKÖZTÜRK

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.



Prof. Dr. İrfan ÇAKIN

Enstitü Müdürü

BİLDİRİM

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

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

31.05.2006

t. Öğretmen

Tuncay ÖĞRETMEN

ÖZET

ÖĞRETMEN, Tuncay. Uluslararası Okuma Becerilerinde Gelişim Projesi (PIRLS) 2001 Testinin Psikometrik Özelliklerinin İncelenmesi: Türkiye-Amerika Birleşik Devletleri Örneği, Doktora Tezi, Ankara, 2006.

Bu çalışmada, ülkemizin 2001 yılında katıldığı Uluslararası Okuma Becerilerinde Gelişim Projesi (PIRLS) kapsamında uygulanan okuma parçalarına ait testlerin psikometrik özellikleri Amerika Birleşik Devletleri ve Türkiye örneklemi bağlamında karşılaştırmalı olarak incelenmiştir.

Araştırmanın birinci aşamasında, PIRLS okuma parçalarına ait testlerin ölçtüğü varsayılan psikolojik yapıların kültürler göre eş değer olup olmadığı çok örneklemlili doğrulayıcı faktör analizi tekniği kullanılarak incelenmiştir. İkinci aşamada ise, test maddelerinin kültürler göre değişen madde fonksiyonu içerip içermediği araştırılmıştır. Bu amaçla, Madde Tepki Kuramı çerçevesinde parametre karşılaştırma ve olabilirlik oranı testi model karşılaştırma yöntemleri ile değişen madde fonksiyonu analizleri yapılmıştır.

Çok örneklemlili doğrulayıcı faktör analizleri sonucunda test maddelerinin ölçtüğü varsayılan kavrama süreçlerine ait faktör yüklerine ve hata varyanslarına ait değerlerin kaynak kültürden elde edilen ölçmeleri ile hedef kültürden elde edilen ölçmeleri arasında anlamlı bir farklılık olduğu ve bu testlerin yapılarının kültürler arasında ölçme eşdeğerliğine sahip olmadığı gözlenmiştir.

Parametre karşılaştırma ve olabilirlik oranı testi model karşılaştırma yöntemleri ile yapılan değişen madde fonksiyonu analizleri sonucunda her iki yöntem göre de PIRLS testlerine ait maddelerin çoğunun kültürler arası karşılaştırmada değişen madde fonksiyonu içerdiği ortaya çıkmıştır

Anahtar Sözcükler

PIRLS, Kısmi Puan Modeli, Değişen Madde Fonksiyonu, Çok Örneklemlili Doğrulayıcı Faktör Analizi, Olabilirlik Oranı Testi Model Karşılaştırma Yöntemi

ABSTRACT

ÖĞRETMEN, Tuncay. Ph.D. Dissertation, The investigation of Psychometric Properties of the Test of Progress in International Reading Literacy (PIRLS) 2001: The Model of Turkey-United States of America, Ankara, 2006.

In this research, psychometric properties of the tests related to reading passages applied in the scope of Progress in International Reading Literacy Study, Turkey joined this Project in 2001, were examined comparatively at the case of USA and Turkey sampling.

At the first stage of research, the assumption of the tests concerning the PIRLS's reading passages measured the psychologic cunstructs as to cultures are equivalence to each other or not was examined by the use of multi sampling confirmatory factor analysis. At the second stage, the test items considering cultures include differential item functioning or not was investigated. For this purpose, in the framework of Item Response Theory, differential item functioning analysis were performed with the parameter comparison and likelihood ratio test model comparison methods.

The result of the multi sampling confirmatory analysis indicated that there was a significant values's difference of processes of comprehensions's, which were assumed to measure by the test items, factor loadings and error variances obtained from the source culture and target culture. In addition, the constructs of the these test did not have any measurement equivalent between the cultures.

At the end of the analyses it was observed that the most of the PIRLS tests's items included differential item functioning considering cross-culture comparison.

Key Words

PIRLS, Partial Credit Model, Differential Item Functioning, Multi Sampling Confirmatory Factor Analysis, Likelihood Ratio Test Model Comparison Method

İÇİNDEKİLER

	SAYFA
KABUL ve ONAY	
BİLDİRİM	
ÖZET.....	i
ABSTRACT.....	ii
İÇİNDEKİLER.....	iii
TABLolar.....	v
BÖLÜM I.....	1
GİRİŞ.....	1
1.1 Problem Durumu.....	1
1.2 Yapısal Eşitlik Modeli ve Çok Örneklemli Doğrulayıcı Faktör Analizi Hakkında Genel Bilgiler.....	13
1.3 Madde Tepki Kuramı Hakkında Genel Bilgiler.....	20
1.4 Değişen Madde Fonksiyonu ve Analiz Yöntemleri.....	25
1.5 Problem Cümlesi.....	30
1.6 Sınırlılıklar.....	31
1.7 Kısaltmalar.....	31
1.8 Araştırmanın Amacı ve Önemi.....	32
1.9 İlgili Araştırmalar.....	32
BÖLÜM II.....	41
YÖNTEM.....	41
2.1. Araştırmanın Türü.....	41
2.2 Araştırmanın Yapıldığı Grup.....	41
2.3. Verilerin Elde Edilmesi.....	42
2.4 Verilerin Çözümlemesi.....	42
2.5 Analizlerde Kullanılan İstatistik Programları.....	52
BÖLÜM III.....	56
BULGULAR ve YORUMLAR.....	56
3.1. Alt Problem 1a.'ya İlişkin Bulgular ve Yorumlar.....	56
3.2. Alt Problem 1b.'ye İlişkin Bulgular ve Yorumlar.....	59

3.3. Alt Problem 1c.'ye İlişkin Bulgular ve Yorumlar.....	61
3.4. Alt Problem 1d.'ye İlişkin Bulgular ve Yorumlar.....	64
3.4. Alt Problem 2a.'ya İlişkin Bulgular ve Yorumlar.....	68
3.4. Alt Problem 2b.'ye İlişkin Bulgular ve Yorumlar.....	80
BÖLÜM IV.....	108
SONUÇ ve ÖNERİLER.....	108
Sonuçlar.....	108
1. Alt Probleme İlişkin Sonuçlar.....	108
2. Alt Probleme İlişkin Sonuçlar.....	109
Öneriler.....	111
KAYNAKÇA.....	113
ÖZGEÇMİŞ.....	122
EKLER.....	123
Ek 1. LISREL 8.54 Programında Sıfır Kodlu PIRLS Testi İçin Oluşturulan Çok Örneklemli Doğrulayıcı Faktör Analizi Komut Dosyaları.....	123
Ek 2. PARSCALE Programında Parametre Karşılaştırma Yöntemi ile Yapılan Değişen Madde Fonksiyonu Analizi İçin Yazılan Komut Dosyası ve Kısaltılmış Analiz Sonuçları.....	125
Ek 3. MULTILOG Programında Olabilirlik Oranı Testi Model Karşılaştırma Yöntemi ile Yapılan Değişen Madde Fonksiyonu Analizi İçin Yazılan Komut Dosyası ve Kısaltılmış Analiz Sonuçları.....	130

TABLOLAR

Tablo 1.1	PIRLS Okuma Becerileri Test Planı	5
Tablo 2.1	Okuma Parçalarına Ait Değerlendirme Testlerine İçin ABD ve Türkiye Örneklemelerinde Temel Bileşenler Faktör Analizi Sonuçları	45
Tablo 3.1	Flowers ve Antarctica Okuma Parçalarından Oluşan Üç Kodlu Alt Test Puanlarının Dağılımına Ait Betimsel İstatistikler	56
Tablo 3.2	Üç Kodlu Alt Testte Bulunan Maddelerin Ölçtüğü Varsayılan Kavrama Süreçlerine Göre Dağılımlarını Gösteren Model	57
Tablo 3.3	Modeller ve Karşılaştırmalarına Ait Çok Örneklemli DFA Sonuçları	58
Tablo 3.4	Mice ve River Okuma Parçalarından Oluşan Altı Kodlu Alt Test Puanlarının Dağılımına Ait Betimsel İstatistikler	59
Tablo 3.5	Altı Kodlu Alt Testte Bulunan Maddelerin Ölçtüğü Varsayılan Kavrama Süreçlerine Göre Dağılımlarını Gösteren Model	60
Tablo 3.6	Modeller ve Karşılaştırmalarına Ait Çok Örneklemli DFA Sonuçları	60
Tablo 3.7	Clay ve Leonardo Okuma Parçalarından Oluşan Sekiz Kodlu Alt Test Puanlarının Dağılımına Ait Betimsel İstatistikler	62
Tablo 3.8	Sekiz Kodlu Alt Testte Bulunan Maddelerin Ölçtüğü Varsayılan Kavrama Süreçlerine Göre Dağılımlarını Gösteren Model	63
Tablo 3.9	Modeller ve Karşılaştırmalarına Ait Çok Örneklemli DFA Sonuçları	63
Tablo 3.10	Hare ve Pufflings Okuma Parçalarından Oluşan Sıfır Kodlu Alt Test Puanlarının Dağılımına Ait Betimsel İstatistikler	65
Tablo 3.11	Sıfır Kodlu Alt Testte Bulunan Maddelerin Ölçtüğü Varsayılan Kavrama Süreçlerine Göre Dağılımlarını Gösteren Model	66
Tablo 3.12	Modeller ve Karşılaştırmalarına Ait Çok Örneklemli DFA Sonuçları	66
Tablo 3.13	Flowers Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular	69
Tablo 3.14	Antarctica Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular	70

Tablo 3.15	Mice Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular	71
Tablo 3.16	River Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular	72
Tablo 3.17	Clay Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular	74
Tablo 3.18	Leonardo Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular	75
Tablo 3.19	Hare Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular	77
Tablo 3.20	Pufflings Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular	78
Tablo 3.21	Okuma Parçalarına Göre Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizi Sonuçları	79
Tablo 3.22	Flowers Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları	81
Tablo 3.23	Flowers Okuma Parçası Değerlendirme Testi I. , II. Ve III. Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modellere Göre Karşılaştırma Sonuçları	83
Tablo 3.24	Flowers Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin -2logL Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları	84
Tablo 3.25	Antarctica Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları	85
Tablo 3.26	Antarctica Okuma Parçası Değerlendirme Testi I. , II. ve III Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modellere Göre Karşılaştırma Sonuçları	86
Tablo 3.27	Antarctica Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin -2logL Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları	87
Tablo 3.28	Mice Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları	88
Tablo 3.29	Mice Okuma Parçası Değerlendirme Testi I., II. ve III. Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modellere Göre Karşılaştırma Sonuçları	89

Tablo 3.30	Mice Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin $-2\log L$ Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları	90
Tablo 3.31	River Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları	92
Tablo 3.32	River Okuma Parçası Değerlendirme Testi I. , II. ,III. ve IV. Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modeller Karşılaştırma Sonuçları	93
Tablo 3.33	River Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin $-2\log L$ Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları	94
Tablo 3.34	Clay Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları	95
Tablo 3.35	Clay Okuma Parçası Değerlendirme Testi I. ,II. ,III. ve IV. Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modeller Karşılaştırma Sonuçları	96
Tablo 3.36	Clay Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin $-2\log L$ Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları	97
Tablo 3.37	Leonardo Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları	98
Tablo 3.38	Leonardo Okuma Parçası Değerlendirme Testi I. ,II. ,III. ve IV. Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modeller Karşılaştırma Sonuçları	99
Tablo 3.39	Leonardo Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin $-2\log L$ Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları	100
Tablo 3.40	Hare Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin I. İterasyon Madde Parametre Tahminleri Ve Olabilirlik Oranı Testi Sonuçları	101
Tablo 3.41	Hare Okuma Parçası Değerlendirme Testi I. ve II. Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modeller Karşılaştırma Sonuçları	102
Tablo 3.42	Hare Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin $-2\log L$ Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları	103
Tablo 3.43	Pufflings Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları	104
Tablo 3.44	Pufflings Okuma Parçası Değerlendirme Testi I. , II. ve III. Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modeller Karşılaştırma Sonuçları	105

Tablo 3.45	Pufflings Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin $-2\log L$ Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları	106
Tablo 3.46	Okuma Parçalarına Göre Olabilirlik Oranı Testi Model Karşılaştırma Yöntemi İle Yapılan DMF Analizi Sonuçları	107
Tablo 4.1	Kavrama Süreçlerine Göre DMF İçeren Maddelerin Dağılımı	109



BÖLÜM I

GİRİŞ

Bu bölümde problem durumu, yapısal eşitlik modelleri kapsamında çok örneklemli doğrulayıcı faktör analizi, madde tepki kuramı hakkında genel bilgiler, madde tepki kuramına dayalı olarak bu tez kapsamında alt problemlerin çözümü için kullanılan değişen madde fonksiyonu analiz yöntemleri, problem cümlesi, sınırlılıklar, kısaltmalar, araştırmanın amacı, önemi ve ilgili araştırmalar yer almaktadır.

1.1 Problem Durumu

Bilim ve teknolojiadaki gelişmeler toplumsal, kültürel, siyasal ve ekonomik alanlarda çok hızlı değişmelere yol açmaktadır. Hızla artan bilgi ve gelişen teknoloji, bireylerin eğitim ihtiyacını da artırmaktadır. Artan bu ihtiyaçlar ülkelerin eğitim sistemlerindeki sorunları ve gelişmeleri sadece ulusal düzeyde değil uluslararası düzeyde de değerlendirme gereksinimleri ortaya çıkarmaktadır. Bu nedenle ülkemiz de 1999 yılından itibaren üyesi olduğu kuruluşlar tarafından düzenlenen uluslararası karşılaştırmalı projelere Milli Eğitim Bakanlığı Eğitimi Araştırma Geliştirme Dairesi (EARGED) aracılığı ile katılmaktadır.

Bu projeler, Uluslararası Eğitim Başarılarını Belirleme Kuruluşu (IEA International Association for the Evaluation of Educational Achievement) tarafından yürütülmekte olan Üçüncü Uluslararası Matematik ve Fen Çalışması (Third International Mathematics and Science Study – TIMSS), İktisadi İşbirliği ve Kalkınma Teşkilatı OECD tarafından yürütülen Uluslararası Öğrenci Başarı Belirleme Programı (Programme for International Student Assessment – PISA) ve Uluslararası Okuma Becerilerinde Gelişim Projesi (Progress in International Reading Literacy Study (PIRLS) adlı projelerdir. Söz konusu projeler katılımcı ülkelerin eğitim uzmanlarına ve araştırmacılara kendi eğitim sistemlerini daha iyi anlamada ve değerlendirmede ulusal ve uluslararası ölçekte önemli bir olanak sağlamaktadır (Mullis ve diğ., 2003). Bu çalışmalardan elde edilen sonuçlar uluslararası karşılaştırmalara imkan verecek kadar

geniş ve detaylıdır. Ülkelere ait örneklemelerin seçkisiz olarak oluşturulması elde edilen sonuçların genellenebilirliğini arttırmaktadır.

Bahsedilen çalışmalar kapsamında katılımcı ülkelerde uygulanacak olan testler genellikle İngilizce olarak hazırlanmış ve diğer ülke dillerine çevrilmiştir.

1999 yılında TIMSS projesi kapsamında yapılan sınavlarda Türkiye 38 ülke arasında fen bilgisinde 33. matematikte ise 31 sırada yer almıştır. 2001 yılında PIRLS projesi kapsamında yapılan sınav sonuçlarına göre ise Türkiye 35 katılımcı ülke arasında 28. sırada yer almıştır. Türkiye'nin puanı uluslararası ortalamanın oldukça altında bulunmaktadır. TIMSS 1999 Türkiye Raporunda, uluslararası platformda öğrenci başarıları açısından beklendik düzeye ulaşamaması, sorunun öğretim programından çok okullardaki öğretim etkinliklerinden ve öğrenci başarı algılarıyla, tutumlarından kaynakladığı şeklinde açıklanmaktadır (TIMSS 1999 Türkiye Raporu, 2003:18).

Öğrenci başarıları açısından beklendik düzeye ulaşamamasına yol açan çok çeşitli nedenler olabilir. Bu nedenlerden biri de çevirisi yapılmış bir testin iki versiyonu arasında psikometrik denkliğin sağlanmasındaki güçlüklerdir. Temel sorun, psikolojik bir ölçeği oluşturan maddelerin bir dildeki formu, bir diğer dilde de aynı psikometrik anlama gelmekte midir? Var olan çeviri değerlendirme işlemleri, bazı çevirilere uygulandığında yararlı olmaktadır; bazılarının ise, çevrilmiş psikolojik ölçeklere uygulanabilirlikleri sınırlıdır. Bu nedenle madde ve ölçeklerin kaynak ve hedef kültürde uygulanan versiyonlarının eşitliği hakkında, yaygın olarak kullanılan tekniklere dayanılarak verilen kararlar ciddi yanlışlar içerebilir (Hulin, 1987). Bu ise, farklı kültürlerdeki örtük özelliklere ilişkin varsayımların karşılanamaması ve ölçeklerin eşitliğinin sağlanamaması anlamına gelebilir. Brislin'e göre (1970 ve 1976) orijinal dilin gramer şekillerini kaybetmeyen çevirilerin tekrar orijinal dile çevrilmesi kolaydır, fakat bu çeviriler, hedef kültürdeki tek dil bilen insanlar için anlamlı olmayabilir.

Çevirisi yapılmış iki versiyon arasındaki psikometrik eşitsizliğin bir diğer kaynağı da, örtük özelliğin hedef kültürdeki kültürel eşitliğinin sağlanamayışındır (Hulin ve Mayer, 1986). Özelliği ölçmek için kullanılan maddelerden ziyade, ölçülmek istenilen özelliğin

kültürel eşitsizliği, örtük özelliklerin deneysel anlamlarının kültürler arası incelemelerinde ortaya çıkabilmektedir.

Kültürel denklik ile ilgili bir diğer sorun ise örtük özelliğin ölçülmesinde kullanılan uyarıcıların (test maddelerinin) hedef kültüre uygun olmaması ya da kötü çevrilmiş olabileceğidir. Eğer ölçülen özellik, her iki dil ve kültürdeki bireyler için anlamlı ve denk ise kültürel denklikle ilgili sorunlar, özellikleri ölçmek için kullanılan maddelerden kaynaklanabilir (Candell ve Roznowski, 1984).

Bazı durumlarda ise maddeler doğru çevrilmiştir fakat kaynak dildeki özellik veya davranışın hedef kültürde karşılığı yoktur. Farklı kültürlerde ortak olan, aynı anlama gelen kavram ve özelliklere, “*etic*” kavramlar; kültüre özgü kavramlara ise “*emic*” kavramlar adı verilir (Osgood, May, Miron, 1975). Belli bir ölçeğin kaynak versiyonunda emic kavramlar içeren birçok maddenin bulunması kültürler arası karşılaştırmaların geçerliğini düşürebilir.

Sireci (1997), eğitimsel testlerin farklı dildeki versiyonlarını cevaplayan farklı kültürlerdeki öğrencilerin başarılarının karşılaştırılmasının çok güç olduğundan bahsetmektedir. Çeviri süreçlerine karışan ve istenmeyen bazı yan etkiler, diller arasında test maddelerinin farklı güçlük dereceleri üretmesine neden olmaktadır. Örneğin, bir test maddesi görelî olarak bir kültürde kolay bir madde gibi işlerken diğer bir dildeki uyarlaması çok zor bir madde olabilmektedir.

Hui ve Triandis (1985) psikolojik yapıların kültürler arası karşılaştırmalarında dört çeşit denkliğin olmasından bahsetmektedirler. Bunlar;

- a) Kavramsal / fonksiyonel eşitlik,
- b) Psikolojik yapıların operasyonel eşitliği,
- c) Madde parametrelerinin eşitliği
- d) Ölçekleme (scalar equivalence) eşitliğidir.

Bu eşitlik türlerinin bir arada sağlanması çok zordur. Sadece bazılarının sağlanması ise psikolojik ölçeğin kültürler arası karşılaştırmaları için yeterli değildir. Kültürler arası karşılaştırma için literatürde önerilen bazı teknikler aşağıda verilmiştir.

- a) Doğrudan karşılaştırma ve rutin çeviri (Gordon ve Kikuchi, 1966)
- b) Regresyon metodları (Poortinga, 1975)
- c) Birlikte puanlama (coscoring) yöntemleri (Cattell, 1957)
- d) Madde tepki kuramı yaklaşımları (Linn, Levine, Hastings ve Wardrop, 1981)
- e) Tepki örüntü metodu (Irvine ve Carroll, 1980; Angoff, 1972).
- f) Çeviri teknikleri (Brislin, 1976 ve 1980)
- g) İç yapı uyumu teknikleri
 - Faktör analizi (Irvine ve Carroll, 1980)
 - Çok boyutlu ölçekleme (Hui ve Triandis, 1985)
 - Doğrulayıcı faktör analizi (Jöreskog, 1971)
 - Korelasyon matrislerinin karşılaştırılması (Jennrich, 1970)
- h) Etic – Emic karışım yaklaşımı (Osgood's 1967; Triandis 1972)

PIRLS projesi:

Uluslararası Okuma Becerilerinde Gelişim Projesine (PIRLS) Türkiye'de dahil olmak üzere 35 ülke katılmaktadır. Bu proje ile ilköğretim 4. sınıf (9-10 yaş grubu) öğrencilerinin okuma becerileri, okuma alışkanlıkları, öğrencilere okuma becerisini kazandırmak için öğretmenlerin uyguladıkları öğretim yöntemleri, öğretim materyallerinin yeterli olup olmadığı, öğrencilerin okuma becerilerini kazanmalarında ailelerin katkıları gibi konular, uluslararası standart test ve anketlerle belirlenmekte ve projeye katılan ülkelerin verileri ile karşılaştırılarak benzerlik ve farklılıklar ortaya çıkarılmaktadır. Bu projenin temel hedeflerinden bir tanesi de ülkelerin sonuçlardan elde ettikleri çıkarımları sisteme yansıtıktan sonra eğitim sistemi ile ilgili olarak zaman içerisinde nelerin değiştiğinin incelenmesidir. Bu nedenle PIRLS testlerinin analizi ile özellikle diğer ülkelerle karşılaştırmalı olarak yapılacak analizlerin okuma becerileri yönünden hem eğitim sistemimizin var olan durumunu ortaya koyması, hem de testlerin

ve test maddelerinin kültürlere göre yapısal ve psikometrik özelliklerinin neler olduğunun araştırılması bakımından çok önemli olduğu anlaşılmaktadır.

PIRLS Okuma Testlerinin Genel Çerçevesi :

PIRLS;

- Okuma amaçları,
- Kavrama süreçleri,
- Okuma alışkanlıkları ve okumaya yönelik tutumlar,
olmak üzere okumanın üç yönüyle ilgilenmektedir.

Okuma amaçları ve kavrama süreçleri PIRLS değerlendirmesinin temellerini oluşturmaktadır. Tablo 1. bu ikisinin birbiriyle etkileşimini göstermektedir.

Tablo 1.1 PIRLS Okuma Becerileri Test Planı

Okuma Amaçları

Okuma deneyimi	%50
Bilgiyi elde etme ve kullanabilme	%50

Kavrama Süreçleri

Açık bir şekilde ifade edilmiş bilgi ve fikirlerin üzerinde durma ve çıkarımlar yapma	%20
Doğrudan çıkarımlar yapma	%30
Fikir ve bilgileri birleştirme ve yorumlama	%30
Metnin öğelerini, içeriğini ve dilini inceleme ve değerlendirme	%20

A-Okuma Amaçları:

PIRLS, genç öğrencilerin okul içinde ve dışında okumalarında daha fazla yer tutan iki amaç üzerine odaklanmıştır. Bunlar:

- Okuma deneyimi kazanmak için okuma,
- Bilgiyi elde etme ve kullanabilme için okumadır.

Bu iki tür okuma erken yaşlar için önem taşımaktadır. PIRLS değerlendirmesinde bu ikisine de aynı oranda yer verilmiştir. Değerlendirme okuma amaçlarına göre farklılık göstermekle birlikte, okurların her iki amaç için kullandığı süreçler ve stratejiler daha benzerdir.

Okuma deneyimi kazanmak amacıyla gerçekleştirilen okumada, okur, kurgulanmış olayların, mekanların, karakterlerin, atmosferin, duygu ve fikirlerin içinde yer almakta ve dilin kendisinden zevk almaktadır. Metni anlamak ve onaylamak için okur kendi deneyimlerini, düşüncelerini, edebi form bilgisini de metne taşımaktadır. Genç okurlar için edebiyat onlara daha önce hissetmedikleri duyguları, karşılaşmadıkları durumları hayali olarak yaşamaları fırsatını sunmaktadır. PIRLS’de kullanılan temel formlardan biri öykülendirmedir.

Bilgiyi elde etme ve kullanabilme için okumada, okur kurgusal bir dünyadan çok gerçek yaşam boyutlarıyla iç içedir. Bilgi veren metin yardımıyla, okur dünyanın nasıl olduğunu, nasıl olacağını, nesnelerin ve oluşumların işleyişini anlayabilir. Okur sadece bilgiyi almakla kalmaz onu mantıksal çıkarımlarda kullanır ve eyleme dönüştürür. Bilgi veren metinleri baştan sona okumak gerekmez. Okur kendisi için gerekli kısmı seçebilir. Bu metinler farklı formlarda karşımıza çıkabilir. Ancak en önemli ayrım bilginin kronolojik sırayla verilip verilmediğidir. Bazen tek bir metinde ikisi de bir arada bulunabilir.

B-Kavrama Süreçleri:

Okurlar açık bir şekilde ifade edilmiş bilgi ve fikirlerin üzerinde durmak ve çıkarımlar yapmak, doğrudan çıkarımlar yapmak, fikir ve bilgileri birleştirip yorumlamak, metnin öğelerini, içeriğini ve dilini inceleyip değerlendirmek gibi farklı yollarla anlamlar çıkarırlar. Kavrama süreçlerinin dört türü aşağıda açıklanmaktadır. Bu süreçler PIRLS’de öğrencilere verilen metinler için kavrama sorularının oluşturulmasında kullanılmıştır. PIRLS’de kullanılan testlerdeki soruların her biri, aşağıda açıklanan süreçlere yönelik hazırlanmış ve tanımlanmıştır. Böyle bir gruplama öğrencilerin yazılı metinlerden anlam çıkarmadaki yetenek ve becerilerini belirleme fırsatı vermektedir.

1-Açık Bir Şekilde İfade Edilmiş Bilgi ve Fikirlerin Üzerinde Durma ve Çıkarımlar Yapma:

Okurların metinde açıkça ifade edilen bilgiye gösterdiği dikkat kişiler arası farklılıklar göstermektedir. Metindeki bazı fikirler özellikle dikkat çekerken diğerleri aynı oranda çekmeyebilir. Örneğin, okurlar metnin içeriği konusunda yapılan tahminlere ya da onu okuma amaçlarıyla örtüşen ya da çelişen fikirlere odaklanabilirler. Ek olarak okurlar çoğunlukla okuma parçasında verilen bir soruyu cevaplamak için verilen metinde açıkça ifade edilen bilgilerden çıkarımlar yapma gereği duyarlar.

Açık bir şekilde ifade edilmiş bilgi ve fikirlerin üzerinde durup, çıkarımlar yaparken okurlar sorulan soruyla ilişkili kısmın içeriğini anlamak için farklı yollar kullanırlar. Metinden uygun çıkarımlar yapmak okurun sadece metinde açıkça ifade edilen bilgiyi anlamasıyla kalmamakta aynı zamanda istenilen bilgiyle nasıl bir ilişkisi olduğunu da anlamasını gerektirmektedir.

Başarılı bir çıkarım yapma oldukça ani ya da otomatik olarak metnin anlaşılmasını gerektirir. Bu süreç neredeyse hiç yorumlama gerektirmez. Anlamda doldurulması gereken boşluklar yoktur. Anlam kesindir ve metinde ifade edilmiştir. Ancak okur istenilen bilgi ile ilgili bilgiyi ayırt edebilmelidir. Bu süreçte metinde odaklanılan kısım cümle düzeyinde kalmaktadır. Okurun birçok farklı bilgiye odaklanması gerekebilir; ancak her defasında bu bilgi bir cümle ya da kelime grubunun içinde yer almaktadır. Bu süreçte yer alan okuma faaliyetleri aşağıdaki gibidir:

- Okumanın amacına ilişkin olan bilginin tanımlanması
- Belirli fikirlerin aranması
- Kelime ya da kelime gruplarının anlamlarının bulunması
- Hikayenin yeri ve zamanının vs. belirlenmesi
- Ana fikrin bulunması (açıkça ifade edilmiş ise)

2-Doğrudan Çıkarımlar Yapma:

Okurlar metinden çıkarımlar yaparken açıkça ifade edilmemiş bilgi ve fikirlerden çıkarımlar yaparlar. Çıkarım yapma okura metin üzerinde gezinmeyi ve metinde genellikle oluşan anlam boşluklarını doldurmayı gerektirir. Bu çıkarımların bazıları metinde yer alan bilgiye dayanmaktadır; okur iki ya da daha fazla fikri ya da bilgiyi birleştirmek durumundadır. Fikirler açıkça ifade edilse de aralarındaki bağlantı açıkça ifade edilmemiştir. Doğrudan çıkarımlar yapma tamamen metine dayalıdır. Metinde açıkça ifade edilmese de, anlam oldukça açıktır.

Becerikli okuyucular bu tür çıkarımları otomatik olarak yapabilir. Onlar metinde açıkça ifade edilmese bile iki ya da daha fazla bilgiyi hemen ilişkilendirebilirler. Birçok durumda yazar metni okurun doğrudan çıkarımlar yapmasına olanak sağlayacak şekilde oluşturur. Örneğin bir hikaye boyunca hikayenin kahramanının davranışı belirli bir karakter yapısını izlemekte ve okurlar bu kahramanın bakış açısı ve kişiliği konusunda aynı sonuca ulaşmaktadır. Bu süreçte, okur cümle seviyesinden fazlasına odaklanmaktadır. Burada odak noktası metnin bir kısmı olabileceği gibi daha geneli ya da metnin tamamını temsil edebilir. Buna ek olarak, bazı doğrudan çıkarımlar bölgesel ve evrensel nitelikleri birleştirmeyi gerektirebilir. Bu süreçte yer alan okuma faaliyetleri aşağıdaki gibidir:

- Neden sonuç ilişkisinin çıkarılması
- Bir dizi tartışma sonucunda ana fikrin ne olduğunun ortaya çıkarılması
- Bir zamirle ifade edilmiş ismin bulunması
- Metinde yapılan genellemelerin belirlenmesi
- İki karakter arasındaki ilişkinin tanımlanması

3-Fikir ve Bilgileri Birleştirme ve Yorumlama:

Bu süreçte okur, yerel ya da evrensel anlamlara bakarak daha doğrudan çıkarımlar yapar ya da ayrıntıları ilişkilendirerek daha genel sonuçlara ulaşabilir. Her durumda, okur cümle seviyesinin ötesinde düşünmektedir.

Okurlar metinde yer alan fikir ve bilgileri birleřtirip yorumlarken genellikle dünyayı algılama biçimlerini ortaya koyarlar. Sadece dolaylı anlatılanlarda deęil, kendi bakış açılarına dayanan yoruma açık olanlar arasında da ilişki kurarlar. Metinde yer alan fikir ve bilgileri birleřtirip yorumlarken, okurlar doğrudan çıkarımlar yapmak yerine kendi deneyimlerinden ve geçmiş bilgilerinden yararlanma gereęi duyabilirler. Bu nedenle okuyucunun fikir ve bilgileri birleřtirip yorumlarken ortaya çıkan anlam kişiden kişiye deęişiklik gösterebilir.

Bu süreçte, okurlar kişisel deneyimleri ve bilgileriyle metni birleřtirerek metni daha belirgin bir şekilde anlamaya çalışırlar. Bu süreçte yer alan okuma faaliyetleri ařaęıdaki gibidir:

- Bir metinden genel bir mesaj ya da tema çıkarılması
- Karakterlerin eylemlerine bir alternatif bulunması
- Metinde yer alan bilginin karřılařtırılması ve kıyaslanması
- Hikayede egemen olan düşüncenin tanımlanması
- Metinde yer alan bilgilerin günlük yaşamda uygulanabilirlięinin yorumlanması

4-Metnin Öğelerini, İçerięini ve Dilini İnceleme ve Deęerlendirme:

Okurlar metnin içerięini, dilini ve yazının öğelerini inceleyip deęerlendirirken üzerinde durulan, metindeki anlamı çıkarmadan, metinde yazılanları kritik bir şekilde ele almaya doğru kaymaktadır. İçerik açısından metinden anladıklarıyla günlük yaşamdaki algıları arasında yorumlar yapar; karřı çıkar, kabul eder ya da metnin içerięi konusunda tepkisiz davranır. Örneęin, okur metinde bahsedilen şeylerin doğruluęunu onaylar ya da karřı çıkar ya da dięer kaynaklarda yer alan bilgiyle karřılařtırır.

Metnin yapısı ve dili gibi öğelerini ele alırken okur metne verilen anlamın nasıl sunulduęunu inceler. Bunu yaparken kendinde var olan metin türü ve yapısı bilgisinden çıkarımlarda bulunur. Aynı zamanda yazarın bilgiyi aktarım yolları, bilginin doğruluęu konusunda yargılarda bulunur ve yazarın amacını, bakış açısını ya da becerisini sorgular.

Bu süreçte okuyucu kendini metinden ayrı tutar ve onu inceler ve değerlendirir. Metnin içeriği oldukça kişisel bir bakış açısıyla ya da nesnel olarak incelenir. Burada okur dünyayı nasıl algıladığına ve önceki bilgilerine dayanarak yorumda bulunur. Bu süreçte yer alan okuma faaliyetleri aşağıdaki gibidir:

- Anlatılan olayların gerçekte olma olasılığının değerlendirilmesi
- Yazarın nasıl sürpriz bir son hazırladığının tanımlanması
- Metindeki bilginin açıklığının ve doğruluğunun yargılanması
- Ana fikir etrafında yazarın bakış açısının ne olduğunun belirlenmesi
- Seçilen sıfatların anlamı nasıl etkilediğinin tanımlanması

PIRLS testi İngilizce olarak hazırlanmış ve 31 ülke diline çevrilmiştir. Beş ülkede iki ayrı dilde hazırlanan testler uygulanmıştır. Yedi ülkede ise testlerin bazıları bir dilde diğerleri birden fazla dilde uygulanmıştır. Türkiye uygulaması için testler Türkçe'ye çevrilerek uygulanmıştır.

Çeviri çalışmaları aşağıdaki aşamaları içermiştir.

- Kültürel adaptasyon ve çeviri için ayrıntılı rehberler geliştirilmiştir.
- Test çevirileri ulusal otoritelerin işbirliği ile birden fazla çeviriye içermiştir.
- Çevirilerin kalitesi bağımsız çalışan çeviri firmalarına onaylatılarak kararlaştırılmıştır.
- PIRLS Uluslararası Çalışma Merkezi gerekli görülen düzeltmeleri yapmıştır.

Her ülkede, pilot uygulama (alan uygulaması) öncesi ve sonrası maddelerde ve okuma parçalarında gerekli görülen düzeltmeler yapılarak PIRLS ölçeklerinin final formları oluşturulmuş ve asıl uygulama için hazırlanmıştır. Pilot uygulamalardan elde edilen verilerin analizleri çeviri hatalarının bulunmasında ve düzeltilmesinde asıl uygulama

öncesinde ülkelere madde istatistiklerinin teşhisi yolu ile çok faydalı olmuştur (Mullis ve diğ., 2003; Gonzalez ve Kennedy, 2003).

PIRLS sonuçları ortalaması 500, standart sapması 100 olan bir standart puan formatında rapor edilmektedir. Bu puan sırasına göre Türkiye 35 katılımcı ülke arasında 28. sırada yer almıştır. Türkiye'nin ortalama standart puanı 449'dur. Türkiye'nin puanı uluslararası ortalamadan 51 puan (yaklaşık yarım standart sapma) daha düşüktür.

PIRLS ile ilgili bazı genel sonuçları şu şekilde özetlemek mümkündür:

- PIRLS 2001'e katılan ülkeler arasında en yüksek başarı puanına sahip olan ülke İsveç'tir. Hollanda, İngiltere ve Bulgaristan başarı sıralamasında bu ülkeyi takip etmektedir. Latvia, Kanada, Litvanya, Macaristan, Amerika Birleşik Devletleri, Almanya ve İtalya diğer ülkelerin pek çoğundan daha iyi bir performans göstermiştir.
- Ülkelerin tamamında, dördüncü sınıf kız öğrenciler, erkek öğrencilerden anlamlı düzeyde daha başarılı olmuştur.
- İsveç ve İngiltere okuma deneyimi amacıyla okumada en yüksek başarı uygulamasına sahip ülkelerdir. İsveç diğer ülkelerden anlamlı düzeyde daha iyi bir performans gösterirken; ikinci sırada yer alan İngiltere; Hollanda, Amerika Birleşik Devletleri ve Bulgaristan'dan anlamlı bir fark göstermemekle birlikte diğer bütün ülkelerden daha iyi bir performans göstermiştir.
- Bilgiyi elde etme ve kullanma amacıyla yapılan okumada İsveç, Hollanda ve Bulgaristan en yüksek başarı ortalamasına sahip ülkelerdir. İsveç diğer bütün ülkelerden daha iyi bir performans gösterirken, Hollanda ve Bulgaristan, Latvia ve İngiltere'den anlamlı bir fark göstermemekle birlikte diğer bütün ülkelerden daha iyi bir performans göstermiştir. Türkiye bilgiyi elde etme ve kullanma içeriği taşıyan sorularda okuma deneyimi sorularına oranla daha başarılı bir sonuç elde etmiştir.

Bu çalışmada öncelikle, PIRLS-2001 testinin ölçtüğü psikolojik yapıların kültürlere göre eş değer olup olmadığı araştırılmıştır. Yapı, insan davranışlarının niteliği hakkında yapılan soyutlamalar ya da teorik açıklamalar olarak tanımlanabilir. Bu nedenle, yapı geçerliği çalışmaları kavramsal yapının tam olarak açığa çıkarılması ile ilgilidir. Ölçek puanlarının ilişkili olduğu yapının ortaya konması ya da bu yapının varlığına ilişkin kanıtlar, ölçme aracının ölçme amacı doğrultusunda çalıştığını göstermek açısından önem taşır. Diğer bir anlatımla, bu geçerlik türü, ölçme aracının soyut bir olguyu (kavramı, boyutu vb.) ne derece doğru ölçebildiğini gösterir. Bunlar doğrudan gözlenmesi ve ölçülmesi güç, hatta olanaksız davranışlardır. Soyut kavramları (zeka, sosyal uyum vb.) ölçebilmek için önce bu kavramların içeriğinin belirlenmesi gerekir. Yapı geçerliği, bir ölçme aracının ve ondan elde edilen puanın ne anlama geldiğini araştırma sürecidir. Ölçülen kavramların içeriğinin, evrenin açıkça bilinmediği ve ölçülecek özelliğin açık-seçik tanımlanamadığı durumlarda daha çok önem kazanmaktadır. Yapı geçerliğini belirlemek için farklı teknikler vardır. Bunlardan en yaygın kullanılanı faktör analizidir (Baykul, 2000).

Araştırmacı ölçmek istediği bilişsel ya da psikolojik bir yapıyı ölçebilmek için, o yapıyı ya da kavramı ölçülebilir, gözlenebilir değişkenlerle açıklamaya çalışır. Bu amaçla, literatüre ve uzman görüşüne dayalı olarak çok sayıda madde yazar. Bu maddelerle ölçmek istediği yapıyı ölçülebilir, gözlenebilir değişkenlerle açıklamaya çalışır. Yapı geçerliğini belirleme süreci, maddelerin aynı yapıyı ölçüp ölçmediğine ilişkin kanıt toplama işidir. Yapı geçerliğini belirleme çalışmalarında, ölçme aracının boyutlarını belirlemede ve boyutları anlamlandırmada en kullanışlı ve etkili bilgiyi veren teknik faktör analizidir. Faktör analizi, aynı yapıyı ya da niteliği ölçen değişkenleri bir araya toplayarak ölçmeyi az sayıda faktör ile açıklamayı amaçlayan bir istatistiksel tekniktir. Faktör analizi, bir faktörleştirme ya da ortak faktör adı verilen yeni kavramları (değişkenleri) ortaya çıkarma ya da maddenin faktör yük değerlerini kullanarak kavramların işlevsel tanımlarını elde etme süreci olarak da tanımlanmaktadır (Şencan, 2005).

Son yıllarda psikolojik testlerin yapı geçerliğinin bir göstergesi olarak açıklayıcı (exploratory) faktör analizleri sonucunda belirlenen yapıların doğrulayıcı faktör analizi

tekniki kullanılarak test edildiği analizler yaygınlaşmıştır (Sümer, 2000). Doğrulatoryı faktör analizi, hem maddelerin temsil gücü hakkında bilgi vermekte hem de boyutları birbirleri ile ilişki içinde deęerlendirerek bunların ölçülen psikolojik özellikler ile ne oranda örtüştüğünü göstermektedir. Aynı ölçeğin farklı örneklem gruplarından elde edilen uygulamalarında faktör yapılarının ve ölçmelerin gruplar arası deęişmezliğini test etmek için ise çok örneklemlı doğrulatoryı faktör analizi teknięi yaygın bir şekilde kullanılmaktadır (Kaplan, 2000; Kline, 1998; Schumacker ve Lomax, 2004; Jöreskog and Sörbom, 1993; Lubke ve Muthén, 2004).

1.2 Yapısal Eşitlik Modeli ve Çok Örneklemlı Doğrulatoryı Faktör Analizi Hakkında Genel Bilgiler

Yapısal eşitlik modelleri (YEM) genellikle hipotetik yapıların gizil deęişken olarak tanımlandığı path modelleri kullanılarak diyagramlarla ifade edilmektedir. Genellikle iki parçası bulunmaktadır: ölçme modeli, yapısal eşitlik modeli. Ölçme modeli; hangi hipotetik yapı ya da faktörlerin gözlenen deęişkenlere dayalı ölçüldüğünü ve bu ölçmelerin geçerlik ve güvenirliklerinin ne olduğunu tanımlamaktadır. Yapısal eşitlik modelleri ise, hipotetik yapılar arasındaki doğrudan ve dolaylı ilişkileri tanımlamakta, ayrıca açıklanan ve açıklanmayan varyans miktarını göstermektedir. YEM’de sürekli ya da sıralı, bir ya da birden fazla bağımsız deęişken ile sürekli ya da sıralı, bir ya da birden fazla bağımlı deęişken arasındaki ilişki aynı anda incelenebilmektedir. Diğer yandan hem bağımlı hem de bağımsız deęişkenler faktör ya da gözlenen deęişkenler (ölçme deęerleri) olabilmektedir (Kline, 1998; Schumacker ve Lomax, 2004).

YEM çok sayıda modele verilen genel bir isimdir. Bu modeller arasında tek ve çok deęişkenli regresyon modelleri, doğrulatoryı faktör analizi ve path analizi modelleri sayılabilmektedir.

YEM analizleri genellikle beş adımdan oluşmaktadır:

İlk adımda model tanımlanmaktadır. Araştırmacı ilk adımda başlangıç olarak kuramsal bir model oluşturmaktadır.

İkinci adımda kuramsal modelde tahmin edilen parametreler için tek bir değer bulunup bulunmadığını araştırır,

Üçüncü aşamada, modelde kullanılan değişkenlerin ölçek tipi ve dağılım özelliklerine uygun olarak maksimum olabilirlik, en küçük kareler, ağırlıklandırılmış en küçük kareler gibi bir tahmin yöntemi seçer,

Dördüncü adımda, modelin uyumu yorumlanır ya da uyum katsayılarının alternatif ya da nested (iç-içe) modeller için karşılaştırmasını yapar.

Son adımda ise, modelin uyum katsayıları zayıf ise, araştırmacı modelde hangi parametreleri tahmin etmenin gereksiz olduğunu ya da hangi pathlerin eklenmesi gerektiğini tartışır, bir karara vardıldıktan sonra alternatif model üzerinde analizler tekrarlanır (Kaplan, 2000; Kline, 1998; Schumacker ve Lomax, 2004).

Doğrulayıcı faktör analizi modeli, ölçmeye konu olan ve doğrudan gözlenemeyen gizil yapıların, belli bir sayıda gözlenen değişkenler aracılığı ile söz konusu gizil yapıları temsil eden faktörler ile bağlanımını veren doğrusal bir regresyon modelidir. Bu model tek bir gruba uygulandığında, bütün faktörlerin ortalamalarının, genellik özelliği kaybolmadan, sıfıra eşit olduğu kabul edilir. i bireyler, j gözlenen değişkenler ve l faktörler olmak üzere, i katılımcısının j maddesindeki puanı şöyle ifade edilir:

$$y_{ij} = \sum_{l=1}^L \lambda_{jl} \eta_{il} + \varepsilon_{ij} \quad (1)$$

Burada, y_{ij} i bireyin Y_j değişkenindeki puanı; λ_{jl} Y_j 'nin η_j faktör puanındaki regresyon doğrusunun eğimi (faktör yükü) ve ε_{ij} i katılımcısının Y_j değişkenindeki artık puanıdır. Değişkenlerin kovaryans matrisi, Σ , aşağıdaki gibidir;

$$\Sigma = \Lambda \Psi \Lambda' + \Theta \quad (2)$$

Λ faktör yüklerinin matrisidir, Ψ , faktör puanlarının ortak kovaryans matrisidir ve Θ artık puanların ortak kovaryans matrisidir. Temel olarak, gözlemlenmiş değişkenlerdeki puanlar, faktör yüklemeleri ve artık puanlarla çarpılan faktör puanlarına bölünür. Faktör puanları ve artık puanların birbiriyle ilişkisiz olduğu, ve gözlenen değişkenlerin hata puanlarının da birbiriyle ilişkisiz olduğu varsayılır. Faktör yükleri ve faktör ortak

kovaryans matrisi gözlemlenmiş değişkenlerin ortak içerikte olmasını sağlar. Bu üç Λ, Ψ, Θ matrisin elemanları modelin parametreleridir ve uygulamada gözlemlenmiş Σ , değişkenlerin ortak kovaryans matrisinden hesaplanır. Model parametrelerinin her birinin tek tek tanımlanıp tanımlanmadığının anlaşılmasını sağlayacak yöntemler Bollen (1989) 'da ele alınmıştır.

Çok örneklemlili doğrulayıcı faktör analizi modeli, daha önce açıklanan ortak kovaryanslar modeline ek olarak ortalamaların dahil edildiği bir modelden oluşur. Grupların her biri için, modelin bu iki parçası tanımlanır. Sonra, bu iki parçanın, tüm gruplar için aynı anda tahmini yapılır. j değişkenindeki i katılımcısının gözlemlenmiş puanı (i katılımcısı g grubunun üyesidir, $g=1, \dots, G$), aşağıdaki gibi ifade edilir;

$$y_{ijg} = \nu_{jg} + \sum_{I=1}^L \lambda_{jlg} \eta_{ilg} + \varepsilon_{ijg} \quad (3)$$

Ortalamaların modeli ve g grubunun kovaryansları aşağıdaki gibi gösterilir:

$$\mu_g = \nu_g + \Lambda_g \alpha_g \quad (4)$$

$$\Sigma_g = \Lambda_g \Psi_g \Lambda_g' + \Theta_g \quad (5)$$

Tek grup modelinde Λ, Ψ ve Θ matrislerinin model parametreleri, gözlemlenmiş ortak kovaryans matrislerinin belli bir yapı oluşturması için kullanılır. Bu parametrelere ek olarak, η gözlemlenmiş ortalamalarının bir yapı oluşturması için, ν regresyon sabiti parametreleri (intercepts) ve α_g faktör ortalamaları denkleme dahil edilir. Regresyon sabiti parametrelerinin gruplar arasında eşit olması gerekmektedir (yani; $\nu_g = \nu$). Ayrıca, faktör ortalamaları grupların birinde sıfıra sabitlenir. Burada, genellikle ilk grup için $\eta_1 = \nu$ olması sağlanır. Bu durumda, faktör ortalaması farkları α vektörünün elemanları olarak birinci gruba göre tanımlanmış olur.

Çok gruplu doğrulayıcı faktör analizi ile gruplar arasındaki ortak kovaryans modelinin ve ortalamanın aynı anda hesaplanması; a) tüm gruplar için parametrelerin eşit kalmasını ve b) grupların faktör ortalamalarına göre karşılaştırılmalarını sağlar. Her ikisi de ölçmelerin gruplar arası değişmezliği analizleri konusunda çok önemli imkanlar sağlamaktadır.

Doğrulayıcı faktör modelinde ölçme değişmezliğinin teorik olarak oluşturulması daha çok Meredith (1964) sayesinde olmuştur. Mellenberg (1989) tarafından da, daha genel bir bağlamda şöyle tanımlanmıştır:

$$f(Y | \eta_s) = f(Y | \eta) \quad (6)$$

Denklem 6, verilen bir faktör puanı için gözlenen puanların olası dağılımının, testi alan bireylerin ait oldukları gruplardan bağımsız olduğuna işaret etmektedir. Bir faktörün belli bir tutumu temsil ettiğini kabul edelim. Ölçme değişmezliğinin anlamı, testi alan bireylerin hangi alt grubunun üyesi olduğunun, tutum testlerinin gözlemlenmiş puanları üzerinde fazladan bir etkisi olmadığıdır. Grup ya da sınıflar, sadece tutum faktörlerinin ortalamalarına ve kovaryanslarına göre farklılık gösterebilir. Burada ölçme değişmezliği, çok gruplu doğrulayıcı faktör modeli kullanılarak test edilebilen bir hipotezdir. Meredith, ölçme değişmezliğini analiz için regresyon doğrusunun kesim noktasının, faktör yükleri ve artık (hata) varyanslarının gruplar ya da sınıflar arasında eşit kalacak şekilde sınırlandırıldığı bir faktör analizi modeli önermektedir. Ölçme değişmezliğinin test edilmesinin bir yolu, tamamıyla sınırlandırılmış bir model ile (örn: regresyon sabitinin, faktör yükleri ve artık varyanslarının tüm gruplarda eşit olduğu bir model) daha az sınırlandırılmış bir modelden hangisinin daha uygun olduğunun, olabilirlik oranı (likelihood ratio) testini kullanılarak karşılaştırılmasıdır. Ölçme değişmezliğini test etmek için bahsedilen parametreler gruplar için serbest olarak hesaplandığında, model uyumunda önemli bir artış olması, testin parametrelerinin gruplar arasında değişmez olmadığını gösterir. Bu şekilde ölçme değişmezliğinin test edilmesi, gözlenen test maddelerinin hangilerinin ölçme değişmezliğini ihlal ettiği konusunda yararlı bilgiler sağlayabilir. Diğer bir yol da, sınırlandırılmış modelin uyum iyiliğini elde ettikten sonra, gruplar arasında parametrelerin serbest bırakıldığı modellerin uyum iyiliği ölçmelerinin değerlendirilmesidir.

Yapısal eşitlik modeli ile ilgili bazı terimler:

1. *Ölçme modeli:* Gözlemlenen değişkenler ve gizil değişkenler arasındaki ilişkiler hakkındaki hipotezleri temsil eden modeldir. Ölçme modelinin amacı göstergelerin gizil değişkenleri ne oranda açıklandığının saptanması ve gizil değişkenler arasındaki korelasyonların belirlenmesidir.
2. *Yapısal model:* Yapısal model araştırmacının sınamak istediği modeldir. Ölçme modelinden en önemli farklılığı gizil değişkenler arasındaki ilişkilerin örüntüsünün ve yönünün tanımlanmış olmasıdır.
3. *Bağlantı (Path) diagramı:* Yapısal Eşitlik Modelinde kesinlikle gerekli olmamasına karşın bağlantı diagramı araştırmacıya çıktıları görsel olarak inceleme olanağı sağlar.
4. *Gösterge (observed) değişkenler:* Gözlenen değişkenler YEM dilinde göstergeler olarak ifade edilir ve bunlar araştırmacının doğrudan ölçtüğü ya da gözlediği değişkenleri ifade eder. Bağımlı ve bağımsız gizil değişkenlerin gösterge değişkenleri x ve y değişkenleri olarak tasarlanır. YEM terminolojisinde göstergeler gizil değişkenleri yordamaz, aksine gizil değişkenler kendi göstergelerini yordar.
5. *Gizil (latent) değişkenler:* Gizil değişkenler YEM'in en önemli kavramlarından biridir ve araştırmacıların ilgilendikleri zeka, güdü, duygu, tutum gibi soyut kavramlar ya da psikolojik yapılara karşılık gelir. Bu yapıları ancak dolaylı olarak, belirli davranışlar ya da göstergeler temelinde ölçülen değişkenler yardımıyla gözlemlenebilir.
6. λ_x (*Lamda x*) ve λ_y (*Lamda y*): λ_x ; gizil bağımsız değişkenlerin yordadığı bağımsız gösterge değişkenlere ilişkin katsayılarıdır.
 λ_y , gizil bağımlı değişkenlerin yordadığı bağımlı gösterge değişkenlere ilişkin katsayılarıdır. λ_x ve λ_y katsayıları aynı zamanda geçerlilik katsayılarıdır. Bu katsayılar ölçme modelinde faktör yüklerini verir. (Jöreskog and Sörbom, 1993).

7. *Gama katsayıları (γ)*: Bağımsız gizil değişkenlerin bağımlı gizil değişkenler üzerindeki yordayıcı regresyon ilişkisini gösteren parametredir (Jöreskog and Sörbom, 1993).
8. *Teta- Delta (δ)*: Gizil bağımsız değişkenlerin yordadığı bağımsız gösterge değişkenlerin hata katsayılarıdır (Jöreskog and Sörbom, 1993).
9. *Teta- Epsilon (ϵ)*: Gizil bağımlı değişkenlerin yordadığı bağımlı gösterge değişkenlerin hata katsayılarıdır (Jöreskog and Sörbom, 1993).

Test edilen modelin analiz edilen veriyle uyumunu sına:

Yapısal Eşitlik Modelinde (YEM) uyum iyiliğinin değerlendirilmesi kullanılan paket programa göre değişebilir. Ancak en yaygın olarak kullanılan Ki Kare testidir. LISREL programı modelin veriye uyumunun farklı yönlerini, farklı ölçütler temelinde değerlendiren çok sayıda uyum indeksi sunmaktadır. Bunlar Ki Kare (χ^2) Uyum Testi (Chi-Square Goodness of fit), Uyum iyiliği Testleri (Goodness of Fit) ve Karşılaştırmalı Uyum İndeksleri (Comparative Fit Indices) olmak üzere üç grupta toplanabilir.

Ki Kare Uyum Testi (Chi-Square χ^2) : Ölçülen değişkenler arasında gözlenen kovaryans matrisi ile model tanımlamayla yordanan kovaryans matrisinin ne oranda benzeştiğine karşılık gelir. Uyum mükemmel ise elde edilen değerin 0'a yakın olması ve anlamlılık değerinin (p değeri) manidar (anamlı) olmaması gerekir. Dolayısıyla geleneksel anlamlılık testinin tersine χ^2 testinde anlamsız bir "p" değeri elde edilmek istenir. χ^2 testi örneklem yeterince genişse ve veri çok değişkenli istatistiğin temel sayıltılarını tam olarak karşılıyorsa doğru bir ölçme verir. Serbestlik dereceside (sd) χ^2 testinde önemli bir ölçüttür. Sd' nin büyük olduğu durumlarında da χ^2 anlamlı sonuçlar verme eğilimindedir. Bu nedenle bazı durumlarda sd' nin χ^2 ' e oranı da yeterlik için bir ölçüt olarak kullanılabilir. Kelloway (1996), χ^2 /sd oranının 5' den küçük olmasını iyi uyumun göstergesi olarak yorumlar.

Uyum İyiliği İndeksi (Goodness of Fit Index GFI) : Model-veri uyumunun örneklem genişliğinden bağımsız olarak değerlendirilebilmesi için geliştirilmiştir. GFI, kurulan modelin örneklemdaki varyans-kovaryans matrisini ne oranda yansıttığını gösterir ve

modelin açıkladığı örneklem varyansı olarak da kabul edilir. Bu nedenle regrasyondaki R^2 ye benzer. GFI değerleri 0 ile 1 arasında değişir ve örneklem genişliğine çok duyarlı olduğu için büyük N’lerde daha küçük değerler verir, 0,90 ve üzeri iyi uyum olarak kabul edilir.

Ayarlanmış/düzeltilmiş Uyum İyiliği İndeksi (Adjusted Goodness of Fit Index, AGFI): Örneklem genişliği dikkate alınarak düzeltilmiş olan bir GFI değeridir. N’in büyük olduğu durumlarda AGFI daha temsili bir uyum indeksidir. AGFI değerleride 0 ile 1 arasında değişir. 0,95 ve üzeri mükemmel uyum 0,90 ve üzeride tatminkar düzeyde uyum olarak kabul edilir.

Yaklaşık Hataların Ortalama Karekökü (Root Mean Square Error of Approximation, RMSEA): Örneklemde gözlenen değişkenler arasındaki kovaryansla modelde önerilen parametreler arasındaki farkın, diğer bir deyişle hatanın büyüklüğünü test eden mutlak uyum indeksidir. GFI ve AGFI’ nin tersine “0” a yakın değerler vermesi beklenir. 0,05’e eşit ya da daha küçük olan değerler mükemmel uyumu, 0,08 ve 0,05 arasındaki değerler de model karmaşıklığı dikkate alınarak kabul edilebilir uyumu gösteren değerler olarak değerlendirilir.

Standardize Edilmiş Hataların Ortalama Karelerinin Karekökü (Standartized Root Mean Square Residual, S-RMR): Gözlenen ve üretilen kovaryans matrisleri arasındaki farkların ortalamasının kareköküdür. LISREL ortalama hataların karekökünü standart bir şekilde yorumlar ve S-RMR adını verir. “0 “ ile “1” arasında bir değer verir, 0,05’ e eşit ya da daha küçük olan değerler kabul edilebilir model-veri uyumunu gösterir.

Karşılaştırmalı Uyum İndeksi (Comparative Fit Index, CFI): Bağımsızlık modelinin (gizil değişkenler arasında ilişkinin olmadığını öngören model) ürettiği kovaryans matrisi ile önerilen YEM modelinin ürettiği kovaryans matrisini karşılaştırır ve ikisi arasındaki oranı yansıtan “0” ile “1” arasında bir değer verir. Değerler 1’ e yaklaştıkça modelin daha iyi bir uyum verdiği kabul edilir, 0,90 ve üzerindeki değerler iyi uyum olarak değerlendirilir.

Normlaştırılmamış Uyum İndeksi (Non-normed Fit Index, NNFI): Modelin karmaşıklığını dikkate alarak bir değer verir. Bunu da karşılaştırdığı modellerin

(bağımsızlık ve önerilen modeller) serbestlik derecelerini hesaba katarak yapar. “0 “ ile “1” arasında bir değer verir, 0,95 ve üzeri mükemmel uyuma 0,90 ve 0,94 arası değerler de kabul edilir uyuma karşılık gelir. Bu araştırmada modelleri değerlendirmek için kullanılan uyum iyiliği indeksleri ve diğer indeksler aşağıdaki çizelgede değerlendirme kriterleri ile birlikte verilmektedir.

Uyum İndeksleri	Kriterler
χ^2/sd	<5
GFI	$\geq 0,90$
AGFI	$\geq 0,90$
CFI	$\geq 0,90$
NFI	$\geq 0,90$
NNFI	$\geq 0,90$
RMSEA	$\leq 0,08$ (orta düzey uyum)
RMSEA	$\leq 0,05$ (yüksek uyum)

Kültürler arası karşılaştırmalarda ölçeklerin ölçme eşdeğerliğinin incelenmesinde çok örneklemli doğrulayıcı faktör analizinin yanı sıra Madde Tepki Kuramını (MTK) kapsamında yapılan değişen madde fonksiyonu analizleride önemli bir yer tutmaktadır. Bu araştırmada PIRLS projesi kapsamında Türkiye ve Amerika Birleşik Devletlerinde uygulanan testlere ait maddelerin değişen madde fonksiyonu içerip içermediğini incelemek üzere MTK çerçevesinde kullanılan yöntemlerden parametre karşılaştırma ve olabilirlik oranı testi model karşılaştırma yöntemleri kullanılmıştır. Bu nedenle, raporun bu bölümünde MTK ve değişen madde fonksiyonu hakkında genel kuramsal bilgiler verilmiştir.

1.3 Madde Tepki Kuramı Hakkında Genel Bilgiler

Madde Tepki Kuramı (IRT) veya Örtük Özellikler Kuramı (Hulin diğer kaynaklar.,1983; Lord,1980; Lord ve Novick,1968), madde cevaplarını, temeldeki potansiyel özelliklere matematiksel olarak bağlamaktadır. Bu kuram genel olarak, θ ile bir maddeye verilen cevapların olasılığını ilintilendiren S-şeklindeki lojistik eğri grubunu belirleyen üç parametreyi gösterir. Madde tepki kuramına göre üç parametrelili modelin matematiksel formülü aşağıda verilmektedir.

$$P_i(\theta) = c_i + (1 - c_i) / 1 + \exp[Da_i(\theta - b_i)] \quad (7)$$

Bu denklemde, $P_i(\theta)$, i maddesine doğru cevap verme olasılığıdır; c_i ise madde karakteristik eğrisinin daha düşük asimptotudur, b_i θ ile aynı metrikte ifade edilen madde zorluğudur (veya, tutum veya duyuş ölçmeündeki aşırılık veya popüleritedir); a_i yükselip alçalma noktasında değerlendirilen madde karakteristik eğrisinin eğimiyle ilgilidir; üç-parametrelili bir madde karakteristik eğrisinin eğimi $(1-c_i) (D/4)$, $\theta = b_i$ 'de hesaplandığında a_i olur; D , madde karakteristik eğrilerinin, normal ogive modelleri ile lojistik madde tepki modelleri arasındaki belirli ilişkilerinden yararlanabilmek için 1,702'ye eşit tutulan bir ölçek sabitidir (Hulin ve diğerleri; 1983 s.29-36). Denklem 7'deki c_i sifıra eşit tutulursa (maddeyi tahminle doğru cevaplama olasılığı sifıra eşitlenirse), Denklem 7 sıfır düşük asimptotu olan iki parametrelili bir model oluşturur. Buna ilave olarak tüm maddelerdeki ayırt etme indeksi a_i sabit bir değere eşit kabul edilirse 1-parametrelili modele ulaşılır.

Madde karakteristik eğrileri, puanlanan madde cevabı olan u_i 'nin bağımlı, θ 'ın ise bağımsız değişken olduğu, doğrusal olmayan regresyon eğrileri olarak yorumlanabilir. Madde karakteristik eğrisinin yüksekliği, potansiyel özellik sürekliliği boyunca θ 'ı olan tüm bireyler arasındaki beklenen madde puanı değeridir. Herhangi bir θ , $E(u_i|\theta)$ noktasındaki madde karakteristik eğrisinin yüksekliği, ölçekteki diğer maddelere veya örnekteki θ dağılımlarına bağlı değildir (Lawley,1943). İki alt popülasyondan elde edilen cevaplara dayanan bir maddenin madde karakteristik eğrileri, doğrusal bir dönüşüme kadar sabittir.

θ 'ya bağlı maddelerin anlamlarının yorumları ve bu maddelerin psikometrik özellikleri, bu sabitlik (invariance) özelliğine dayanmaktadır. Hedef dili kullanan farklı bir kültürden gelen bir grup birey, maddelere kendi orijinal dilinde cevap veren popülasyondan daha farklı bir alt popülasyon olarak görülebilir. Diller ve kültürler arası sabit madde parametreleri, madde anlamlarının ve psikometrik özelliklerinin eşitliğine kanıt olarak yorumlanacaktır. Madde karakteristik eğrileri açısından sabit olmayan maddeler madde tepki kuramının ölçme varsayımlarını karşılamaz (veya kötü çevrilmiştir), ve eşit olmadıkları söylenir. Böyle maddeler genellikle eşit olmayan ölçekler yaratacaktır.

Madde cevap eğrileri normal bir ogive'e göre de örneklenebilir:

$$P_i(\theta) = \frac{1}{\sqrt{2}} \int_{-\infty}^{\zeta} e^{-y^2/2} dy \quad (8)$$

Denklem 8, bir $-\infty$ z-puanından, bir $\zeta=a_1(\theta-b)$ z-puanına standartlaştırılmış normal bir eğrinin altındaki alanı ifade eder. Bu şekilde kullanıldığında, yukarıdaki denklemde verilen θ 'daki normal ogive'in yüksekliği, temel özelliği θ miktarında olan bireyler arasında i maddesine verilen doğru veya olumlu cevap olasılığı olarak yorumlanabilir. Normal ogive'de sıfır düşük asymptotu bulunur ve, Denklem 7'nin iki parametrelili lojistik eğri versiyonundan, normal θ sırası içinde önemsiz miktarlarla ($<|.01|$) ayrılır.

Madde tepki kuramları tek boyutluluk (unidimensionality) ve yerel bağımsızlık olmak üzere iki temel varsayım üzerine kurulmaktadır. Daha önce de belirtildiği gibi, bir madde tepki modelinde, bir maddeye doğru yanıt verilmesi olasılığı, madde karakteristik fonksiyonu yoluyla söz konusu yetenek, madde güçlüğü ve madde ayrıncılığı ile ilişkilendirilir. Fakat yerel bağımsızlık, kestirilen madde parametrelerinin sayısına bakılmaksızın varsayılır. Yani, yanıtlayanın yeteneği göz önünde bulundurulduğunda, yanıtlarının başka maddelerle hiçbir ilişkisinin olmaması için, söz konusu yetenek, maddeye verilen yanıtı etkileyen tek faktör olmalıdır. Bu nedenle, eğer örtük özellikte çok küçük aralıklar alınacak ve bu özellikleri ölçen maddeler ilişkilendirilecek olursa (*correlate*), tüm korelasyonlar 0 çıkmalıdır.

Belli bir yanıtlayıcı için (ya da belli bir yetenek düzeyindeki tüm yanıtlayıcılar için) bir dizi maddeye verilen yanıt örüntüsünün birleşik olasılığı, bu yanıtlayıcının maddelere verdiği yanıtlara ilişkin olasılıkların çarpımına eşittir.

$$Prob(u_1, u_2, \dots, u_n | \theta) = P(u_1 | \theta) P(u_2 | \theta) \dots P(u_n | \theta) = \prod_{i=1}^n P(u_i | \theta) \quad (9)$$

Burada u_i , rasgele seçilmiş bir yanıtlayıcının i maddesine verdiği yanıt ve $P(u_i|\theta)$, θ yeteneğine sahip rasgele seçilmiş bir yanıtlayıcının yanıt/tepki olasılığıdır (bkz. Lord ve Novick, 1968).

Yerel bağımsızlık için yeterli şart olan tek boyutluluk, doğru tepki olasılığının sadece bir değişkenin fonksiyonu olduğu anlamına gelir; böylece, ölçmeün başarısını sadece tek bir temel özellik açıklamış olur. Model varsayımlarının sınanması, temel bileşenler çözümlemesinden elde edilen özdeğerlere karşı bileşen sırası (*component rank*) grafiğinin incelenmesiyle mümkün olabilmektedir. Reckase (1979) temel bileşenler analizi ya da benzer faktör analizleri sonucunda ilk faktörün varyansın % 20'sini açıklaması koşulu ile ilk faktörün özdeğerinin ikinci faktörün özdeğerinden birkaç kat daha büyük olması durumunda tek boyutluluk varsayımının karşılandığını belirtmektedir (akt. Smith, 2002). Bu yöntemin dışında doğrulayıcı faktör analizi tekniği kullanılarak da tek boyutluluk varsayımını sınamak mümkün olabilmektedir.

Sadece madde güçlüğü içeren “bir parametrelili lojistik model, 1-PLM”, madde güçlük ve ayırt edicilik parametrelerini kapsayan “iki parametrelili lojistik model, 2-PLM” ve bu iki parametre ile birlikte doğru cevabı tahmin parametrelerini içeren “üç parametrelili lojistik model, 3-PLM” modeller iki kategori (1-0 şeklinde dikotomik) olarak cevaplanan verileri analiz etmede kullanılan madde tepki kuramı modelleridir. Diğer taraftan, iki cevap seçeneğinden daha fazla sayıda seçeneği olan çoklu kategorili (polythomus) maddeler için de bazı modeller geliştirilmiştir. Bu modeller arasında, bir parametrelili lojistik modelin (Rasch modeli) bir uzantısı olan ve Masters (1982) tarafından geliştirilen “Kısmi Puan Modeli (Partial Credit Model)”, Muraki (1992) tarafından geliştirilen “Genelleştirilmiş Kısmi Puan Modeli (Generalized Partial Credit Model)”, iki parametrelili lojistik modelin bir uzantısı olan ve Samejima (1969) tarafından geliştirilen “Ağırlıklandırılmış Tepki Modeli (Graded Response Model)”, Andrich (1978) tarafından geliştirilen “Dereceli Ölçek Modeli (Rating Scale Model)” ve Bock (1972) tarafından geliştirilen “Sınıflamalı Tepki Modeli (Nominal Response Model)” sayılabilir. Modeller arasındaki farklılık madde karakteristik eğrisinin matematiksel tanımlarından kaynaklanmaktadır.

PIRLS projesinin alt testlerinde bulunan okuma parçalarına ait değerlendirme sorularının çoğu kısmi olarak puanlandığı için madde tepki kuramına dayalı olarak yapılacak olan değişen madde fonksiyonu analizleri için kısmi puan modeli kullanılacaktır. Diğer taraftan, parametre eşitleme yöntemi ile değişen madde

fonksiyonu analizleri PARSCALE programı ile ve olabilirlik oranı testi model karşılaştırma yöntemi ile yapılacak olan değişen madde fonksiyonu analizleri MULTILOG programı kullanılarak yapılacağından kısmi puan modeli olarak her iki programda geliştirilmiş puan modelini kullandığından geliştirilmiş kısmi puan modeli aşağıda kısaca açıklanmıştır.

Genelleştirilmiş Kısmi Puan Modeli

Genel Model:

Genelleştirilmiş kısmi puan modeli, k ya da $k-1$ puanlarını elde etme olasılığı üzerinden k puanı elde etme olasılığının iki kategorili bir modele uyduğunu varsayan bir ilke üzerine kurulmaktadır. Söz konusu varsayım aşağıdaki gibi ifade edilebilir:

$$\frac{P_{jk}(\theta)}{P_{j(k-1)}(\theta) + P_{jk}(\theta)} = \frac{1}{1 + \exp[-D_{aj}(\theta - b_{jk})]} \quad (10)$$

Aslında, puanın $k-1$ ya da k olduğu göz önünde tutulduğunda, bir k puanı elde eden her bir olaya da bir 2 parametrelili lojistik model yerleştirilmiştir. $k=1,2,\dots,m$, ve m de tepki sayısı ya da puan kategorileri olduğunda, $\sum P_{jk}(\theta) = 1$ sınırlandırmasıyla, bütün olası puanlar için denklem kümelerinin çözülmesiyle bir k puanı elde etme olasılığı sağlanabilir. Bu m tanımı, m 'nin kategori sayısının bir eksik olduğu ağırlıklandırılmış tepki modeli denklemlerindeki tanımdan farklıdır. Aslında;

$$P_{jk}(\theta) = \frac{\exp[D_{ajk}(\theta - b_{jk})]}{\sum_{u=1}^{m_j} \exp[D_{aju}(\theta - b_{ju})]} \quad (11)$$

Denklem 11 daha genel bir model olan nominal modelin (Bock, 1972) bir ifadesidir. Burada a_{jk} 'nin yerini $T_{jk}a_j$ aldığında nominal model geliştirilmiş kısmi puan modeli halini alır:

$$P_{jk}(\theta) = \frac{\exp[D_{T_{jk}}(\theta - b_{jk})]}{\sum_{u=1}^{m_j} \exp[D_{T_{ju}}(\theta - b_{ju})]}, \quad (12)$$

T_j 'nin doğrusal bir vektör olması gerekir. Bu formülasyonda T_j puanlama fonksiyonu olarak adlandırılır.

1.4 Değişen Madde Fonksiyonu ve Analiz Yöntemleri

Değişen madde fonksiyonu (DMF), bir tek test maddesine aynı yetenek düzeyinde fakat farklı gruplardan gelen iki bireyin doğru cevap verme olasılığının aynı olmaması olarak tanımlanabilir (Adams and Rowe, 1988; Mellenberg, 1989; Hambleton, Swaminathan and Rogers, 1991).

Tanımdan da anlaşılacağı gibi, değişen madde fonksiyonu analizleri her bir test maddesinin aynı evrenden elde edilen farklı alt gruplarda benzer şekilde işleyip işlemediği sorunu üzerine odaklanmaktadır. Diğer bir deyişle, bu analizler eğitimsel ve psikolojik yapıların ölçmelerinin gruplara göre farklı olup olmadığını araştırmak amacıyla yapılmaktadır (Mellenberg, 1983). Eğer bir test çok heterojen bir evren için kullanılacaksa, değişen madde fonksiyonu analizleri madde seçimi sürecinin en önemli parçası durumuna gelir. Çünkü, bu ölçme işleminde test uzmanı için en önemli kriter farklı alt gruplardan gelen ve testi alan bireyler için en adil ve en doğru sonuçları elde etmek olacaktır.

Değişen madde fonksiyonu analizlerinin iki temel amacı vardır. Bunlardan birincisi, test puanlarının değişik gruplardan gelen farklı değişkenlik kaynaklarından etkilenip etkilenmediğinin araştırılması, ikincisi ise bu ilişkisiz kaynakların bazı alt popülasyonlar için adil olmayan bir avantaj sağlayıp sağlamadığıdır (Crocker ve Algina, 1986). Eğer test ile ölçülen özellik ilgilenilen farklı alt gruplar için eşit ise testteki maddelerin farklı alt gruplardaki aynı değişkenlik kaynağından etkilendiğini söyleyebiliriz, ve buna ek olarak, aynı yetenek düzeyindeki cevaplayıcılar aynı test tarafından ölçülen psikolojik yapı açısından benzer düzeyde olacaklardır, ilişkisiz

değişkenlik kaynağının dağılımı bütün alt popülasyonlar için aynı ya da benzer olacaktır (Crocker ve Algina, 1985; s.376-377).

Test maddelerinin fonksiyonel farklılıklarını testin uygulandığı popülasyonun alt gruplarından gelen öğrencilere göre değerlendirmek, test geliştirme uzmanları için madde analizleri ve geçerlik çalışmalarının önemli bir parçasını oluşturmaktadır. Değişen madde fonksiyonu analizleri her bir test maddesinin aynı popülasyondan elde edilen farklı alt gruplara benzer şekilde davranıp davranmadığı sorunu üzerine odaklanmaktadır. Diğer bir deyişle, bu analizler eğitimsel ve psikolojik yapıların ölçmelerinin gruplara göre farklı olup olmadığını araştırmak amacıyla yapılmaktadır (Mellenberg, 1983). Bu açıklamaya göre değişen madde fonksiyonu analizleri kullanılarak testlerin geçerlikleri sorgulanmaktadır. Eğer bir test çok heterojen popülasyonların test edilmesi için kullanılacaksa, değişen madde fonksiyonu analizleri madde seçimi sürecinin en önemli parçası durumuna gelir. Çünkü, bu ölçme işleminde test uzmanı için en önemli kriter farklı alt gruplardan gelen ve testi alan bireyler için en adil ve en doğru sonuçları elde etmek olacaktır. Testlerdeki her bir maddenin aynı popülasyondan gelen farklı alt gruplar için benzer özellikler gösterip göstermediğini değerlendirmek amacıyla test ve madde düzeyinde yapılan değişen madde fonksiyonu analizleri, grupların birlikte test edildiği uygulamaların artmasından dolayı, daha da önem kazanmıştır.

Testlerdeki yanlış maddeleri araştırmak için bir dizi yöntem önerilmektedir. Bunlardan bazıları Klasik Test Teorisi tekniklerinin kapsamına girmektedir. Bu teknikler genellikle, madde istatistiklerinin (madde güçlük indeksi ve madde ayırtıcılık gücü indeksi) alt gruplara göre karşılaştırılması üzerine kuruludur. Klasik test teorisi kapsamında başvuru olan diğer bir teknik ise chi-square yaklaşımıdır. Bu yöntem cevaplayıcı farklarının, ilgilenilen gruplarda bir maddeye doğru cevap verme oranlarının toplam test puanları kategorilerine göre hesaplanmasını içermektedir (Adams ve Rowe, 1988).

Bu teknikler, klasik teorelin getirdiği bazı dezavantajlar göz önünde bulundurularak bir çok araştırmacı tarafından eleştirilmektedir (Shepard, Camilli and Williams, 1984;

Rodney and Drasgow, 1990). Şöyle ki, klasik yöntemlerle hesaplanan madde parametrelerinin bir gruptan diğerine değişkenlik göstermesi yani sabit parametreler olmaması ve örnekleme bağımlı olması nedeniyle yüksek oranlarda hatalı ölçme sonuçları doğuracağı ve bu yüzden, yapılan değişen madde fonksiyonu analizlerinin hatalı yorumlara yol açacağı ileri sürülmektedir. Diğer taraftan, değişen madde fonksiyonu analizi ile bağlantılı olarak Madde Tepki Kuramı (MTK) modellerinin kullanılmasının Klasik Test Teorisi ve tekniklerine göre üstün tarafları vardır. Bu modellerle elde edilen madde karakteristik eğrisi (Item Characteristic Curve) ve madde parametrelerinin sabit olması yani bir gruptan diğerine değişmeyen değerler alması, araştırmacıya maddenin yanlı olup olmadığı konusunda kanıt sağlamaktadır. Bir test maddesinin belirli bir yetenek düzeyinde örneklem farkına rağmen aynı olması yani aynı madde için farklı gruplardan elde edilen madde karakteristik eğrilerinin benzer olması gerekir (Rodney ve Drasgow, 1990; Devine ve Raju, 1982). Bu tanımla uyumlu olarak, madde tepki kuramı modelleri değişen madde fonksiyonu analizleri için gerekli çerçeveyi sağlayabilir, çünkü bu teoride, madde karakteristik eğrileri kapsanan yetenek düzeyinde doğru bir tepkide bulunma olasılığı ile madde özelliklerini ilişkilendirmektedir. Bu anlamda, değişen madde fonksiyonu analizini madde tepki kuramı çerçevesinde yapmak ancak bir test maddesinin iki ayrı gruptan elde edilen madde karakteristik eğrilerinin karşılaştırılması ile olasıdır (örneğin, cinsiyet gruplarına göre veya kültürler arası). Bu karşılaştırmalar çeşitli şekillerde yapılabilir. Madde parametreleri karşılaştırılarak aynı parametrelere sahip madde karakteristik eğrilerinin de aynı olduğu kabul edilebilir veya madde karakteristik eğrileri arasında kalan alan hesaplanarak değişen madde fonksiyonu açısından delil olarak kabul edilebilir (Lord, 1980; Rudner ve diğerleri, 1980). İkinci yöntemde aynı madde için iki ayrı gruptan elde edilen madde karakteristik eğrileri arasında kalan alanın hesaplanması yöntemi ile değişen madde fonksiyonu analizi yapılmaktadır. Hesaplanan alan değeri sıfır olduğunda, o maddenin alt gruplara göre değişen fonksiyon göstermediğine karar verilir (Rudner ve diğerleri., 1980; Raju, 1988).

Diğer taraftan değişen madde fonksiyonu, tek biçimli (uniform) değişen madde fonksiyonu ve çok biçimli (non-uniform) değişen madde fonksiyonu olarak da gözlemlenebilir. Şöyle ki; iki gruba ait madde karakteristik eğrileri temelde b

parametrelerinde farklı olmak üzere aşağı yukarı paralel ise bu çeşit yanlılığa tek biçimli değişen madde fonksiyonu denir. Tek biçimli değişen madde fonksiyonunda maddenin doğru cevaplandırılma olasılığındaki farklılık tüm yetenek düzeylerinde iki grup için değişmemektedir. Bu tür değişen madde fonksiyonunda madde tüm yetenek düzeylerinde aynı grubun lehine işlemektedir. Eğer madde karakteristik eğrileri kesişiyorsa bunun anlamı maddenin doğru cevaplandırılma olasılığının alt yetenek düzeyinde bir grup için yüksek, üst yetenek düzeyinde ise diğer grup için yüksek olduğudur. Olasılıklardaki farklılık yetenek düzeylerinde gruplara göre değişkenlik gösterdiği için (sadece bir grubun lehine devamlılık göstermediği için) bu değişen madde fonksiyonu türü, çok biçimli değişen madde fonksiyonu olarak adlandırılmaktadır. Böyle durumlarda, madde karakteristik eğrilerinin kesişme noktalarının aşağısındaki ve yukarısındaki alanların toplamı sıfır olabilir ve bu yüzden değişen madde fonksiyonu ayırt edilemez. Bu durumda, işaretli ve işaretsiz alan indeksleri birlikte değerlendirilir.

Madde tepki kuramı modelleri değişen madde fonksiyonun incelenmesi için ve daha az yanlı ölçeklerin geliştirilebilmesi için çekici bir yöntem sunmaktadır. Çünkü, belli koşullar altında bu modeller evrene bağımlı olmayan sabit parametreler üretirler. Bu nedenle, kuramsal olarak da parametreler düzensizlik veya madde yaygınlığına ilişkin oran farklılıklarından daha az etkilenirler. Klasik test kuramından farklı olarak, madde tepki kuramı, madde güçlüğünü yetenek düzeyine bağlar; maddeler ve kişiler ortak bir ölçeğe yerleştirilirler.

DMF, doğru tepkinin şartlı olasılığının gruplar arası farklılaşması durumunda meydana gelir. MTK modeline göre, farklı alt gruplardan, ancak aynı yetenek düzeyindeki kişilerin belli bir maddeye olumlu tepki verme ihtimallerinin eşit olmadığı durumda bir madde DMF gösterir. Örneğin, ortalama derecede bunamış, ortalama eğitim düzeyindeki kişilerin yüksek eğitim düzeyinde olup benzer derecede zihinsel geriliği olan kişiler ile belli bir test maddesine benzer derecede doğru cevap vermesini bekleriz. Diğer bir deyişle, maddeye verilecek tepkinin olasılığını söz konusu nitelik ile ilişkilendiren madde karakteristik eğrilerinin tüm ilgi grupları için aynı olması gerekir.

Modelin varsayımları karşılanmışsa, ve iki veya daha fazla eğrinin alanlarında büyük farklar yaratacak derecede eğimler oluşmuşsa, bu DMF için bir delil olarak kabul edilir.

Değişen madde fonksiyonunu (DMF) değerlendirmede madde tepki kuramına dayalı yöntemler:

Camilli ve Shepard (1994) madde tepki kuramı kapsamında değişen madde fonksiyonunun tespit edilmesinde birbiri ile ilişkili ancak birbirinden farklı iki yaklaşımdan bahsetmektedirler. Birinci yaklaşımda değişen madde fonksiyonu belirlenmesi ön planda iken, ikinci yaklaşımda gözlenen değişen madde fonksiyonunun anlamlı olup olmadığı veya ne kadar güvenilir olduğu istatistiki olarak test edilmektedir. Birinci yaklaşım dahilinde hesaplanan değişen madde fonksiyonunun büyüklüğünü ölçmede dört yöntem önerilmektedir. Sırası ile, a-gruplarda elde edilen madde karakteristik eğrileri arasındaki farkı bulmaya yönelik temel alan indeksleri yöntemi, b-olasılık fark indeksleri, c-madde güçlük parametrelerini karşılaştıran yöntem ve d-küçük örneklemelerin madde karakteristik eğrisi yöntemidir (Camilli ve Shepard, 1994: s:46-100).

İkinci yaklaşımda ise değişen madde fonksiyonu farklılığının istatistiki anlamlılık testleri önem kazanmakta ve karşılaştırma yapılan gruplara ait parametreler arasında gözlenen farkların güvenilir bir biçimde ölçüldüğü kabul edilmektedir (Holland ve Wainer, 1993). Bu yaklaşım kapsamında beş yöntemden bahsedilmektedir. Bu yöntemler; madde güçlüğü farkı, madde eğilim modeli (item drift model), Lord'un χ^2 testi, değişen madde fonksiyonu indekslerini belirlemede ampirik örneklem dağılımları (amprical sampling distributions for DIF indices) ve model karşılaştırma ölçmeleri (model comparisons meaasures) yöntemleridir.

Bu çalışma kapsamında değişen madde fonksiyonu analizleri, yukarıda bahsedilen yaklaşımlardaki; a-karşılaştırma yapılan gruplardan elde edilen parametreleri madde düzeyinde analiz eden parametre karşılaştırma yöntemi ve b-yine madde düzeyinde olabilirlik oranı testine dayalı olan model karşılaştırma yöntemleri ile yürütülmüştür.

Testlerin psikolojik yapıları ile ilgili denkliğin ve ölçme eşitliğinin sağlanamaması, bu testlerin puanlarına dayalı olarak yapılan kültürler arası değerlendirmelerin anlamlı olmayacağı sonucunu doğurmaktadır. Çünkü bu tür sorunların olduğu karşılaştırmalarda testlerin kültürler arası uygulanan versiyonlarının geçerliği tartışılır olmaktadır. Sonuçta bu tür testlerin psikometrik yapısı kültürlere göre değişiklik gösterebilmekte ve test maddeleri kültürlere göre değişen madde fonksiyonu içerebilmektedir. Bu nedenle ülkemizin katıldığı ve arzu edilen öğrenci başarılarına ulaşamadığı rapor edilen sınavlarda uygulanan testlerin yapısının ve psikometrik özelliklerinin kültürler arası karşılaştırmalı olarak incelenmesi gerekmektedir.

Bu çalışma kapsamında, ülkemizin de 2001 yılında katıldığı PIRLS projesine ait testlerin psikometrik özellikleri kaynak kültür (Amerika Birleşik Devletleri) ve hedef kültür (Türkiye) kapsamında karşılaştırmalı olarak incelenmiştir.

1.5 Problem Cümlesi

PIRLS testlerinin yapıları ve psikometrik özellikleri kültüre bağlı olarak değişim göstermekte midir?

Alt Problemler

1-PIRLS Projesi kapsamında Türkiye ve Amerika Birleşik Devletlerinde uygulanan ;

- a) Flowers ve Antarctica okuma parçalarından oluşan alt testin faktör yapıları kültürler arası değişim göstermekte midir?
- b) Mice ve River okuma parçalarından oluşan alt testin faktör yapıları kültürler arası değişim göstermekte midir?
- c) Clay ve Leonardo okuma parçalarından oluşan alt testin faktör yapıları kültürler arası değişim göstermekte midir?

d) Hare ve Pufflings okuma parçalarından oluşan alt testin faktör yapıları kültürler arası değişim göstermekte midir?

2-PIRLS projesi kapsamında uygulanan alt testleri oluşturan 8 okuma parçasına ait değerlendirme testlerindeki maddeler kültürler arasında;

- a) Parametre Karşılaştırma Yöntemine göre değişen madde fonksiyonu göstermekte midir?
- b) Olabilirlik Oranı Testi Model Karşılaştırma Yöntemine göre değişen madde fonksiyonu göstermekte midir?

1.6 Sınırlılıklar

Bu çalışma, Uluslararası Eğitim Başarılarını Belirleme Kuruluşu (IEA)'nın yapmış olduğu PIRLS-2001 testinin uygulandığı örnekleme dahil edilen Türkiye ve Amerika Birleşik Devletleri İlköğretim Okulları 4. sınıf öğrencileri ile sınırlıdır.

1.7 Kısaltmalar

IEA	: Uluslararası Eğitim Başarılarını Belirleme Kuruluşu
EARGED	: Eğitimi Araştırma Geliştirme Dairesi
TIMMS	: Üçüncü Uluslararası Fen Bilgisi ve Matematik Çalışması
PISA	: Uluslararası Öğrenci Başarısını Belirleme Programı
PIRLS	: Uluslararası Okuma Becerilerinde Gelişim Projesi
YEM	: Yapısal Eşitlik Modelleri
MTK	: Madde-Tepki Kuramı
DMF	: Değişen madde fonksiyonu
ABD	: Amerika Birleşik Devletleri

1.8 Araştırmanın Amacı ve Önemi

Araştırmanın amacı öncelikle Okuma Becerilerinde Gelişim Projesi PIRLS-2001 testinin farklı kültürlerde aynı psikolojik yapıyı ortaya çıkarıp çıkarmadığını tespit etmektir. Diğer bir amaç ise benzer psikolojik yapılar ölçülüyor olsa bile ölçekteki maddelerin farklı alt popülasyonlarda benzer şekilde fonksiyon gösterip göstermediğinin belirlenmesidir.

Eğer, PIRLS-2001 testinin ölçtüğü psikolojik yapılar kültürler göre eş değer değil ise ve test maddeleri kültüre göre yanlılık gösteriyorsa testin geçerliği tartışılabilir olacaktır. Bu durumda, bu testlerden elde edilen sonuçlar üzerinden yapılan kültürler arası değerlendirmeler anlamlı olmayacaktır (Hambleton, 1994).

1.9 İlgili Araştırmalar

Hulin (1987), Mesleki Memnuniyet Ölçeğinin dört ayrı dile (İspanyolca, Tagalogca, İbranice ve Kanada Fransızcası) çevrilen versiyonlarını çeviri kuramları ve değerlendirme metodları açısından incelemiştir. Çalışanlar ölçeklere ya kaynak dilde (İngilizce) ya da hedef dilde cevap vermişlerdir. Araştırmacı, kültürler arası ve diller arası karşılaştırmada, çevirilerin diller arası eşitliğinin hesaplanmasına olanak verdiği için madde karakteristik eğrilerini kullandığını belirtmektedir. Sonuçlar her çeviride ve karşılaştırmada yeterli sayıda eşit olmayan madde bulunduğunu göstermiştir.

Sireci ve Berberoğlu (2000), Kurs Değerlendirme Ölçeğini (Course Evaluation Form, CEF) kullanarak Orta Doğu Teknik Üniversitesindeki kursları ve öğretim elemanlarını değerlendirmişlerdir. Ölçek 4'lü Likert ölçeği türünde ve 53 maddeden oluşmaktadır. Cevaplayıcı öğrencilerin iki dili profesyonelce kullandığı (Öğretim dili İngilizce ve ana dili Türkçe) varsayılmıştır. Orijinal ölçek İngilizce olup iki bağımsız çevirmen tarafından Türkçe'ye çevrilmiştir. İki ayrı dilde oluşturulan form dönüşümlü olarak öğrencilere uygulanmıştır. Araştırma sonucunda yanlılık gösteren maddeler ortaya çıkmıştır.

Ellis (1989), 205 Alman yüksek okul öğrencisine ve 217 Amerikalı birinci ve ikinci sınıf kolej öğrencilerine Career Ability Placement Survey (CAPS) ve Wilde-Intelligenz-Test (WIT) testlerini uygulamıştır. Testler öncelikle diğer dillere çevrilmiştir. CAPS testi İngilizce'den Almanca'ya ve WIT testi Almanca'dan İngilizce'ye çevrilmiştir. WIT testinin orijinal versiyonu ve çevirisi yapılan CAPS testi Alman öğrencilere, orijinal CAPS testi ve çevirisi yapılan WIT testi Amerikalı öğrencilere uygulanmıştır. Her iki ülke uygulamasında da testler öğrencilere dönüşümlü sıralarla uygulanmıştır. Araştırma sonuçları CAPS testinde 8 maddenin, WIT testinde ise 2 maddenin yanlılık gösterdiğini ortaya koymuştur. Araştırmacı çeviri adaptasyon ve geri çeviri adaptasyon işlemleri ne kadar dikkatli ve sancılı yapılırsa yapılsın kültürler arası karşılaştırmalarda ölçmenin eşitliğinin garantilenemediğini belirtmektedir.

Reise, Wideman ve Pugh 1993 yılında yaptıkları bir çalışmada, çok örneklemli doğrulayıcı faktör analizi ile madde tepki kuramı modelleri (olabilirlik oranı testini kullanarak) kapsamında ölçme eşdeğerliğini belirlemede her iki yöntemin benzer ve farklı yönlerini incelemişlerdir. Araştırmada beş dereceli NA5 “negatif duygu” ölçeği Amerikalı (n=540) ve Çinli (n=598) üniversite öğrencilerine uygulanarak veriler elde edilmiştir. Yapılan karşılaştırmalar, hem doğrulayıcı faktör analizi hem de olabilirlik oranı testi sonucunda iki maddenin gruplar arasında ölçme eşdeğerliğinin olduğunu gösterirken bir maddenin ise ölçme eşdeğerliğinin sağlanamadığını göstermiştir. Diğer taraftan olabilirlik oranı testi diğer kalan iki maddenin de gruplar arasında madde işlev farklılığını ortaya çıkarırken, doğrulayıcı faktör analizi bu iki madde hakkında madde işlev farklılığı olduğuna dair bulgular sağlayamamıştır. Yazarlar bu durumun, doğrulayıcı faktör analizinin daha çok doğrusal ilişkiler içinde sınıflanması buna karşın madde tepki kuramı modellerinin ise doğrusal olmayan lojistik ilişkileri içermesi nedeniyle ortaya çıktığını savunmaktadırlar. Doğrulayıcı faktör analizi ile elde edilen regresyon katsayıları madde tepki kuramı modellerindeki madde ayırt edicilik parametresine karşılık gelse de tam olarak aynı olmadığı ve hem madde güçlük hem de madde ayırt etme parametre değerlerinin bilinmesinin psikolojik yapılar bakımından son derece gerekli olduğu belirtilmektedir. Araştırmacılar MTK modellerinin de DFA' da olduğu gibi daha fazla ve pratik olarak yorumlanabilir uyum indekslerine sahip olmaları durumunda daha verimli bir şekilde değerlendirilebileceğini belirtmektedirler.

Değişen madde fonksiyonunun incelendiği bir diğer araştırma Huang, Church ve Katigbak tarafından 1997 yılında yapılmıştır. Araştırmada NEO kişilik envanterinin İngilizce versiyonunun kültürler arası ölçme değişmezliği analizi Filipinli (n=432) ve Amerikalı (n=610) üniversite öğrencisi ile yapılan uygulamanın verileri üzerinden yürütülmüştür. Araştırmada madde tepki kuramı modellerine uygun olarak parametre eşitleme yöntemi ile model karşılaştırma yöntemi ve klasik test teorisine dayalı yöntem olarak da Mantel-Haenszel yöntemi kullanılmıştır. Uygulması yapılan ölçek 180 madde ve 5'li Likert tipinde derecelmeli bir ölçek olduğu halde araştırmacılar maddelere verilen cevapları iki kategoriye (1-0) dönüştürmüşlerdir. Analizler sonucunda parametre eşitleme yönteminde 30, model karşılaştırma yönteminde 62 ve MH yöntemi ile yapılan analiz sonucunda ise 79 maddenin değişen madde fonksiyonu gösterdiği ortaya çıkmıştır ($p<0.001$). Model karşılaştırma yönteminde tekrarlayıcı madde ayıklama işlemi uygulanırken, diğer yöntemler için bu işleme başvurulmamıştır. Araştırmacılar parametre eşitleme yöntemi ile daha az sayıda yanlış madde bulunmasının örneklem büyüklüğüne bağlı bir sonuç olabileceğini öne sürmektedirler. Çünkü parametre eşitleme yönteminde ayrı ayrı parametre tahminleri yapılırken model karşılaştırma ve MH yönteminde odak ve referans grupların madde parametre tahminleri aynı anda yapıldığı için toplam örneklem sayısı parametre eşitleme yöntemine göre daha fazla olmaktadır. Araştırmada ayrıca, ölçeğin orijinal Likert tipi puanlamasından elde edilen ham puanlar, değişen madde fonksiyonu gösteren maddeler belirlenmeden önceki tüm maddeler üzerinden hesaplanan θ kişilik özelliği tahminleri ve değişen madde fonksiyonu saptanan maddelerin çıkarılmasından sonra hesaplanan θ kişilik özelliği tahminleri arasında kültürlere ve cinsiyet değişkenlerine göre MANOVA analizi yapılmıştır. Bu analizler sonucunda ise özellikle ölçek puanlarının orijinal Likert formatı ile iki kategorili hale getirilen formatı arasında bahsedilen değişkenler açısından bir farklılık gözlenmemiştir.

Teresi, Kleinman ve Welikson (2000) tarafından yapılan bir araştırmada, bilişsel tarama testlerinin ve test maddelerinin, eğitim düzeyi, ırk ve etnik köken açısından farklılık gösteren gruplarda farklı sonuçlar doğurduğu ortaya konmuştur. Bu çalışma kapsamında, Mattis Bunama Ölçeği'nin dikkat boyutuna ilişkin geliştirilmiş olan ölçek

incelenmiştir. İki ve üç parametrelili lojistik tepki modeli ve politomus tepki modeli de dahil, pek çok madde tepki kuramı modelleri, bu çalışma kapsamında karşılaştırılmıştır. (Log olabilirlik oranı testleri, üç parametrelili modelin iki parametrelili modelden daha gelişmiş bir model olmadığını göstermiştir). Test maddelerinin pek çoğu farklı eğitim gruplarında benzer sonuçlar doğurmuş; madde karakteristik eğrileri neredeyse özdeş bulunmuştur. Ancak dört maddede sorun yaşanmıştır. Bir madde doğrulayıcı faktör analizinde DFA büyük bir hata terimi göstermiştir; BIMAIN kullanılarak geliştirilen madde uyumlu ki kare istatistikleri IRT modellerinin sonuçlarını doğrulamaktadır. Genelde, en üst eğitim düzeyindeki bireylerin, θ 'nın çoğu düzeyi için, maddeyi doğru cevaplandırmada daha büyük bir olasılığa sahip olduğu gözlenmiştir. MULTILOG kullanılarak çözümlenen, DMF'in model dayanaklı testleri, düşük de olsa, anlamlı DMF'e sahip üç madde belirlenmiştir. Ortalama olarak tüm puan aralığında, düşük eğitim düzeyine sahip grubun maddeyi doğru yanıtlama olasılığı, yüksek eğitim düzeyine sahip olan gruptan 0.11 yüksek çıkmıştır. Daha büyük alt gruplar ile yapılan çapraz geçerlik çalışması bu ölçme içim küçük DMF'in genel sonuçlarını onaylamıştır. Değişen madde fonksiyonunu ortaya çıkarabilmek için kullanılan yöntemler (ki bu yöntemler sırayla yanlılığın göstergesi olabilirler) nöro-psikolojik bir yan teste uygulanmıştır. Bu yöntemler daha önce de eğitim ve etnik köken ve ırk tarama ölçmelerinde alt gruplarında yanlılığı incelemek üzere kullanıldılar. Yazarlar tanı koymada kullanılan tarama ölçmelerinin ve nöro-psikolojik testlerin yanlılıktan arındırılması ve bu sayede kültürel anlamda daha adil sınıflamaların yapılabilmesinin sağlanması gibi önemli epidemiolojik uygulamalara ek olarak, bu yöntemler çoklu bölgelerde gerçekleştirilen klinik deneylerinde bölgeler arası farklılıkların incelenebilmesi açısından yararlı olacağını belirtmektedirler.

Ülkemizde yapılan bir araştırmada Somer (2004), parametre karşılaştırma yöntemi ile değişen madde fonksiyonu analizlerini PARSCALE programı kullanarak gerçekleştirmiştir. Araştırmacı 1807 kişilik öğrenci örnekleminde, bir kişilik ölçeğinin Yumuşak Başlılık alt ölçeğinden elde edilen verileri ele almış ve iki kategorili (1-0) olarak puanlanan 16 ölçek maddesini cinsiyet grupları açısından karşılaştırmıştır. Analizlerde, 2 parametrelili lojistik modele göre madde parametre tahminleri yapılmış, değişen madde fonksiyonu analizleri için madde ayırt etme ve yerleşim parametreleri

kız ve erkek gruplarının her ikisinde de serbest bırakılmış ve ayırt etme parametresine göre gruplar arasında değişen madde fonksiyonu gösteren madde bulunmamıştır. Diğer taraftan, madde güçlük/yerleşim parametresi açısından karşılaştırma grupları arasında 5 maddenin değişen madde fonksiyonu gösterdiği ortaya çıkmıştır. Araştırmacı değişen madde fonksiyonu gösteren bu beş maddeyi ölçme aracından çıkararak ölçekte kalan 11 madde üzerinden analizleri tekrarlamış ve bu tekrarlayıcı işlemleri gruplar arasında değişen madde fonksiyonu göstermeyen maddeler elde edilene kadar sürdürmüştür. Değişen madde fonksiyonu gösteren maddeler çıkarılıp ölçme eşdeğerliği sağlanan toplam puanlar üzerinden yapılan t testi sonucuna göre kızlar ve erkeklerin puan ortalamaları arasında anlamlı bir fark çıkmamıştır. Diğer taraftan ölçeğin bütün maddelerinden elde edilen puanların aritmetik ortalamaları karşılaştırıldığında ise kız öğrencilerin puan ortalamalarının erkeklerin puan ortalamalarından anlamlı bir şekilde farklı olduğu gözlenmiştir. Bu nedenle yazar ölçek eşdeğerliğinden emin olunmadan yapılacak grup karşılaştırmalarında, elde edilen farklılıkların gruplar arasındaki gerçek farklılıklardan mı yoksa ölçme yanlılığından mı kaynaklandığını bilmenin mümkün olduğunu ve bu farklılıklarla ilgili olarak yapılan yorumlarda yanlış kanılara ulaşmanın mümkün olabileceğini belirtmektedir.

Kazelskis, Thames ve Reeves (2004) tarafından yapılan bir araştırmada, McKenna ve Kear tarafından geliştirilen İlköğretim Okuma Tutum Araştırması (The Elementary Reading Attitude Survey "ERAS") ölçeğinin yapı geçerliği çok örneklemli doğrulayıcı faktör analizi ile çalışılmıştır. Ölçekte 10 tanesi okul görevleri dışında zevk için okumaya karşı tutum ve diğer 10 tanesi akademik okumaya karşı tutum olmak üzere 20 madde bulunmaktadır. Uygulama ilköğretim 4. ve 6. sınıf öğrencilerinden oluşan 371 tanesi erkek 341 tanesi kız olmak üzere toplam 728 birimlik bir öğrenci grubu üzerinde yürütülmüştür. Örneklemdeki 353 öğrenci Avrupa kökenli Amerikalı ve 375 tanesi ise Afrika kökenli Amerikalı öğrencilerdir. Çok örneklemli doğrulayıcı faktör analizi çalışmaları cinsiyet ve etnik farklılıklara göre yürütülmüştür. Model veri uyumu ve faktör yapılarının değişmezliği için khi-kare, khi-kare/sd oranı, GFI, AGFI, RMSEA indekslerinin alternatif modeller arasında değerlendirilmesi yöntemi ile yapılmıştır. Analiz sonuçlarına göre okuma tutum ölçeğinin cinsiyet gruplarına göre yapılan doğrulayıcı faktör analizine ait bulgular ölçeğin faktör kovaryanslarının gruplar arasında

ölçme değişmezliğine sahip olduğunu göstermiştir. Diğer taraftan öğrencilerin aile orijini gruplarına göre yapılan karşılaştırmalarda ise okul görevleri dışında zevk için okuma boyutundaki maddelerin faktör yüklerinin anlamlı bir şekilde farklı olduğunu göstermiştir. Akademik okuma boyutunda ise bir farklılık gözlenmemiştir. Bu bulgular ölçeğin zevk için okuma tutumu boyutuna ait yapısının kavramlaştırılması konusunda sorunlar olduğunu ima etmektedir. Araştırmacılar etnik farklılıklarla ilgili bu yapı geçerliği konusunun yeni araştırmalar yapılarak yeniden değerlendirilmesinin faydalı olabileceğini belirtmişlerdir.

Bertrant ve Boiteau (2003) 1995 ve 1999 yılında yapılan TIMSS uygulamalarında ortak olan 48 adet matematik sorusunun Amerika ve Japonya örneklemelerinde değişen madde fonksiyonu gösterip göstermediğini incelemişlerdir. DMF analiz tekniği olarak Mantel-Haenszel, Lojistik regresyon, MTK temelli alan indeksleri tekniği ve Raju'nun NCDIF indeksi olmak üzere dört ayrı teknik kullanmışlardır. Analiz sonuçları hem 1995 hem de 1999 uygulamalarında dört farklı DMF tekniğine göre 48 adet sorunun en az %80'inin kültürlerarası uygulamada DMF içerdiğini göstermiştir. Araştırmacılar ayrıca, 1995 ve 1999 yıllarındaki TIMSS uygulamalarında DMF içeren maddeleri tesbit etmede Mantel-Haenszel yöntemin diğer yöntemlere göre daha tutarlı sonuçlar verdiğini belirtmişlerdir.

Salzberger, Sinkovics ve Schlegelmilch (2000) ölçmenin denkliliğini test etmede başvurulacak uygun istatistiksel yöntem ve tekniklerden çok gruplu doğrulayıcı faktör analizi ve örtük özellikler teorisine dayalı olarak 1 parametrelili (Rasch) politomik tepki modelini bir simulasyon çalışmasına dayalı olarak karşılaştırarak bu yöntemleri karşılaştırmışlardır. On maddelik bir tutum ölçeğinin 400 kişilik veri seti için 1-5 maddeler gruplar arasında sınırlandırılarak (yani gruplar için regresyon katsayıları eşit alınarak) 6-10. maddeleri ise serbest bırakılarak analiz edilmiştir. İkinci setteki maddelere kültürler arası yanlılık gösterme açısından hem çok örneklemli doğrulayıcı faktör analizi hem de Rasch modeli (çok kategorili puanlanan maddeler için) uygulanmıştır. Bulgular DFA modellerini kurarken eşitlenen maddelerin ve sınırlamaların körü körüne yapılmasının hatalı sonuçlar doğuracağını göstermiştir. Ayrıca örtük özellikler kuramına dayalı modellerin daha sağlam sonuçlar ürettiği belirtilmektedir. Özellikle Rasch modelinin madde ayırt edicilik parametresini değişmez

bir deęer olarak alması nedeniyle madde karakteristik eęrilerinin kltre zg cevap eęilimlerinden arınık olarak elde edildięi ve sadece konum parametresi ile alıřmanın kiři ve madde parametrelerinin alıřılan rneklemden baęımsız olması kořulunu saęladıęı ifade edilmektedir. Yazarlar, kltrler arası lme deęiřmezlięi iin yapılan analizlerde yukarıda verilen gerekelere dayalı olarak klasik test teorisine dayalı model analizlerinin yeterli olmadıęını, rtk zellikler teorine dayalı modeller ile de analizlerin mutlaka tekrarlanması gerektięini vurgulayarak bu alanda alıřan akademisyen ve pratisyenlere aęırıda bulunmuřlardır.

Childs ve Chen (1999) yılında yaptıkları bir alıřmada, Maine eyaletinde 11. sınıflara uygulanan okuma ve matematik testlerinin ok kategorili olarak puanlanan maddelerine ait verileri aęırlıklandırılmıř tepki modeli ve kısmi puan modelini kullanarak MULTILOG ve PARSCALE programlarında analiz etmiřlerdir. Bu iki programdan elde edilen madde parametrelerinin kıyaslanabilirlięini arařtırmıřlardır. Bařlangıta aęırlıklandırılmıř tepki modeli ve kısmi puan modeli iin her iki programdan elde edilen madde parametre deęerlerinin kıyaslanabilir olmadıęı grlmřtır. Arařtırmacılar bunun nedenini, MULTILOG ve PARSCALE analiz programlarının, bu modelleri farklı řekilde parametrik olarak ifade etmeleri olduęunu belirtmiřlerdir. Bu nedenle, her iki model iin iki ayrı programdan elde edilen parametre tahminlerini kıyaslanabilir hale getirmek amacıyla bir yntem ortaya koymuřlardır. nce aęırlıklandırılmıř tepki modeline gre maddeler hem Parscale hem de Multilog programında analiz edilmiřtir. Multilog programdan elde edilen madde parametrelerini Parscale programından elde edilen parametrelerle kıyaslayabilmek iin IRTGRAPH programının Childs ve Scumpher tarafından geliřtirilen zel bir versiyonu kullanılarak gerekli dnřtrmeleri yapmıřlardır. Bu dnřtrme yapıldıktan sonra madde parametre tahminlerinin her iki programdan elde edilen deęerleri arasında a parametresi iin 0,01 ile 0,03 arasında deęiřen farklar gzlenirken, b parametresi iin 0,00 ile 0,11 arasında deęiřen deęer farkları bulunmuřtur. Dnřtrme iřleminden elde edilen madde parametre deęerlerinin benzerlięinin olduka yksek olduęu belirtilmiřtir. Gerekli parametre dnřtrmeleri yapıldıktan sonra aęırlıklandırılmıř tepki modelinde olduęu gibi, genellenmiř kısmi puan modeli iinde Multilog ve Parscale madde parametre deęerleri her iki test iinde benzer ıkmıřtır. a ve c parametreleri iin her iki programdan elde edilen parametre

değerleri arasındaki ilişki 0,99 dan daha büyük çıkmıştır. Madde parametre değerleri arasındaki farklar beklendiği gibi çok küçük çıkmıştır (a parametreleri için farklar 0,00 ile 0,06 ve c parametreleri için ise 0,00 ile 0,08 arasında değişmiştir). Araştırma raporunun ekinde farklı modellere (GRM ve GPCM) göre oluşturulan komut dosyaları (Multilog ve Parscale programları için) verilmiştir.

Korkmaz (2005), Beş Faktör Kişilik Envanterinin Uzlaşma alt ölçeğinde 500, 1000 ve 1500 kişilik örneklem gruplarında cinsiyete açısından ağırlıklandırılmış cevap modelini kullanarak üç işlev farklılık yöntemine göre (olabilirlik Oranı Testi Model Karşılaştırma Yöntemi “LR”, Madde Test İşlev Farklılık (DFIT) Yöntemi ve Parametre Karşılaştırma Yöntemi) işlevsel madde farklılığı olup olmadığını belirlemek amacıyla bir çalışma yapmıştır. Analizler sonucunda, $n=500$ örneklem genişliğinde olabilirlik oranı testi ve madde parametre karşılaştırma yöntemlerinin işlev farklılığını saptamada güçsüz kaldığı, örneklem sayısının artmasıyla her iki yöntemin benzer sonuçlar verdiği ve hem 0,001 hem de 0,05 anlamlılık düzeylerinde gruplar arasında işlev farklılığı gösteren maddelere karşı bu yöntemlerin hassasiyetlerinin arttığı gözlenmiştir. Buna karşın DFIT yöntemi tüm örneklem genişliklerinde daha tutarlı sonuçlar vermiştir ve test düzeyinde gruplar arasında işlev farklılığı saptanmamıştır. Araştırmada ölçeğin 6’ncı ve 7’inci maddelerinin en belirgin madde işlev farklılığı gösterdiği ve uzlaşma özelliği açısından değerlendirildiğinde; 6’ncı maddede kadınların daha örtülü bir antagonistik kişilik özelliği varken, 7’inci maddede ise erkeklerin açık ve saldırgan bir ifadeyi daha kolay benimseyebildikleri görülmüştür. Araştırmacı bulguların, psikolojik yapıların içerik ile ilgili incelemelerinde madde işlev farklılığı analizlerinin önemli bir katkısı olduğunu gösterdiğini belirtmektedir.

Alkan (2005), PIRLS kapsamında uygulanan anket verilerinin okuma becerileri başarısını yordama gücünü incelemiştir. Araştırmacı, anket sorularına verilen cevaplardan evde yapılan okuma etkinlikleri indeksi, okulda yapılan okuma sonrası etkinlikler indeksi, öğrencinin okumaya karşı tutum indeksi, öğrencilerin okuma konusunda kendilerine duydukları güven indeksi ve evde bulunan eğitim kaynakları indekslerini oluşturarak bu indekslerin bölgelere ve cinsiyete göre dağılımını analiz etmiş ve bu değişkenlerin başarıyı yordamada ne kadar etkili olduğunu araştırmıştır.

Yapılan çoklu regresyon analizine göre başarıyı en iyi yordayıcı değişkenin öğrencinin okumada kendine duyduğu güven olduğu ortaya çıkmıştır. Diğer taraftan evde yapılan etkinliklerde veli öğrenci etkileşiminin istenilen düzeyde olamadığı görülmüştür. Gerek velilerin gerekse öğrencilerin okumaya ayırdıkları sürenin oldukça az olduğu, evde bulunan toplam kitap sayısı ve çocuk kitapları sayısının oldukça az olduğu anlaşılmıştır. Kız öğrencilerin daha başarılı oldukları gözlenmiştir. Teknoloji ve eğitim materyalleri açısından Doğu ve Güney Doğu Anadolu bölgelerindeki okulların çok yetersiz kaldığı ortaya çıkmıştır.



BÖLÜM II

YÖNTEM

Bu bölümde araştırmanın türü, araştırmanın yapıldığı gruplar, verilerin elde edilmesi verilerin analizinde kullanılan yöntemler ve analizlerde kullanılan istatistik paket programları açıklanmıştır.

2.1. Araştırmanın Türü

Bu araştırmada, Uluslararası Okuma Becerilerinde Gelişim Projesi PIRLS-2001 testinin ABD ve Türkiye'deki uygulamaları analiz edilerek, testin farklı kültürlerdeki psikometrik özellikleri ortaya konmuş, daha sonra kaynak ve hedef kültürlerdeki yapıların eşdeğer olup olmadığı incelenmiştir. Son aşamada ise testlerdeki maddelerin kültürle göre değişen madde fonksiyonu gösterip göstermediği araştırılmıştır.

Bu yönleri ile ele alındığında PIRLS-2001 testinin geçerliği ve özellikle kültürler arası geçerliği saptanmaya çalışılacaktır.

Yukarıda açıklanan amaçlar göz önüne alındığında, bu araştırmanın betimsel araştırma niteliğinde olduğu görülmektedir.

2.2 Araştırmanın Yapıldığı Grup

PIRLS-2001 uygulamasında ülkelere ait örneklemeler Canada Statistics tarafından yansız örnekleme yöntemi ile belirlenmiş, şehir merkezleri ve kırsal kesimlerdeki farklı okul tipleri örnekleme tabakalandırılarak dahil edilmiştir. Türkiye örnekleminde 5390 öğrenci hem öğrenci anketlerini, hem de 10 ayrı formdan oluşan başarı testlerini cevaplandırmışlardır. Sonuçta 5125 Türk öğrenciye ait veriler değerlendirmeye alınmıştır. Amerika Birleşik Devletleri örneklemine ise 4091 öğrenci alınmış, bunlara da aynı yöntemle anketler ve başarı testleri uygulanmıştır. Sonuçta 3763 Amerika Birleşik Devletleri öğrencisinin verileri değerlendirmeye alınmıştır.

PIRLS testlerinin temelini 8 adet okuma parçası ve bu parçalara ait sorular oluşturmaktadır. Bu sekiz okuma parçası ve her parçanın bitiminde bulunan değerlendirme soruları 10 adet test kitapçığına dağıtılmıştır. Uygulama kapsamına giren her bir öğrenci sadece bir alt testin kapsadığı 2 okuma parçasına ait değerlendirme sorularını cevaplamıştır. Her bir okuma parçası için 40 dakika olmak üzere bir alt test için verilen süre toplam 80 dakika olmuştur. Bu tez çalışması kapsamında üç, altı, sekiz ve sıfır kodlu alt testlere ait veriler kullanılmıştır. Söz konusu alt testlerin seçilmesinin nedeni PIRLS 2001 testindeki bahsedilen 8 okuma parçasının tamamının ikiye bölünebilir olarak bu alt testlerde kapsanıyor olmasıdır.

2.3. Verilerin Elde Edilmesi

Araştırma, Uluslararası Eğitim Başarılarını Belirleme Kuruluşu (IEA)'nın PIRLS-2001'e katılan Türkiye ve Amerika Birleşik Devletleri öğrencilerine uygulanan okuma başarı testlerinden elde edilen veriler üzerinde yürütülmüştür. Çalışma için gerekli olan veri [ISC (International Study Center), 2003]'ün web sitesinden elde edilmiştir.

2.4 Verilerin Çözümlemesi

Araştırmanın birinci alt probleminin a, b, c ve d maddelerinde belirtilen problemlerin çözümü için PIRLS projesi kapsamında Türkiye ve Amerika Birleşik Devletlerinde uygulanan üç, altı, sekiz ve sıfır kodlu alt testlerin faktör yapılarının kültürler arası eş değeri yapısal eşitlik modelleri yöntemlerinden çok örneklemli doğrulayıcı (confirmatory) faktör analizi yöntemi ile incelenmiştir.

Analizlere geçilmeden önce verilerin normal dağılıp dağılmadığı, tek değişkenli-çok değişkenli aykırı gözlem (outlier) değer içerip içermediği, maddelerin her birinin birleşimlerinin doğrusal olup olmadığı ve artıkların (residual) normal dağılıp dağılmadığı incelenmiştir (Tabachnick ve Fidell, 2001:115-122).

Bu amaçla, veri setindeki maddelerin her birinde tek değişkenli ve birleşimlerinde çok değişkenli, aykırı gözlem ve maddeler arası çoklu bağlantılılık olup olmadığı incelenmiştir (Stevens, 2001; Tabachnick ve Fidell, 2000). Bireylerin üç, altı, sekiz ve

sıfır kodlu testlerdeki toplam test puanlarına ait z puanlarının sırasıyla -2,16 ile 1,93; -2,40 ile 2,09; -2,29 ile 1,79 ve -1,87 ile 2,27 aralığında değiştiği gözlenmiştir. $z = \pm 3,20$ aralığını aşan aykırı standart test puanına rastlanmamıştır. Maddelerin varyans şişme oranlarının yüksek olmadığı, condition indekslerde maddeler arasında yüksek değerlere rastlanmadığı, tolerans değerlerinin 0'a yakın olduğu bu nedenle maddeler arasında çok yüksek doğrusal bağlantılılığın olmadığı ve Mahalanobis uzaklık testi sonuçlarına göre çok değişkenlik açısından aykırı gözlem bulunmadığı gözlenmiştir. Artıkların yani hata puanlarının normal dağıldığı ve ilişkisiz oldukları gözlenmiştir. Söz konusu bu analizler PIRLS projesi kapsamında uygulanan ve çok örneklemli doğrulayıcı faktör analizi yapılan diğer altı, sekiz ve sıfır kodlu alt testler içinde tekrar edilmiştir. Analizler sonucunda bu alt testlere ait veriler içinde benzer bulgulara ulaşılmıştır.

Analizlerin sonraki aşamasında her bir alt test için ölçme modeli oluşturulmuş ve çok örneklemli doğrulayıcı faktör analizi ile alt testlerin faktör yapılarının kültürler arası eş değeri test edilmiştir. (Byrne, 1994: 159-191; Jöreskog ve Sörbom, 1993: 51-61).

Bir veri seti için çok örneklemli doğrulayıcı faktör analizi aynı veri seti için oluşturulan dört adet iç-içe (nested) modelin analiz edilmesi ve bu modeller arasında karşılaştırmalar yapılarak uyum iyiliği en yüksek olan model üzerinden yorumlar yapılması aşamalarını kapsamaktadır. Şöyle ki, ModelA'da teste bulunan maddelerin ait oldukları faktörler belirlenir (ModelA ilgili literatürde baseline model, kuramsal model, temel model isimleriyle adlandırılmaktadır). ModelA'da maddeler ve gizil yapılar (faktörler) arasında kurulan yapısal ilişkiler araştırmacının mevcut olduğunu iddia ettiği ve bu analiz tekniğini kullanarak doğrulamaya çalıştığı kuramsal modele ait yapısal ilişkileri içeren ölçme modelidir. ModelA'da karşılaştırma yapılan gruplar için faktör yükleri, hata varyansları ve faktör korelasyonlarının sabit olduğu bir model kurulur ve analiz edilir. ModelB ModelA'daki ölçme modeline hiçbir değişken eklemekten veya değişken çıkarmadan oluşturulur. Fakat bu modelde gruplar arasında sadece faktör yüklerinin değişmesine izin verilir ve hata varyansları ile faktör korelasyonları sabit tutulur. Böylece ModelB maddelerin ait oldukları faktörlerdeki faktör yüklerinin karşılaştırma yapılan gruplarda aynı olup olmadığı ortaya çıkarır. ModelC'de faktör yüklerinin ve hata varyanslarının karşılaştırma yapılan grupların ölçme modelindeki değerlerinin farklılaşıp farklılaşmadığını anlamak için serbest bırakılırlar. Son olarak

ModelD'de sadece hata varyanslarının değerlerinin karşılaştırma yapılan gruplarda serbestçe değişmesine izin verilir (Jöreskog ve Sörbom, 1993; Klein, 1998; Schumacker ve Lomax, 2004). ModelB, C ve D için elde edilen uyum iyiliği ve diğer karşılaştırma indeksleri temel modele (ModelA) ait indekslerle karşılaştırılır. Model karşılaştırmalarında $\Delta\chi^2$ fark testinin anlamlı çıkması ve daha düşük χ^2 değeri veren modelin diğer uyum indekslerinin de aranan kritik değerleri sağlaması durumunda bu koşulları sağlayan modelin diğerine göre daha geçerli bir model olduğunu söylemek mümkün olmaktadır (Kline, 1998 , s: 224-228). Bu karşılaştırmalar için hazırlanan bir tablo üzerinden alt testlerin çok örneklemlili doğrulayıcı faktör analizi sonuçları özetlenerek yorumlanmıştır.

Araştırmanın ikinci alt probleminin a ve b maddelerinde belirtilen problemlerin çözümü için PIRLS projesi kapsamında Türkiye ve Amerika Birleşik Devletlerinde uygulanan üç, altı, sekiz ve sıfır kodlu alt testleri oluşturan 8 okuma parçasına ait testlerin maddelerinin kültürler arasında (Türkiye ve Amerika Birleşik Devletlerindeki uygulamalarında) değişen madde fonksiyonu gösterip göstermediği incelenmiştir. Bu amaçla, parametre karşılaştırma yöntemi ve olabilirlik oranı testi model karşılaştırma yöntemi kullanılarak alt testlerdeki değişen madde fonksiyonları incelenmiştir. Alt testlerdeki maddelerin bir kısmı iki kategorili, bir kısmı da kısmi puanlama yapılarak çok kategorili olarak puanlanmıştır. Bu durum göz önüne alınarak, her iki yöntemle yapılan DMF analizlerinde iki kategorili MTK modellerinden Rasch model, çok kategorili MTK modellerinden kısmi puanlama modeli kullanılmıştır (Embretson ve Reise, 2000). Söz konusu bu analizlere geçilmeden önce madde tepki kuramının varsayımlarının sağlanıp sağlanmadığı kontrol edilmiştir.

Tek boyutluluk:

PIRLS projesi kapsamında Türkiye ve Amerika Birleşik Devletlerinde uygulanan üç, altı, sekiz ve sıfır kodlu alt testleri oluşturan 8 okuma parçasına ait testlerin tek boyutluluk özelliğini sağlayıp sağlamadığını belirlemek için temel bileşenler faktör analizi tekniği kullanılmıştır. Her bir okuma parçasına ait testin hem kaynak kültürdeki örneklem verilerine hem de hedef kültürden elde edilen verilerine ayrı ayrı temel

bileşenler analizi yapıldığından toplam 8 okuma parçası verisi için 16 ayrı analiz yapıldı. Analiz sonuçları Tablo 2.1’de verilmiştir.

Tablo 2.1
Okuma parçalarına ait değerlendirme testleri için ABD ve Türkiye örneklemelerinde temel bileşenler faktör analizi sonuçları.

Okuma Parçası		ABD Örnekleme			Türkiye Örnekleme		
		I. faktör	II. faktör	III. faktör	I. Faktör	II. faktör	III. faktör
Flowers	Özdeğer	4,08	1,11	1,00	3,11	1,15	1,07
	Varyansı Açıklama %	31,39	8,54	7,72	23,97	8,91	8,29
Antarctica	Özdeğer	3,96	1,08	-	3,82	-	-
	Varyansı Açıklama %	36,07	9,85	-	34,75	-	-
Mice	Özdeğer	5,15	1,16	-	4,39	1,18	1,10
	Varyansı Açıklama %	36,83	8,32	-	31,35	8,48	7,86
River	Özdeğer	3,22	1,11	-	3,34	1,15	-
	Varyansı Açıklama %	29,35	10,12	-	30,41	10,52	-
Clay	Özdeğer	5,28	-	-	3,88	1,02	1,00
	Varyansı Açıklama %	39,64	-	-	29,88	7,91	7,75
Leonardo	Özdeğer	3,51	1,15	1,00	3,23	1,08	1,06
	Varyansı Açıklama %	29,26	9,65	8,33	26,98	9,02	8,8
Hare	Özdeğer	3,78	1,06	-	3,51	-	-
	Varyansı Açıklama %	34,38	9,65	-	31,94	-	-
Pufflings	Özdeğer	3,76	1,10	-	3,12	1,17	1,02
	Varyansı Açıklama %	28,98	8,51	-	23,99	9,06	7,88

Tek boyutluluk varsayımının karşılanıp karşılanmadığını incelemek üzere temel bileşenler faktör analizi yöntemi ile döndürülmemiş faktör çözümü uygulanmış ve özdeğerleri 1’den büyük olan boyutlar tabloda verilmiştir. Örneğin Flowers okuma parçasına ait testte, ABD örnekleminde birinci faktörün özdeğeri 4,08 ve varyansı açıklama miktarı % 31,39, ikinci faktörün özdeğeri 1,11 ve varyansı açıklama miktarı 8,54, üçüncü faktörün özdeğeri 1,00 ve varyansı açıklama miktarı 7,72 dir. Benzer şekilde, Flowers okuma parçasına ait testte, Türkiye örnekleminde birinci faktörün

özdeğeri 3,11 ve varyansı açıklama miktarı % 23,97, ikinci faktörün özdeğeri 1,15 ve varyansı açıklama miktarı 8,91, üçüncü faktörün özdeğeri 1,07 ve varyansı açıklama miktarı 8,29 dur. Birinci faktörde yer alan maddelerin faktör yüklerinin 0,35-0,70 aralığında değiştiği gözlenmiştir. Testlerin tamamına ait analizlerde ilgili örneklem gruplarında birinci faktör ile ikinci faktöre ait özdeğerler arasında ani bir kopma olduğu ve ikinci faktörlerle diğer faktörlerin özdeğerleri arasında önemli bir fark olmadığı gözlenmiştir. Ayrıca, analizlerin tümünde birinci faktörün varyansı açıklama miktarı %23'ün üzerinde çıkmıştır. Bu nedenle Lord (1980) ve Reckase (1979) tarafından tek boyutluluk için önerilen kriterler karşılanmış olduğundan söz konusu testlere ait verilerin ilgili örneklem gruplarında tek boyutluluk varsayımını sağladıkları anlaşılmıştır.

Yerel Bağımsızlık:

Madde test kuramında test maddelerinin cevaplanmasının birbirinden bağımsız olması gereklidir. Yani, yanıtlayamın yeteneği göz önünde bulundurulduğunda, yanıtlarının başka maddelerle hiçbir ilişkisinin olmaması için, söz konusu yetenek, maddeye verilen yanıtı etkileyen tek faktör olmalıdır. Bu nedenle, eğer örtük özellikte çok küçük aralıklar alınacak ve bu özellikleri ölçen maddeler ilişkilendirilecek olursa, tüm korelasyonlar 0 çıkmalıdır. Yerel bağımsızlık için yeterli şart olan tek boyutluluk, doğru tepki olasılığının sadece bir değişkenin fonksiyonu olduğu anlamına gelir; böylece, ölçme sonucunu sadece tek bir temel özellik açıklamış olur. Yukarıda verildiği gibi testlerin tamamı için tek boyutluluk varsayımı karşılanmış olduğundan yerel bağımsızlık varsayımı için yeterli koşulun sağlanmış olduğu düşünülmektedir.

Bu çalışma kapsamında değişen madde fonksiyonu analizleri, yukarıda bahsedilen yaklaşımlardaki; a-karşılaştırma yapılan gruplardan elde edilen parametreleri madde düzeyinde analiz eden parametre karşılaştırma yöntemi ve b-yine madde düzeyinde olabilirlik oranı testine dayalı olan model karşılaştırma yöntemleri ile yürütülmüştür. Bu nedenle her iki yöntem aşağıda kısaca açıklanmaktadır.

Değişen madde fonksiyonunun belirlenmesinde madde parametrelerini karşılaştırma yöntemi:

Madde düzeyinde değişen madde fonksiyonunun hesaplanmasında öncelikle her bir madde için referans ve odak (fokal) gruplardan hesaplanan madde ayırt etme ve madde güçlük parametre farklılık istatistikleri aşağıda verilen eşitlikler yardımıyla hesaplanır (Reise, Smith ve Furr, 2001; Smith, 2002).

$$\text{Değişen madde parametre farkı} = \hat{a}_{i(R)} - \hat{a}_{i(F)} \quad (13)$$

$$\text{Değişen madde parametre farkı} = b_{i(R)} - b_{i(F)} \quad (14)$$

Bu eşitlikler, bir madde için, maddenin referans gruptaki ayırt etme ya da madde güçlük parametre tahmin değerinden odak gruptaki parametre tahmin değerinin çıkarılmasıyla madde tahmin değerleri farkının doğrudan bir ölçmeünü verir. Bir sonraki aşmada parametrelerden elde edilen bu fark değerleri standardize edilir. Örneğin, madde güçlük parametresi için standardize değişen madde fonksiyonu fark (SDIF) değeri aşağıda verilen eşitlik ile sağlanır (Reise, Smith ve Furr, 2001; Smith, 2002).

$$SDIF = \frac{DIF}{\sqrt{\text{var } \hat{b}_{i(R)} + \text{var } \hat{b}_{i(F)}}} \quad (15)$$

Standardize değişen madde fonksiyonu fark istatistiği standart puanlara benzer ve karşılaştırma yapılan grupların madde parametre tahminleri arasındaki farkın (contrast) bir ölçmeünü verir. Standardize edilen fark değerinin karesi, 1 serbestlik derecesinde χ^2 istatistiği olarak değerlendirilebilmektedir (du Toit, 2003). Bu durumda eğer bir madde 0,01 veya 0,05 alfa manidarlık düzeyinde anlamlı bir χ^2 değerine sahipse maddenin gruplar arasında değişen madde fonksiyonu gösterdiğine karar verilir. Bu çalışma kapsamında madde parametrelerini karşılaştırma yöntemine göre değişen madde fonksiyonu analizleri PARSCALE programı ile yapılacaktır.

Camilli ve Shepard (1994), yntemsel alıřmalar dıřında, a ve c parametrelerinin deęiřen madde fonksiyonu analizlerinin alıřılmasını nermemektedirler. Bu grřlerine gereke olarak, a ve c parametrelerinin b parametresine gre tahminlerinin daha zayıf olduęunu bu nedenle bu iki parametre farkına ait yorumların daha az gvenilir olduęunu belirtmektedirler. Buna ek olarak, madde ayırt etme ve řans parametrelerine gre grup farklarını yorumlamada temel kavramsal ve pratik sorunların olduęunu, fakat dięer taraftan bir maddenin neden bir grup iin dięer gruba gre daha kolay ya da zor olduęunu aıklamanın arařtırmacılar tarafından daha bařarılı bir řekilde yapılabildięini sylemektedirler. Burada verilen gerekelere dayalı olarak bu alıřma kapsamında deęiřen madde fonksiyonu analizleri hem madde parametrelerini karřılařtırma ynteminde hem de olabilirlik oranı testi ynteminde sadece b parametresinin gruplarda deęiřen madde fonksiyonu gsterip gstermedięi analiz edilerek incelenecektir. Bu nedenle madde tepki kuramı modellerinden bir parametrelili kısmi puan modeli kullanılarak her iki DMF yntemi iin leklemeler ve parametre tahminleri gerekleřtirilecektir.

Parscale programı (Muraki ve Bock, 1996) madde fonksiyonu farklılıęının incelenmesinde madde ayırt etme, madde glk ve řans parametrelerinin referans ve odak gruplar arasındaki karřılařtırmaya dayalı olarak deęiřen madde fonksiyonu zmlemesi yapmaktadır. Program, arařtırmacıya karřılařtırma yapılacak gruplarda madde parametrelerinin sınırlandırılması ya da serbest bırakılması ve parametrelerin gsterdikleri farklılıkların ayrı ayrı test edilmesi iin olanak saęlamaktadır. Bir bařka deyiřle arařtırmacı istedięi sayıda parametre iin deęiřen madde fonksiyonu analizi yapabilmekte ve buna baęlı olarak uygun madde tepki kuramı modeli ile lekleme yapabilmektedir. Bu arařtırmada parametre karřılařtırma yntemine ile yapılan deęiřen madde fonksiyonu incelemelerinde sadece madde glk parametresine gre test maddelerinin deęiřen madde fonksiyonu gsterip gstermedikleri incelenmiřtir. Karřılařtırma gruplarında madde glk parametrelerinin deęerleri aısından anlamlı bir farklılařma olup olmadıęı 0,05 ve 0,01 anlamlılık dzeyinde ve "1" serbestlik derecesinde χ^2 testi ile deęerlendirilmiřtir.

Değişen madde fonksiyonunun belirlenmesinde olabilirlik oranı testine dayalı olan model karşılaştırma yöntemi:

İki kategorili (*dichotomous*) ve çok kategorili (*plythomous*) tepki modellerinin her ikisi için de DMF'nin incelenmesinde bir başka yöntem ise, dar ve geniş model olarak tanımlanan iki modelin olabilirlik oranı farklarının anlamlılığı üzerine kurulan model karşılaştırma yöntemidir (Thissen, Steinberg ve Wainer, 1993). Yöntemin gösterimi denklem 18. de verilmektedir.

$$\text{Olabilirlik Oranı (LR)} = L^*(\text{DarModel}) / L^*(\text{GenişModel}) \quad (16)$$

Geniş modelin kestirim için serbest bırakılan bir ya da daha fazla parametresi vardır. Örneğin, tipik olarak, a ve b parametrelerinin test edilen maddeler için gruptan gruba farklılık göstermesine izin verilir. Serbestlik dereceleri geniş ve dar modeldeki parametre sayısının farkı olarak hesaplanır. Elde edilen G^2 istatistiği ki-kare olarak dağılır; eğer bu değer ki-kare tablo değerinden büyükse, iki grup arasında madde parametrelerinde dikkate değer farklılıkların olduğu kabul edilir. Eğer sonuç anlamlı değilse, geniş modeldeki parametrelerden hiçbirinin 0'dan farklı olmadığı sonucuna varılır (Kim, Cohen, DiStefano ve Kim, 1998; Meade ve Lautenschlager, 2004; Teresi, Kleinman, Welikson, 2000). Bu model karşılaştırmasında amaç geniş modele ilave edilen parametrelere gerçekten ihtiyaç olup olmadığını ve bu ilave parametrelerin modelin uyum derecesini artırmada sıfırdan anlamlı bir biçimde farklılaşıp farklılaşmadığını test etmektir. Burada H_0 hipotezinin dar modeli yalnızca N sayıda madde parametresini içerdiğini ve H_1 hipotezininde geniş modeli yani N parametrelerine ilave olarak M sayıda parametreleri içerdiğini varsaydığımızda; dar model daha sade bir model olarak geniş modelden daha az parametreye sahip olacaktır.

Olabilirlik oranı testi ile örneklem verileri H_0 hipotezini desteklemekte midir sorusuna cevap aranır. Diğer bir deyişle ilave edilen M kadar parametre modelin uyum iyiliğini artırmakta mıdır? Olabilirlik oranının -2 defa logaritmik dönüşümü alındığında M serbestlik derecesinde özellikle geniş örneklem büyüklüklerinde χ^2 dağılımı gösteren bir test istatistiği elde edilir.

$$\chi^2 (M) \approx -2\ln(LR) = [-2\ln L^*(\text{Dar Model})] - [-2\ln L^*(\text{Geniş Model})] \quad (17)$$

Bu sonuca $\ln(x/y) = \ln x - \ln y$ matematiksel eşitliği ile varılmaktadır. Olabilirlik oranı testi model karşılaştırma yöntemi dar ve geniş modellerin -2 defa logaritmik dönüşümünden elde edilen değerlerinin birbirinden çıkarılmasını içermektedir.

Olabilirlik oranı testi ile yapılan DMF analizinden önce, incelenen maddelerin DMF içermediğini varsayan bir dar model oluşturulur. Dar modelde, iki grup için madde parametreleri eşitlenerek ölçekleme yapılır. İkinci aşamada DMF incelenen madde sayısı kadar geniş model oluşturulur. Her modelde bir maddenin DMF içerdiği varsayılarak bu maddenin parametreleri referans ve odak gruplarda serbest bırakılır, diğer maddelerin madde parametreleri eşitlenir. Böylece bir tane dar model için, k tane geniş model için olmak üzere k+1 sayıda analiz yapılır (Meade ve Lautenschlager, 2004; Holland ve Wainer, 1993).

Olabilirlik oranı testi ile değişen madde fonksiyonu incelenirken karşılaştırma yapılan grupların parametrelerini aynı metrik üzerinde eşitlemek üzere iç kriter olarak testin kendi maddelerinden oluşan bağ maddelerine (anchor items) ihtiyaç duyulur (Thissen, Steinberg ve Wainer, 1993). Burada madde parametrelerinin gruplara göre farklı olmadığını varsayan dar modelin olasılıkları, bağ maddeleri dışında kalan diğer maddelerin değişen madde fonksiyonu için incelendiği geniş modelin olasılıkları ile karşılaştırılır. Bağ maddelerin değişen madde fonksiyonu göstermeyen maddeler olduğu varsayılır ve bu maddelere değişen madde fonksiyonu analizi yapılmaz (Kim ve Cohen, 1995).

Bu çalışmada analizi yapılan testlerin bağ maddelerini seçmek için Kim ve Cohen (1995) tarafından önerilen tekrarlayıcı madde seçme işlemi kullanılmıştır. Bu işlemin basamakları aşağıda verilmektedir;

- 1- Tüm madde parametrelerinin karşılaştırma yapılan gruplarda eşitlendiği dar modelin $-2 \log L (G_1^2)$ değeri hesaplanır.

2- Her bir madde için parametrelerin karşılaştırma yapılan gruplarda serbest bırakıldığı geniş modellerin $-2 \log L (G_2^2)$ değerleri hesaplanır.

3-Testteki her madde için parametrelerinin serbest bırakıldığı geniş modelden elde edilen $-2 \log L$ değerleri başlangıçta tüm maddelerin parametrelerinin gruplar için eşitlendiği (parametrelerin karşılaştırma yapılan gruplarda sabit olduğu) dar modelin $-2 \log L$ değerinden çıkarılır. Yani $G_i^2 = G_{dar}^2 - G_{geniş}^2$ test istatistik değerleri bulunur. Elde edilen değer ilgili serbestlik derecesinde χ^2 tablo değeri ile karşılaştırılır ve anlamlı bir G_i^2 değeri bulunmuyorsa işlem bu adımda sonlandırılır. Eğer G_i^2 değeri anlamlı çıkan maddeler var ise bu değeri en büyük olan test maddesi testten çıkarılır.

4-Kalan maddeler için yeniden dar model oluşturulur ve $-2 \log L$ değeri elde edilir.

5-Kalan madde sayısı kadar yeni geniş modeller kurulur ve bu geniş modellerin $-2 \log L$ değerleri hesaplanır.

6- $G_i^2 = G_{dar}^2 - G_{geniş}^2$ test istatistik değerleri tekrar hesaplanır ve yalnızca en büyük G_i^2 değeri veren madde testten çıkartılır.

Bu tekrarlayıcı madde ayıklama işlemi testte kalan hiçbir maddenin anlamlı G_i^2 değeri vermediği aşamaya kadar sürdürülür. Sonuçta değişen madde fonksiyonu göstermeyeceği tahmin edilen bu maddelerden bağ maddeleri seti oluşturulur.

Olabilirlik oranı testi ile değişen madde fonksiyonu analizlerinin devamında, önce sadece bağ maddelerinin parametrelerinin karşılaştırılan gruplarda eşitlendiği diğer değişen madde fonksiyonu göstermeye aday maddelerin parametrelerinin gruplar için serbest bırakıldığı bir geniş model kurularak analiz yapılır ve bu geniş model için $-2 \log L$ değeri elde edilir. Daha sonra seçilen bağ maddelerinin dışında testte kalan ve değişen madde fonksiyonu göstermeye aday her bir madde için ayrı ayrı madde parametrelerinin karşılaştırma yapılan gruplarda sınırlandırıldığı (eşitlendiği) dar modeller kurulur. Her bir madde için dar modelinden elde edilen $-2 \log L$ değeri sadece bağ maddelerinin eşitlendiği geniş modelin $-2 \log L$ değerinden çıkarılarak G_i^2 değerleri hesaplanır. Elde edilen G_i^2 değeri ilgili serbestlik derecesinde χ^2 tablo değeri ile

karşılaştırılır ve tablo değerinden büyük G_i^2 değeri veren maddelerin değişen madde fonksiyonu gösterdiği belirtilerek analizler sonlandırılır.

2.5 Analizlerde Kullanılan İstatistik Programları

LISREL:

K. Jöreskog ve D. Sörbom (1993) tarafından geliştirilen LISREL LInear Structural RELations kelimelerin kısaltılması ile adlandırılan ve ölçmelerdeki kovaryans yapılarını analiz eden bir programdır. Yapılar arasındaki nedensel ilişkileri araştıran hipotezlerin test edilmesinde kullanılır. Yazılım kullanılarak regresyon analizinde olduğu gibi değişkenler arasındaki doğrusal ilişkiler araştırılır. Ayrıca gizil değişkenlerin özelliklerini incelemek amacıyla da kullanılır. Teorik ve soyut kavramlar dolaylı olarak, gizil değişkenler tanımlanarak ölçülür. Araştırmacı geliştirdiği nedensel ilişkilere dayanan ölçme modelinin gözlenen verilerle desteklenip desteklenmediğini analiz eder. Lisrel’de ölçme modelinin geçerlik ve güvenilirliğini hesaplamak için doğrulayıcı faktör analizi yöntemi kullanılır. Lisrel genel amaçlı bir istatistiki analiz programı değil, bir dizi yapısal eşitlik katsayılarını tahmin eden bir programdır. Lisrel programı ile ölçme modelleri ve doğrulayıcı faktör analizi, nedensel ilişki modelleri, gizil değişkenler için yapısal eşitlik modelleri, çok örneklemli analizler, gizil büyüme eğrileri modeli vb. gibi spesifik analizler yapılabilmektedir. Ayrıca yazılım, sıralı ve eşit aralıklı ölçek verileri için polikorik ve poliseriyal korelasyon matrisleri ile çalışma imkanı sağlamaktadır. Bu tez çalışması kapsamında birinci alt problemin çözümü için gerekli olan çok örneklemli doğrulayıcı faktör analizleri LISREL 8.54 yazılımı kullanılarak yapılmıştır. Ek 1.’ de sıfır kodlu PIRLS testi için oluşturulan çok örneklemli doğrulayıcı faktör analizi komut dosyaları örnek olarak verilmiştir.

PARSCALE:

Araştırmada parametre karşılaştırma yöntemine göre değişen madde fonksiyonu analizleri PARSCALE 4.1 programı kullanılarak yapılmıştır. Muraki ve Bock tarafından 1996 yılında geliştirilen program hem iki kategorili hem de çok kategorili puanlanmış

veriler için lojistik ve normal ogive yöntemlere uygulanabilen madde tepki kuramı modelleri ile madde ve test bazında ölçekleme yapılabilen bir programdır (du Toit, 2003). Program çoklu grup karşılaştırmaları ve değişen madde fonksiyonu çalışmalarında analiz imkanı sağlamaktadır. Özellikle değişen madde fonksiyonu çalışmalarında araştırmacıya istediği sayıda parametre ile analiz imkanı tanımaktadır. Program, değişik sayıda çoklu kategorili puanlanan maddelere uygulanabilme, madde parametrelerinin gruplar arasında kolayca sınırlayabilme, değişen madde fonksiyonu için istatistiksel testleri uygulayabilme ve sonuçları kolayca yorumlayabilme olanaklarını sunmaktadır.

PARSCALE programında genelleştirilmiş kısmi puan modeli aşağıda gösterildiği gibi parametreleştirilmektedir;

$$P_{jk}(\theta) = \frac{\exp[Da_j(\theta - b_j + d_{jk})]}{\sum_{v=1}^{m_j} \exp[Da_j(\theta - b_j + d_{jv})]} \quad (18)$$

Parscale programı ile parametrelerin elde edilmesi, Multilog ile yapılan parametre tahminlerinden üç önemli noktada farklılaşmaktadır.: 1- madde j sadece bir eğim parametresine sahiptir, a_j ; 2- toplama işlemi modelin belirlenmesinde puanlama fonksiyonunun amacına hizmet eder, böylece kontrastların tanımlanmasına ihtiyaç duyulmaz; 3-madde j 'nin, her bir tepki seçeneği ya da puan ($d_{j1}=0$) için ilave eşik parametreleriyle, d_{jk} , birlikte tek bir konum parametresi vardır, b_j . Parscale programında bir parametrelili (Rasch) kısmi puan modeli ile parametre tahmin yapılmak istenirse PRIORS, SPRIOR ve PRIORREAD komutları kullanılarak tüm maddeler için a_j parametresinin değeri 1 olarak sabitlenebilmektedir. Bu parametrelerin aynı veri setinden Multilog ve Parscale ile elde edilen tahminlerinin doğrudan kıyaslanması mümkün değildir, ancak her iki programın parametre tahminlerini birbirine dönüştüren işlemler yapıldıktan sonra kıyaslama mümkün olabilmektedir (Childs ve Chen, 1999). Ek 2.'de PARSCALE programında parametre karşılaştırma yöntemi ile yapılan değişen

madde fonksiyonu analizi için yazılan bir komut dosyası ve kısaltılmış analiz sonuçları ile birlikte verilmiştir.

MULTILOG:

Bu araştırma kapsamında model karşılaştırma yöntemine dayalı olarak olabilirlik oranı testi ile değişen madde fonksiyonu analizleri MULTILOG 7.03 programı ile yürütülmüştür. Madde tepki kuramına göre iki kategorili ve çok dereceli puanlanan ölçek maddelerinin parametrelerini lojistik parametre kestirimlerini yapan bir yazılımdır (Thissen, 1991). Programın derecelenmiş ölçekler için Samejima lojistik modelini, çok kategorili maddeler için nominal ve multinominal lojistik tepki modelini, çoktan seçmeli maddeler için çoktan seçmeli modeli kullanarak ölçekleme yapma imkanı sağlamaktadır. Program iki ya da daha fazla sayıdaki örnekleme ait madde tepki fonksiyonlarının tahmin edilmesinde marjinal maksimum olabilirlik (MML) yöntemini kullanarak, değişen madde fonksiyonunun incelenmesinde referans ve odak gruplar için eşzamanlı parametre tahmini yapabilmektedir (du Toit, 2003).

Multilog yazılımında ağırlıklandırılmış kısmi puan modeli aşağıdaki tanımlanmaktadır;

$$P_{jk}(\theta) = \frac{\exp(a_{jk}\theta + c_{jk})}{\sum_{u=1}^{m_j} \exp(a_{ju}\theta + c_{ju})}, \quad (19)$$

Bu parametrik ifade denklem 18'dekine benzemektedir, ama $c_{jk} = -a_{jk}b_{jk}$ ve puanlama fonksiyonu T_j kontrastlar kullanılarak ele alınmaktadır. Özellikle genelleştirilmiş kısmi puan modeli parametrelerini elde etmek için Multilog kullanımı parametreler arasındaki kontrastların (karşıtlıkların) tahminini gerektirir (Childs ve Chen, 1999). Bu karşıtlıkları tanımlamak için TMATRIX ve FIX komutları, komut dosyasına aşağıdaki örnekte verildiği gibi dahil edilmelidir (du Toit, 2003).

```
>TMATRIX ITEMS = (1,2,3,...,N), AK, POLYNOMIAL;
>FIX ITEMS = (1,3,7,...,N), AK = 2, VALUE = 0.0;
>FIX ITEMS = (2,4,5,...,N), AK = (2,3), VALUE = 0.0;
```

a_{jk} parametreleri için doğrusal puanlama fonksiyonuna, quadratik ve daha yüksek kontrastların sıfırda sabitlenmesiyle bu parametreler için polinom bir T matrisi tespit edilerek ulaşılır. Terk edilen doğrusal kontrastlar daha sonra puanlama fonksiyonu olarak işlem görür.

Multilog kullanarak, uygun olan genelleştirilmiş kısmi puan modeli madde parametre tahminleri üretmek için kontrastlar c_{jk} parametreleri için de aşağıdaki gibi tanımlanmalıdır.

>TMATRIX ITEMS = (1,2,3,.....,N), CK, TRIANGLE;

Denklem 12'deki a_{jk} ve c_{jk} parametreleri Multilog çıktısında bulunmaktadır. Multilog tarafından sağlanan kontrastlar (eksik olan T sapma matrisleri) bu parametreleri tahmin için kullanılır. Bir madde için a_{jk} parametreleri, tek bir kontrast değerinin tepki sayısı ya da puan kategorileri için kullanılan T matrisinin ilk sırası ile çarpılması yoluyla elde edilir. c_{jk} parametreleri ise, kontrast vektörünün uygun olan T sapma matrisi ile çarpılmasıyla hesaplanır. Multilog'da normal metric ogive yerine lojistik model kullanılır, bu nedenle D sabiti 1,7 değil D=1 olarak tanımlanmıştır. Yukarıda bahsedildiği gibi Multilog doğrudan genelleştirilmiş kısmi puan modelini kullanmaz. Sınıflamalı tepki modeli (nominal response model) üzerinden yukarıda verilen TMATRIX ve FIX komutları vasıtasıyla genelleştirilmiş kısmi puan modeline ulaşılır. Eğer araştırmacı genelleştirilmiş kısmi puan modeli yerine bir parametrelili Kısmi Puan Modeli ile çalışmak isterse a_{jk} parametresine tüm maddeler için sabit bir değer vermek zorundadır. Multilog programının EQUAL komutu kullanılarak istenilen durum sağlanabilmektedir. Ek 3.'de MULTLOG programında olabilirlik oranı testi model karşılaştırma yöntemi ile yapılan değişen madde fonksiyonu analizi için yazılan bir komut dosyası ve kısaltılmış analiz sonuçları verilmiştir.

Diğer betimleyici analizler ve temel bileşenler analizleri SPSS 10.0 programı kullanılarak yapılmıştır.

BÖLÜM III

BULGULAR ve YORUMLAR

Bu bölümde, araştırma sonucunda elde edilen bulgular tablo, grafik ve açıklamalarla verilmiş ve yorumlanmıştır.

Araştırmanın birinci alt probleminin a, b, c ve d maddelerinde belirtilen problemlerin çözümü için PIRLS projesi kapsamında Türkiye ve Amerika Birleşik Devletlerinde uygulanan 3, 6, 8 ve 0 kodlu alt testlerin faktör yapılarının kültürler arası eş değeri yapısal eşitlik modelleri yöntemlerinden çok örneklemlili doğrulayıcı (confirmatory) faktör analizi yöntemi ile incelenmiştir.

3.1. Alt Problem 1a.'ya İlişkin Bulgular ve Yorumlar:

Bu alt problemde incelenen Flowers ve Antarctica okuma parçalarından oluşan üç kodlu testin dağılımına ait betimleyici istatistikler ülkelere göre Tablo 3.1'de verilmektedir.

Tablo 3.1
Flowers ve Antarctica Okuma Parçalarından Oluşan Üç Kodlu Alt Test
Puanlarının Dağılımına Ait Betimsel İstatistikler

Betimsel İstatistikler	ABD örneklemlili	Türkiye örneklemlili
Aritmetik Ortalama	22,46	14,61
Ortanca	24,00	14,00
Standart Sapma	6,80	6,78
Varyans	46,24	46,01
Bağıl Değişim Katsayısı	30	46
Çarpıklık	-0,98	-0,02
Basıklık	0,41	-0,89
Ranj	31	29
En küçük puan	1	0
En yüksek puan	32	29
Madde Sayısı	24	24
Birey Sayısı	315	423
Testin Güvenirlik Katsayısı	0,87	0,85

Teste ait toplam puanlara bakıldığında ABD örneklemindeki öğrencilerin puan ortalamalarının Türkiye örneklemindeki öğrencilerin puan ortalamasından oldukça yüksek olduğu, bağıl değişim katsayıları incelendiğinde ise ABD örneklemlili ait puan

dağılımının daha homojen ve simetrik olduğu görülmektedir. Türkiye örnekleminde en yüksek puan 29 iken ABD örnekleminde en yüksek puan 32 olarak gözlenmiştir. Çarpıklık katsayılarına bakıldığında, ABD örnekleminin puan dağılımının daha fazla sola çarpık olduğu yani ABD’li öğrencilerin testte daha başarılı oldukları anlaşılmaktadır. Testin güvenirlik katsayıları ABD örnekleminin için Cronbach alfa=0,87 Türkiye örnekleminin için ise 0,82 olarak hesaplanmıştır. Çarpıklık ve basıklık katsayılarına ait değerlerin ± 1 aralığında değişiyor olduğundan iki ülke örnekleminin ait puanların normal dağılıma sahip olduklarını söylemek mümkün görünmektedir (George ve Mallery, 2001).

Flowers ve Antarctica okuma parçalarından oluşan üç kodlu test (booklet) için çok örneklemlili doğrulayıcı faktör analizi sonuçları:

Flowers ve Antarctica okuma pasajlarını ait değerlendirme sorularından oluşan üç kodlu testteki maddelerin ölçtüğü varsayılan kavrama süreçlerine (boyutlar) göre dağılımı Tablo 3.2’de verilmektedir (Gonzalez ve Kennedy, 2003; Mullis, Martin, Gonzalez ve Kennedy, 2003).

Tablo 3.2
Üç Kodlu Alt Testte Bulunan Maddelerin Ölçtüğü Varsayılan
Kavrama Süreçlerine Göre Dağılımlarını Gösteren Model

KAVRAMA SÜREÇLERİ/ BOYUTLAR	Okuma Parçaları ve maddeler	
	Flowers	Antarctica
Açık bir şekilde ifade edilmiş bilgi ve fikirlerin üzerinde durma ve çıkarımlar yapma	4,9	1,2,3,6,8
Doğrudan çıkarımlar yapma	2,3,5,6,10	5,7
Fikir ve bilgileri birleştirme ve yorumlama	1,7,8,12	4,9
Metnin öğelerini, içeriğini ve dilini inceleme ve değerlendirme	11,13	10,11

Tablo 3.2’de görüldüğü gibi Flowers okuma parçasında 13 soru Antarctica okuma parçasında ise 11 soru vardır. Teste ait bu yapının kaynak kültürde ve hedef kültürde eşdeğer olup olmadığını incelemek için ABD örnekleminin ve Türkiye örnekleminin ait veriler üzerinden çok değişkenli doğrulayıcı faktör analizi yapılmıştır. Modellere ait analiz sonuçlarını özetlemek ve gerekli karşılaştırmaları yapabilmek için aşağıda verilen Tablo 3.3 hazırlanmıştır.

Tablo 3.3 incelendiğinde ModelA, B ve D'nin uyum iyiliği ve diğer model değerlendirme indekslerinin iyi bir modelde bulunması gereken kritik değerlere sahip olmadıkları görülmektedir.

Tablo 3.3
Modeller ve Karşılaştırmalarına Ait Çok Örneklemli DFA Sonuçları

Modeller	χ^2_{sd}	χ^2 / sd	GFI	CFI	NFI	SRMR	RMSEA	$\Delta\chi^2_{sd}$
ModelA	7927,45 ₍₅₄₆₎	14,51	0,94	0,97	0,96	0,09	0,19	-
ModelB	8486,47 ₍₅₂₂₎	16,25	0,96	0,98	0,97	0,08	0,20	*559,02 ₍₂₄₎
ModelC	1264,75 ₍₄₉₈₎	2,53	0,98	0,99	0,98	0,04	0,05	*6662,70 ₍₄₈₎
ModelD	1897,15 ₍₅₂₂₎	3,63	0,95	0,98	0,97	0,06	0,08	*6030,3 ₍₂₄₎

ModelA: Faktör Yükleri, Faktör Korelasyonları ve Hata Varyansları Sabit.

ModelB: Faktör Yükleri serbest (Faktör Korelasyonları ve Hata Varyansları Sabit).

ModelC: Faktör Yükleri ve Hata Varyansları serbest (Faktör Korelasyonları Sabit).

ModelD: Hata Varyansları serbest (Faktör Yükleri ve Faktör Korelasyonları Sabit).

P<0,01

Tablo 3.3'de görüldüğü gibi ModelB, ModelC ve ModelD nin temel model olan ModelA ile yapılan ikili karşılaştırmalarında, maddelerin faktör yükleri ile hata varyanslarının serbest ve faktör korelasyonlarının sabit tutulduğu ModelC'nin uyum iyiliği indeksleri en yüksek ve diğer model değerlendirme kriterlerini sağlayan en iyi model olduğu anlaşılmaktadır. Uyum iyiliğinde en yüksek artış ModelC'de olmuştur. ModelA'nın χ^2 değeri 7927,45 iken ModelC'de bu değer 1264,75 olarak gerçekleşmiştir. Bunun anlamı ise, Flowers ve Antarctica okuma parçalarına ait değerlendirme sorularından oluşan 3 kodlu testin yapısının kaynak kültürde ve hedef kültürde eşdeğer olmadığı yani bu test için kültürler arası ölçme değişmezliğinin sağlanamamış olduğudur. Çünkü, ModelA'da kurulan ölçme modeli uyum iyiliği kriterlerini sağlayamayan bir modeldir, yani test maddeleri kuramsal olarak önerildiği gibi dört ayrı kavrama süreci ile belirtilen boyutlarla iddia edilen ilişkileri vermemektedir. Test maddelerin kaynak kültür ve hedef kültür için kurulan doğrulayıcı faktör analizi modelinde faktör yükleri ve hata varyanslarına ait değerlerinin eşit olmadığı, bu değerlerin kültürlerle göre anlamlı bir şekilde farklılık gösterdiği anlaşılmaktadır.

3.2. Alt Problem 1b.'ye İlişkin Bulgular ve Yorumlar:

Bu alt problemde incelenen Mice ve River okuma parçalarından oluşan altı kodlu testin dağılımına ait betimleyici istatistikler ülkelere göre Tablo 3.4'de verilmektedir.

Tablo 3.4
Mice ve River Okuma Parçalarından Oluşan Altı Kodlu Alt Test
Puanlarının Dağılımına Ait Betimsel İstatistikler

Betimsel İstatistikler	ABD örnekleme	Türkiye örnekleme
Aritmetik Ortalama	23,15	15,11
Ortanca	24,00	15,00
Standart Sapma	7,24	7,80
Varyans	52,44	60,91
Bağıl Değişim Katsayısı	31	51
Çarpıklık	-0,90	0,18
Basıklık	0,41	-0,85
Ranj	33	33
En küçük puan	2	1
En yüksek puan	35	34
Madde Sayısı	25	25
Birey Sayısı	308	422
Testin Güvenirlilik Katsayısı	0,88	0,86

Testin toplam puanlarına bakıldığında ABD örneklemindeki öğrencilerin puan ortalamalarının Türkiye örneklemindeki öğrencilerin puan ortalamasından oldukça yüksek olduğu, bağıl değişim katsayıları incelendiğinde ise ABD örneklemine ait puan dağılımının daha homojen ve simetrik olduğu görülmektedir. Türkiye örnekleminde en yüksek puan 34 iken ABD örnekleminde en yüksek puan 35 olarak gözlenmiştir. Çarpıklık katsayılarına bakıldığında, ABD örnekleme puan dağılımının daha fazla sola çarpık olduğu yani ABD'li öğrencilerin testte daha başarılı oldukları anlaşılmaktadır. Basıklık katsayılarına bakıldığında ise ABD örnekleme puan dağılımının daha dik ve homojen Türkiye örnekleme puan dağılımının ise daha basık ve heterojen olduğu görülmektedir. Testin güvenilirlik katsayıları ABD örnekleme için Cronbach alfa=0,88 Türkiye örnekleme için ise 0,86 olarak hesaplanmıştır. Çarpıklık ve basıklık katsayılarına ait değerlerin ± 1 aralığında değişmesi sonucu her iki ülke örneklemine ait puanların normal dağılıma sahip olduklarını söylemek mümkün görünmektedir.

Mice ve River okuma parçalarından oluşan altı kodlu test (booklet) için çok örneklemli doğrulayıcı faktör analizi sonuçları:

Mice ve River okuma parçalarına ait değerlendirme sorularından oluşan altı kodlu testteki maddelerin ölçtüğü varsayılan kavrama süreçlerine göre dağılımı Tablo 3.5’de verilmektedir (Gonzalez ve Kennedy, 2003; Mullis, Martin, Gonzalez ve Kennedy, 2003).

Tablo 3.5
Altı Kodlu Alt Testte Bulunan Maddelerin Ölçtüğü Varsayılan
Kavrama Süreçlerine Göre Dağılımlarını Gösteren Model

KAVRAMA SÜREÇLERİ/ BOYUTLAR	Okuma Parçaları ve maddeler	
	Mice	River
Açık bir şekilde ifade edilmiş bilgi ve fikirlerin üzerinde durma ve çıkarımlar yapma	2,5,10	3,6
Doğrudan çıkarımlar yapma	1,3,7,9,	4,5,7
Fikir ve bilgileri birleştirme ve yorumlama	4,6,11,12	8,9,10
Metnin öğelerini, içeriğini ve dilini inceleme ve değerlendirme	8,13,14	1,2,11

Tablo 3.5’de görüldüğü gibi Mice okuma parçasında 14 soru River okuma parçasında ise 11 soru vardır. Teste ait bu yapının kaynak kültürde ve hedef kültürde eşdeğer olup olmadığını incelemek için ABD örnekleme ve Türkiye örnekleme ait veriler üzerinden çok değişkenli doğrulayıcı faktör analizi yapılmıştır. Modellere ait analiz sonuçları Tablo 3.6’da verilmektedir.

Tablo 3.6
Modeller ve Karşılaştırmalarına Ait Çok Örneklemli DFA Sonuçları

Modeller	χ^2_{sd}	χ^2 / sd	GFI	CFI	NFI	SRMR	RMSEA	$\Delta\chi^2_{sd}$
ModelA	3192,50 ₍₅₉₄₎	5,37	0,95	0,96	0,95	0,08	0,11	-
ModelB	3145,69 ₍₅₆₉₎	5,52	0,97	0,99	0,97	0,07	0,11	*46,81 ₍₂₅₎
ModelC	3312,43 ₍₅₄₄₎	6,08	0,97	0,99	0,98	0,06	0,11	*119,93 ₍₅₀₎
ModelD	2630,67 ₍₅₆₉₎	4,62	0,95	0,97	0,95	0,09	0,10	*561,83 ₍₂₅₎

ModelA: Faktör Yükleri, Faktör Korelasyonları ve Hata Varyansları Sabit.

ModelB: Faktör Yükleri serbest (Faktör Korelasyonları ve Hata Varyansları Sabit).

ModelC: Faktör Yükleri ve Hata Varyansları serbest (Faktör Korelasyonları Sabit).

ModelD: Hata Varyansları serbest (Faktör Yükleri ve Faktör Korelasyonları Sabit).

* P<0,01

Tablo 3.6’da görüldüğü gibi ModelB, ModelC ve ModelD nin temel model olan ModelA ile yapılan ikili karşılaştırmalarında, maddelerin hata varyanslarının serbest ve faktör yükleri ile faktör korelasyonlarının sabit tutulduğu ModelD’nin uyum iyiliği indeksleri nispeten kabul edilebilir düzeyde ve diğer model değerlendirme kriterlerini sağlayan en iyi model olduğu anlaşılmaktadır. ModelD için hesaplanan χ^2/sd indeksi diğerlerini göre en uygun oranı vermiştir. Uyum iyiliğinde en yüksek artış da yine ModelD’de olmuştur. ModelA’nın χ^2 değeri 3192,50 iken ModelD için bu değer 2630,67 olarak gerçekleşmiştir. Bunun anlamı ise, Flowers ve Antarctica okuma parçalarına ait değerlendirme sorularından oluşan 6 kodlu testin yapısının kaynak kültürde ve hedef kültürde eşdeğer olmadığı yani bu test için kültürler arası ölçme değişmezliğinin sağlanamamış olduğudur. Çünkü, ModelA’da kurulan ölçme modeli uyum iyiliği kriterlerini sağlayamayan bir modeldir, yani test maddeleri kuramsal olarak önerildiği gibi dört ayrı kavrama süreci ile belirtilen boyutlarla iddia edilen ilişkileri vermemektedir.

Test maddelerinin kaynak kültür ve hedef kültürdeki yapısal ilişkileri için kurulan doğrulayıcı faktör analizi modelinde (ölçme modeli) faktör yükleri ve hata varyanslarına ait değerlerinin eşit olmadığı, bu değerlerin kültürlerle göre anlamlı bir şekilde farklılık gösterdiği anlaşılmaktadır.

3.3. Alt Problem 1c.’ye İlişkin Bulgular ve Yorumlar:

Bu alt problemde incelenen Clay ve Leonardo okuma parçalarından oluşan sekiz kodlu testin dağılımına ait betimleyici istatistikler ülkelere göre Tablo 3.7’de verilmektedir. Testin toplam puanlarına bakıldığında ABD örneklemindeki öğrencilerin puan ortalamalarının Türkiye örneklemindeki öğrencilerin puan ortalamasından oldukça yüksek olduğu görülmektedir. Türkiye örnekleminde en yüksek puan 32 iken ABD örnekleminde en yüksek puan 34 olarak gözlenmiştir. Bağıl değişkenlik karsayıları incelendiğinde ise ABD örneklemine ait puan dağılımının daha homojen ve simetrik olduğu görülmektedir. Çarpıklık katsayılarına bakıldığında, ABD örnekleminin puan dağılımının daha fazla sola çarpık olduğu yani ABD’li öğrencilerin testte daha başarılı oldukları anlaşılmaktadır. Türkiye örnekleminin puan dağılımının daha fazla sağa çarpık

olduğu yani öğrencilerin testte daha başarısız oldukları anlaşılmaktadır. Basıklık katsayılarına bakıldığında ise ABD örnekleme puan dağılımını ile Türkiye örnekleme puan dağılımı benzerlik göstermektedir. Her iki dağılımda basık ve heterojendir.

Tablo 3.7
Clay ve Leonardo Okuma Parçalarından Oluşan Sekiz Kodlu
Alt Test Puanlarının Dağılımına Ait Betimsel İstatistikler

Betimsel İstatistikler	ABD örnekleme	Türkiye örnekleme
Aritmetik Ortalama	19,49	12,31
Ortanca	21,00	11,00
Standart Sapma	7,93	6,98
Varyans	62,96	48,74
Bağıl değişim Katsayısı	40	56
Çarpıklık	-0,51	0,48
Basıklık	-0,66	-0,62
Ranj	33	31
En küçük puan	1	1
En yüksek puan	34	32
Madde Sayısı	25	25
Birey Sayısı	321	435
Testin Güvenirlik Katsayısı	0,89	0,85

Testin iç tutarlık güvenirlik katsayıları ABD örnekleme için Cronbach alfa=0,89 Türkiye örnekleme için ise 0,85 olarak hesaplanmıştır. Çarpıklık ve basıklık katsayılarına ait değerlerin ± 1 aralığında değişmesi sonucu her iki ülke örnekleme puanlarının normalden kabul edilebilir sınırlar içinde sapma gösterdiklerini söylemek mümkün görünmektedir.

Clay ve Leonardo okuma parçalarından oluşan sekiz kodlu test için çok örneklemlili doğrulayıcı faktör analizi sonuçları:

Clay ve Leonardo okuma parçalarına ait değerlendirme sorularından oluşan sekiz kodlu testteki maddelerin ölçtüğü varsayılan kavrama süreçlerine göre dağılımı Tablo 3.8’de verilmektedir.

Tablo 3.8
Sekiz Kodlu Alt Testte Bulunan Maddelerin Ölçtüğü Varsayılan
Kavrama Süreçlerine Göre Dağılımlarını Gösteren Model

KAVRAMA SÜREÇLERİ/ BOYUTLAR	Okuma Parçaları ve maddeler	
	Clay	Leonardo
Açık bir şekilde ifade edilmiş bilgi ve fikirlerin üzerinde durma ve çıkarımlar yapma	1,8,9	1,3,9
Doğrudan çıkarımlar yapma	2,3,4,5,7	2,5,6,8
Fikir ve bilgileri birleştirme ve yorumlama	6,10,11,13	4,7,10
Metnin öğelerini, içeriğini ve dilini inceleme ve değerlendirme	12	11,12

Tablo 3.8’de görüldüğü gibi Clay okuma parçasında 13 soru Leonardo okuma parçasında ise 12 soru vardır. Teste ait bu yapının kaynak kültürde ve hedef kültürde eşdeğer olup olmadığını incelemek için ABD örnekleme ve Türkiye örnekleme ait veriler üzerinden çok değişkenli doğrulayıcı faktör analizi yapılmıştır. Modellere ait analiz sonuçlarını özetlemek ve gerekli karşılaştırmaları yapabilmek için aşağıda verilen Tablo 3.9 hazırlanmıştır.

Tablo 3.9
Modeller ve Karşılaştırmalarına Ait Çok Örneklemli DFA Sonuçları

Modeller	χ^2_{sd}	χ^2 / sd	GFI	CFI	NFI	SRMR	RMSEA	$\Delta\chi^2_{sd}$
ModelA	4782,81 ₍₅₉₄₎	8,05	0,84	0,84	1,00	0,17	0,14	-
ModelB	2988,55 ₍₅₆₉₎	5,26	0,91	0,91	1,00	0,21	0,11	*1783,26 ₍₂₅₎
ModelC	1233,49 ₍₅₄₄₎	2,26	0,97	0,97	1,00	0,04	0,05	*3549,32 ₍₅₀₎
ModelD	3667,27 ₍₅₆₉₎	6,44	0,88	0,88	1,00	0,09	0,12	*1115,54 ₍₂₅₎

ModelA: Faktör Yükleri, Faktör Korelasyonları ve Hata Varyansları Sabit.

ModelB: Faktör Yükleri serbest (Faktör Korelasyonları ve Hata Varyansları Sabit).

ModelC: Faktör Yükleri ve Hata Varyansları serbest (Faktör Korelasyonları Sabit).

ModelD: Hata Varyansları serbest (Faktör Yükleri ve Faktör Korelasyonları Sabit).

* P<0,01

Tablo 3,9 incelendiğinde ModelA, B ve D’nin uyum iyiliği ve diğer model değerlendirme indekslerinin iyi bir modelde bulunması gereken kritik değerlere sahip olmadıkları görülmektedir. Model karşılaştırmalarında ise $\Delta\chi^2$ fark testinin anlamlı çıkması ve daha düşük χ^2 değeri veren modelin diğer uyum indekslerinin de aranılan kritik değerleri sağlaması durumunda bu koşulları sağlayan modelin diğerine göre daha geçerli bir model olduğunu söylemek mümkün olmaktadır (Kline, 1998 , p=224-228).

Tabloda görüldüğü gibi ModelB, ModelC ve ModelD nin temel model olan ModelA ile yapılan ikili karşılaştırmalarında, maddelerin faktör yükleri ile hata varyanslarının serbest ve faktör korelasyonlarının sabit tutulduğu ModelC'nin uyum iyiliği indeksleri en yüksek ve diğer model değerlendirme kriterlerini sağlayan en iyi model olduğu anlaşılmaktadır. Uyum iyiliğinde en yüksek artış ModelC'de olmuştur. ModelC için uyum indeksleri değerleri GFI= 0,97, CFI= 0,88, NFI= 1,00, SRMR= 0,04 ve RMSEA= 0,05 olarak elde edilmiştir. χ^2 /sd oranı en iyi olan modelde yine 2,26 değeri ile ModelC dir. ModelA'nın χ^2 değeri 4782,81 iken ModelC'de bu değer 1233,49 olarak gerçekleşmiştir ve χ^2 değerinde en yüksek iyileşme 3549,32 değeri ile ModelC için elde edilmiştir. Bunun anlamı ise, Clay ve Leonardo okuma parçalarına ait değerlendirme sorularından oluşan sekiz kodlu testin yapısının kaynak kültürde ve hedef kültürde eşdeğer olmadığı yani bu test için kültürler arası ölçme değişmezliğinin sağlanamamış olduğu anlaşılmaktadır. Çünkü, ModelA'da kurulan ölçme modeli uyum iyiliği kriterlerini sağlayamayan bir modeldir, yani test maddeleri kuramsal olarak önerildiği gibi dört ayrı kavrama süreci ile belirtilen boyutlarla iddia edilen ilişkileri vermemektedir.

Test maddelerinin kaynak kültür ve hedef kültür için kurulan doğrulayıcı faktör analizi modelinde (ölçme modeli) faktör yükleri ve hata varyanslarına ait değerlerinin eşit olmadığı, bu değerlerin kültürlerle göre anlamlı bir şekilde farklılık gösterdiği anlaşılmaktadır.

3.4. Alt Problem 1d.'ye İlişkin Bulgular ve Yorumlar:

Bu alt problemde incelenen Hare ve Pufflings okuma parçalarından oluşan sıfır kodlu testin dağılımına ait betimleyici istatistikler ülkelere göre Tablo 3.10'da verilmektedir. Testin toplam puanlarına bakıldığında ABD örneklemindeki öğrencilerin puan ortalamalarının Türkiye örneklemindeki öğrencilerin puan ortalamasından oldukça yüksek olduğu görülmektedir. Bağıl değişkenlik karsayıları incelendiğinde ise ABD örneklemine ait puan dağılımının daha homojen ve simetrik olduğu görülmektedir.

Tablo 3.10
Hare ve Pufflings Okuma Parçalarından Oluşan Sıfır Kodlu
Alt Test Puanlarının Dağılımına Ait Betimsel İstatistikler

Betimsel İstatistikler	ABD örnekleme	Türkiye örnekleme
Aritmetik Ortalama	20,26	14,30
Ortanca	22,00	14,00
Standart Sapma	6,50	6,48
Varyans	42,29	41,99
BağılDeğişim Katsayısı	32	45
Çarpıklık	-0,76	0,13
Basıklık	-0,60	-0,76
Ranj	31	31
En küçük puan	1	1
En yüksek puan	32	32
Madde Sayısı	24	24
Birey Sayısı	926	1283
Testin Güvenirlik Katsayısı	0,86	0,83

Çarpıklık katsayılarına bakıldığında, ABD örnekleme puan dağılımının daha fazla sola çarpık olduğu, ABD’li öğrencilerin testte daha başarılı oldukları anlaşılmaktadır. Türkiye örnekleme puan dağılımının daha fazla sağa çarpık olduğu yani öğrencilerin testte daha başarısız oldukları anlaşılmaktadır. Basıklık katsayılarına bakıldığında ise ABD örnekleme puan dağılımını ile Türkiye örnekleme puan dağılımı benzerlik göstermektedir. Her iki dağılımda basık ve heterojendir.

Testin iç tutarlık güvenirlik katsayıları ABD örnekleme için Cronbach Alfa= 0,86 Türkiye örnekleme için ise 0,83 olarak hesaplanmıştır. Çarpıklık ve basıklık katsayılarına ait değerlerin ± 1 aralığında değişmesi sonucu her iki ülke örneklemine ait puanların normalden kabul edilebilir sınırlar içinde sapma gösterdiklerini söylemek mümkün görünmektedir.

Hare ve Pufflings okuma parçalarından oluşan sıfır kodlu (booklet) test için çok örneklemlili doğrulayıcı faktör analizi sonuçları:

Bu alt problemde incelenen Hare ve Pufflings okuma parçalarından oluşan sıfır kodlu testin dağılımına ait betimleyici istatistikler ülkelere göre Tablo 3.11’de verilmektedir.

Tablo 3.11
Sıfır Kodlu Alt Testte Bulunan Maddelerin Ölçtüğü Varsayılan
Kavrama Süreçlerine Göre Dağılımlarını Gösteren Model

KAVRAMA SÜREÇLERİ/ BOYUTLAR	Okuma Parçaları ve maddeler	
	Hare	Pufflings
Açık bir şekilde ifade edilmiş bilgi ve fikirlerin üzerinde durma ve çıkarımlar yapma	1,2,4	1,2,3,5,9
Doğrudan çıkarımlar yapma	5,7	10
Fikir ve bilgileri birleştirme ve yorumlama	6,8,9,10,11	6,7,8,11,12,4
Metnin öğelerini, içeriğini ve dilini inceleme ve değerlendirme	3	13

Hare okuma parçasında 11 soru Pufflings okuma parçasında ise 13 soru vardır. Teste ait bu yapının kaynak kültürde ve hedef kültürde eşdeğer olup olmadığını incelemek için ABD örnekleme ve Türkiye örnekleme ait veriler üzerinden çok değişkenli doğrulayıcı faktör analizi yapılmıştır. Modellerle ait analiz sonuçlarını özetlemek ve gerekli karşılaştırmaları yapabilmek için aşağıda verilen Tablo 3.12 hazırlanmıştır.

Tablo 3.12
Modeller ve Karşılaştırmalarına Ait Çok Örneklemli DFA Sonuçları

Modeller	χ^2_{sd}	χ^2 / sd	GFI	CFI	NFI	SRMR	RMSEA	$\Delta\chi^2_{sd}$
ModelA	2225,46 ₍₅₄₆₎	4,07	0,93	0,76	0,75	0,09	0,09	-
ModelB	2083,34 ₍₅₂₂₎	3,99	0,98	0,95	0,93	0,06	0,09	*142,12 ₍₂₄₎
ModelC	1751,83 ₍₄₉₈₎	3,51	0,99	0,98	0,96	0,05	0,06	*473,59 ₍₄₈₎
ModelD	2876,77 ₍₅₂₂₎	5,51	0,96	0,87	0,85	0,08	0,11	*651,31 ₍₂₄₎

ModelA: Faktör Yükleri, Faktör Korelasyonları ve Hata Varyansları Sabit.

ModelB: Faktör Yükleri serbest (Faktör Korelasyonları ve Hata Varyansları Sabit).

ModelC: Faktör Yükleri ve Hata Varyansları serbest (Faktör Korelasyonları Sabit).

ModelD: Hata Varyansları serbest (Faktör Yükleri ve Faktör Korelasyonları Sabit).

* P<0,01

Tablo 3.12 incelendiğinde ModelA, B ve D'nin uyum iyiliği ve diğer model değerlendirme indekslerinin iyi bir modelde bulunması gereken kritik değerlere sahip olmadıkları görülmektedir. Model karşılaştırmalarında ise $\Delta\chi^2$ fark testinin anlamlı çıkması ve daha düşük χ^2 değeri veren modelin diğer uyum indekslerinin de aranılan

kritik deęerleri saęlaması durumunda bu kořulları saęlayan modelin dięerine gre daha geerli bir model olduęunu sylemek mmkn olmaktadır (Kline, 1998 , p=224-228). Tabloda grldę gibi ModelB, ModelC ve ModelD nin temel model olan ModelA ile yapılan ikili karřılařtırmalarında, maddelerin faktr ykleri ile hata varyanslarının serbest ve faktr korelasyonlarının sabit tutulduęu ModelC'nin uyum iyilięi indeksleri en yksek ve dięer model deęerlendirme kriterlerini saęlayan en iyi model olduęu anlařılmaktadır. Uyum iyilięinde en yksek artıř ModelC'de olmuřtur. ModelC iin uyum indeksleri deęerleri GFI= 0,99, CFI= 0,98, NFI= 0,96, SRMR= 0,05 ve RMSEA= 0,06 olarak elde edilmiřtir. χ^2/sd oranı en iyi olan modelde yine 3,51 deęeri ile ModelC dir. ModelA'nın χ^2 deęeri 2225,46 iken ModelC'de bu deęer 1751,83 olarak gerekleřmiřtir ve χ^2 deęerinde en yksek iyileřme 473,59 deęeri ile ModelC iin elde edilmiřtir. Bunun anlamı ise, Hare ve Pufflings okuma paralarına ait deęerlendirme sorularından oluřan 0 kodlu testin yapısının kaynak kltrde ve hedef kltrde eřdeęer olmadıęı yani bu test iin kltrlr arası lme deęiřmezlięinin saęlanamamıř olduęudur. nk, ModelA'da kurulan lme modeli uyum iyilięi kriterlerini saęlayamayan bir modeldir, yani test maddeleri kuramsal olarak nerildięi gibi drt ayrı kavrama sureci ile belirtilen boyutlarla iddia edilen iliřkileri vermemektedir.

Test maddelerinin kaynak kltr ve hedef kltr iin kurulan doęrulayıcı faktr analizi modelinde (lme modeli) faktr ykleri ve hata varyanslarına ait deęerlerinin eřit olmadıęı, bu deęerlerin kltrlere gre anlamlı bir řekilde farklılık gsterdięi anlařılmaktadır.

ok rneklemlili doęrulayıcı faktr analizleri sonucunda btn alt testlerde faktr yklerinin ve hata varyanslarına ait deęerlerin kaynak kltrden elde edilen lmeleri ile hedef kltrden elde edilen lmeleri arasında anlamlı bir farklılık olduęu yani bu testlerin kltrlr arası lme eřdeęerlięi gstermedikleri anlařılmıřtır. Bu durum sz konusu test maddelerinin kltrlere gre deęiřen madde fonksiyonu gsterme eęilimi tařıdıklarını aęrıřtırmaktadır.

3.4. Alt Problem 2a.'ya İlişkin Bulgular ve Yorumlar:

Araştırmanın ikinci alt probleminin a maddesinde belirtilen problemlerin çözümü için PIRLS projesi kapsamında Türkiye ve Amerika Birleşik Devletlerinde uygulanan üç, altı, sekiz ve sıfır kodlu alt testleri oluşturan 8 okuma parçasına ait testler için ayrı ayrı parametre karşılaştırma yöntemi ile değişen madde fonksiyonu analizi yapıldı bu testlerdeki maddelerin kültürler arasında (Türkiye ve Amerika Birleşik Devletlerindeki uygulamalarında) değişen madde fonksiyonu gösterip göstermediği incelendi. Bu nedenle ikinci alt problemin a maddesinde belirtilen problemin çözümü için parametre karşılaştırma yöntemi ile yapılan değişen madde fonksiyonu analizlerine ait bulgular okuma parçalarına göre sırası ile aşağıda verilmiştir.

Flowers okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında parametre karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Tablo 3.13'de görüldüğü gibi Flowers okuma parçasında 13 adet değerlendirme sorusu vardır. Bu sorulardan 10 tanesi iki kategorili, 3 tanesi ise kısmi puanlama yapılarak puanlanmıştır. Parametre karşılaştırma yöntemi ile ABD ve Türkiye örneklemeleri üzerinden yapılan DMF analizi sonuçları Tablo 3.13'de verilmiştir.

Tabloda görüldüğü gibi 5, 10 ve 13 maddelerin 0,01 manidarlık düzeyinde DMF gösteren maddeler olduğu gözlenmiştir. 10 ve 13. maddelerin ABD örnekleme lehine DMF gösterdikleri yani söz konusu maddelerin aynı yetenek düzeyinde olan ABD örneklemindeki öğrenciler için daha kolay Türkiye örneklemindeki öğrenciler için ise daha zor maddeler olarak işlev gösterdikleri ortaya çıkmaktadır. Testin 5'ci maddesi ise Türkiye örneklemindeki öğrencilerin lehine DMF göstermiştir. Değişen madde fonksiyonu gösteren maddeler incelendiğinde 3 maddenin de 0-1 puanlama yapılan çoktan seçmeli maddeler olduğu görülmektedir.

Tablo 3.13

Flowers Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular

Madde No	Madde Güçlük Değerleri		Fark (Contrast)	χ^2 (sd=1)	p
	ABD örnekleme	Türkiye örnekleme			
1 (0-1)	-2,72	-2,54	0,17	0,43	0,51
2 (0-1)	-2,04	-2,47	-0,37	2,72	0,09
3 (0-1)	-2,50	-1,84	0,66	7,05	0,08
4 (0-1)	-2,45	-2,77	-0,31	1,58	0,20
5 (0-1)	-0,90	-3,62	2,72	171,32**	0,00
6 (0-1)	-1,28	-1,52	-0,24	1,37	0,23
7 (0-1-2)	-0,07	-0,16	-0,08	0,34	0,56
8 (0-1)	-1,31	-1,50	0,12	0,73	0,39
9 (0-1-2)	-1,63	-1,51	2,03	0,52	0,47
10 (0-1)	-3,88	-1,84	-2,03	34,00**	0,00
11 (0-1)	-0,93	-1,02	-0,09	0,22	0,64
12 (0-1-2)	0,05	0,26	0,21	1,90	0,16
13 (0-1)	-2,17	-1,36	-0,804	11,83**	0,00
TOPLAM				234,05**	0,00

*p<0,05 ; **p<0,01

Not: parantez içindeki değerler maddenin puanlanma biçimini göstermektedir.

Antarctica okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında parametre karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Antarctica okuma parçası için parametre karşılaştırma yöntemi ile ABD ve Türkiye örneklemeleri üzerinden yapılan DMF analizi sonuçları Tablo 3.14’de verilmiştir.

Tablo 3.14’de görüldüğü gibi Antarctica okuma parçasında 11 adet soru vardır. Bu sorulardan 7 tanesi iki kategorili, 3 tanesi ise 0-1-2 şeklinde kısmi puanlanan ve 1 tanesi de 0-1-2-3 şeklinde kısmi puanlanma yapılan maddelerdir.

Parametre karşılaştırma yöntemi ile ABD ve Türkiye örneklemeleri üzerinden yapılan DMF analizi sonucunda 1,2,3,4,7,8,9 ve 10 maddelerin DMF gösteren maddeler olduğu gözlenmiştir. Bu maddelerden 3,4,7,8,9 ve 10. maddelerin 0,01 anlamlılık düzeyinde DMF gösteren maddeler olduğu anlaşılmaktadır.

Tablo 3.14

Antarctica Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular

Madde No	Madde Güçlük Değerleri		Fark (Contrast)	χ^2 (sd=1)	p
	ABD örnekleme	Türkiye örnekleme			
1 (0-1)	-1,68	-2,11	-0,42	4,15*	0,03
2 (0-1)	-0,68	-0,20	0,48	4,34*	0,03
3 (0-1)	-1,16	-1,98	-0,82	12,85**	0,00
4 (0-1-2)	-0,48	-0,04	0,44	9,21**	0,00
5 (0-1)	-1,88	-1,58	0,29	1,83	0,17
6 (0-1)	-2,54	-2,12	0,42	2,00	0,15
7 (0-1-2-3)	-0,51	-0,84	-0,33	12,21**	0,00
8 (0-1-2)	-0,92	-1,53	-0,69	13,75**	0,00
9 (0-1-2)	-0,49	0,14	0,64	17,56**	0,00
10 (0-1)	-0,21	-0,62	-0,41	9,20**	0,00
11 (0-1)	0,18	0,50	0,31	2,19	0,13
TOPLAM				88,33**	0,00

*p<0,05 ; **p<0,01

Not: parantez içindeki değerler maddenin puanlanma biçimini göstermektedir.

Değişen madde fonksiyonu gösteren maddeler incelendiğinde 1,2,3 ve 10. maddelerin de 0-1 puanlama yapılan çoktan seçmeli maddeler olduğu 4,7,8 ve 9. maddelerin ise kısmi puanlama yapılan maddeler olduğu gözlenmiştir. Puan kategorilerine göre cevap oranları ülkelere göre karşılaştırıldığında DMF gözlenen maddelerin tamamının ABD örnekleme lehine avantaj sağladığı gözlenmektedir. Örneğin; 6 madde 0-1-2 şeklinde puanlanan bir maddedir. Bu maddenin puan kategorilerinde düşük puandan yüksek puana doğru ABD örnekleminde sıralama % 7,9 , % 68,3 ve % 23,8 olarak gözlenirken aynı sıralama Türkiye örnekleme için % 38,5 , % 50,8 ve % 10,6 olmuştur.

İki kategorili olarak 0-1 şeklinde puanlanan 2. maddeyi ise ABD örnekleminde % 62,5 doğru cevaplanırken Türkiye örnekleminde doğru cevaplayanların oranı sadece % 40,4 olarak gözlenmiştir. DMF gözlenen tüm maddelerde Türkiye örneklemindeki öğrencilerin daha yüksek θ yetenek seviyesi gerektiren kategorilere geçmede ABD örneklemindeki öğrencilere göre daha fazla zorlandıkları ortaya çıkmıştır.

Mice okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında parametre karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Mice okuma parçası için parametre karşılaştırma yöntemi ile ABD ve Türkiye örneklemeleri üzerinden yapılan DMF analizi sonuçları Tablo 3.15’de verilmiştir.

Tablo 3.15

Mice Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular

Madde No	Madde Güçlük Değerleri		Fark (Contrast)	χ^2 (sd=1)	p
	ABD Örnekleme	Türkiye Örnekleme			
1 (0-1)	-1,36	-1,71	-0,35	4,89*	0,02
2 (0-1)	-1,97	-1,80	0,17	0,71	0,40
3 (0-1)	-0,90	-0,85	0,04	0,11	0,73
4 (0-1)	0,56	0,34	0,29	4,17*	0,03
5 (0-1)	-1,44	-1,48	-0,40	0,06	0,79
6 (0,1,2)	-0,82	-1,10	-0,28	6,02*	0,01
7 (0-1)	-1,05	-1,23	-0,17	1,48	0,22
8 (0-1)	-0,38	0,30	-0,68	24,48**	0,00
9 (0-1)	-1,31	-1,30	0,01	0,00	0,90
10 (0-1)	-1,41	-2,07	-0,66	17,22**	0,00
11 (0-1)	-0,03	0,02	0,57	0,16	0,68
12 (0,1,2,3)	0,07	-0,45	-0,53	27,35**	0,00
13 (0,1)	-1,16	-0,88	0,28	3,87*	0,04
14 (0,1)	-1,11	-0,61	0,14	12,12**	0,00
TOPLAM				102,02**	0,00

*p<0,05 ; **p<0,01

Not: parantez içindeki değerler maddenin puanlanma biçimini göstermektedir.

Mice okuma parçasında 14 soru vardır. Bu sorulardan 12 tanesi iki kategorili, 2 tanesi ise kısmi puanlama yapılarak puanlanan maddelerdir. Parametre karşılaştırma yöntemi ile ABD ve Türkiye örneklemeleri üzerinden yapılan DMF analizi sonucunda 1,4,6,8,10,12,13 ve 14 maddelerin DMF gösteren maddeler olduğu gözlenmiştir. Bunlardan 8,10,12 ve 14. maddeler 0,01 anlamlılık düzeyinde DMF göstermektedir. Puan kategorilerine göre cevap oranları ülkelere göre karşılaştırıldığında DMF gözlenen maddelerin tamamının ABD örnekleme lehine olduğu gözlenmiştir. Örneğin; 6 madde 0-1-2 şeklinde kısmi puanlanan bir maddedir. Bu maddenin puan kategorilerine göre ABD örnekleminde sıralama % 12, %24,7 ve %63,3 olarak gözlenirken aynı sıralama Türkiye örnekleme için % 36, % 16,6 ve %47,4 olmuştur. İki kategorili olarak

puanlanan 8. maddeyi ise ABD örnekleminde % 79,5 doğru cevap alınırken Türkiye örnekleminde doğru cevaplayanların oranı sadece % 38,6 olarak gözlenmiştir. DMF gözlenen tüm maddelerde Türkiye örneklemindeki öğrencilerin yüksek θ yetenek seviyesi gerektiren cevapları vermekte ABD örneklemindeki öğrencilere oranla daha fazla zorlandıkları ortaya çıkmıştır.

River okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında parametre karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

River okuma parçası için parametre karşılaştırma yöntemi ile ABD ve Türkiye örneklemeleri üzerinden yapılan DMF analizi sonuçları Tablo 3.16'da verilmiştir.

Tablo 3.16

River Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular

Madde No	Madde Güçlük Değerleri		Fark (Contrast)	χ^2 (sd=1)	p
	ABD örneklemini	Türkiye örneklemini			
1 (0-1)	-0,35	0,53	0,89	43,61**	0,00
2 (0-1)	0,22	0,07	-0,14	1,03	0,31
3 (0-1)	-1,39	-1,47	-0,08	0,22	0,64
4 (0-1)	-1,34	-0,78	0,55	11,10**	0,00
5 (0-1)	-1,01	-0,76	0,25	2,60	0,10
6 (0-1-2)	0,05	-0,00	-0,06	0,25	0,61
7 (0-1)	0,25	-0,05	-0,30	4,67*	0,02
8 (0-1-2)	0,27	0,23	-0,03	0,10	0,74
9 (0-1)	0,42	-0,44	-0,02	0,02	0,84
10 (0-1-2-3)	0,24	-0,55	-0,79	44,16**	0,00
11 (0-1-2-3)	0,17	-0,06	-0,24	4,15*	0,03
TOPLAM				102,96**	0,00

*p<0,05 ; **p<0,01

Not: parantez içindeki değerler maddenin puanlanma biçimini göstermektedir.

River okuma parçasında 11 adet değerlendirme sorusu vardır. Bu sorulardan 7 tanesi iki kategorili puanlanan diğer 2 tanesi ise 0-1-2 şeklinde kısmi puanlanan ve 2 tanesi de 0-1-2-3 şeklinde kısmi puanlanma yapılan maddelerdir. Parametre karşılaştırma yöntemi ile ABD ve Türkiye örneklemeleri üzerinden yapılan DMF analizi sonucunda 1,4,7,10 ve

11. maddelerin DMF gösteren maddeler olduğu gözlenmiştir. Bunlardan 1,4 ve 10. maddelerin 0,01 anlamlılık düzeyinde DMF içerdiği görülmüştür.

Puan kategorilerine göre cevap oranları ülkelere göre karşılaştırıldığında DMF gözlenen maddelerin tamamının ABD örneklemini lehine olduğu gözlenmiştir. Örneğin; 11 madde 0-1-2-3 şeklinde puanlanan bir maddedir. Bu maddenin puan kategorilerine göre ABD örnekleminde sıralama %14, %25,3 , %45,1 ve %15,6 olarak gözlenirken aynı sıralama Türkiye örneklemini için %51,2, % 18 , %19,9 ve %10,9 olmuştur. Açıkça görüldüğü gibi bu soruya yanlış cevap verenlerin oranı ABD örnekleminde %14 iken Türkiye örnekleminde yanlış cevap oranı %51,2 olarak gözlenmiştir.

İki kategorili olarak 0-1 şeklinde puanlanan 1. maddeyi ise ABD örnekleminde % 69,5 doğru cevap alınırken Türkiye örnekleminde doğru cevaplayanların oranı sadece % 22,7 olarak gözlenmiştir. DMF gözlenen tüm maddelerde Türkiye örneklemindeki öğrencilerin daha yüksek θ yetenek seviyesi gerektiren kategorilere geçmede ABD örneklemindeki öğrencilere oranla daha fazla zorlandıkları ortaya çıkmıştır.

Clay okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında parametre karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Clay okuma parçası için parametre karşılaştırma yöntemi ile ABD ve Türkiye örneklemeleri üzerinden yapılan DMF analizi sonuçları Tablo 3.17’de verilmiştir.

Clay okuma parçasında 13 soru vardır. Bu sorulardan 10 tanesi iki kategorili puanlanan diğer 2 tanesi ise 0-1-2 şeklinde kısmi puanlanan ve 1 tanesi de 0-1-2-3 şeklinde kısmi puanlanma yapılan maddelerdir. Parametre karşılaştırma yöntemi ile ABD ve Türkiye örneklemeleri üzerinden yapılan DMF analizi sonucunda 1,2,4,6,8 ve 11. maddelerin DMF içeren maddeler olduğu gözlenmiştir. Bu maddelerden 2,4,6 ve 8. maddelerin 0,01 anlamlılık düzeyinde DMF gösteren maddeler olduğu görülmektedir. Puan kategorilerine göre cevap oranları ülkelere göre karşılaştırıldığında DMF gözlenen maddelerin tamamının ABD örneklemini lehine avantaj sağladığı gözlenmektedir.

Tablo 3.17

Clay Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular

Madde No	Madde Güçlük Değerleri		Fark (Contrast)	χ^2 (sd=1)	p
	ABD örnekleme	Türkiye örnekleme			
1 (0-1)	-0,74	-0,37	0,36	5,00*	0,02
2 (0-1)	0,15	-0,50	-0,66	20,24**	0,00
3 (0-1)	-1,15	-1,07	0,07	0,22	0,64
4 (0-1)	-0,56	-0,15	0,40	7,16**	0,00
5 (0-1)	-1,19	-1,00	0,19	1,40	0,23
6 (0-1)	-0,51	0,11	0,63	15,29**	0,00
7 (0-1)	-1,02	-1,14	-0,12	0,62	0,43
8 (0-1-2)	0,05	-0,44	-0,50	20,59**	0,00
9 (0-1)	-0,02	-0,00	0,02	0,02	0,85
10 (0-1-2-3)	-0,10	-0,27	-0,17	3,12	0,07
11 (0-1-2)	-0,17	-0,48	-0,30	6,06*	0,01
12 (0,1)	-0,98	-0,74	0,24	2,54	0,10
13 (0,1)	-0,41	-0,58	-0,16	1,31	0,24
TOPLAM				83,645**	0,00

*p<0,05 ; **p<0,01

Not: parantez içindeki değerler maddenin puanlanma biçimini göstermektedir.

DMF gözlenen 11 madde 0-1-2 şeklinde kısmi puanlanan bir maddedir. Bu maddenin puan kategorilerine göre ABD örnekleminde sıralama %20,2, %47,7 ve %32,1 olarak gözlenirken aynı sıralama Türkiye örnekleme için %56,2, %23,2 ve %20 olmuştur. Açıkça görüldüğü gibi bu soruya yanlış cevap verenlerin oranı ABD örnekleminde %20,2 iken Türkiye örnekleminde yanlış cevap oranı %56,2 olarak gözlenmiştir. İki kategorili olarak 0-1 şeklinde puanlanan 6. maddeyi ise ABD örnekleminde % 79,3 doğru cevap alınırken Türkiye örnekleminde doğru cevaplayanların oranı sadece % 20,7 olarak gözlenmiştir. DMF gözlenen tüm maddelerde DMF gözlenen maddelerin tümünde aynı yetenek düzeyinde fakat farklı örneklem grubunda bulunan öğrencilerin maddeleri doğru cevaplama olasılıklarının farklılaştığı gözlenmektedir. ABD örneklemindeki öğrencilerin daha yüksek θ yetenek seviyesi gerektiren kategorilere Türkiye örneklemindeki öğrencilere oranla daha kolay geçebildikleri görülmüştür.

Leonardo okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında parametre karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Analiz sonuçları Tablo 3.18’de verilmektedir.

Tablo 3.18

Leonardo Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular

Madde No	Madde Güçlük Değerleri		Fark (Contrast)	χ^2 (sd=1)	p
	ABD örnekleme	Türkiye örnekleme			
1 (0-1)	-1,32	-1,81	-0,48	8,05**	0,00
2 (0-1)	-0,10	-0,16	-0,60	0,16	0,68
3 (0-1)	-0,36	-0,44	-0,08	0,34	0,56
4 (0-1-2-3)	0,36	-0,45	0,81	47,05**	0,00
5 (0-1)	-0,82	-0,01	0,80	29,09*	0,01
6 (0-1-2)	-0,17	0,17	0,34	5,69*	0,01
7 (0-1)	-0,24	0,01	0,25	3,07	0,07
8 (0-1-2)	0,15	0,19	0,03	0,07	0,00
9 (0-1)	-0,82	-1,39	-0,57	15,35	0,00
10 (0-1-2)	0,26	0,23	-0,02	0,03*	0,82
11 (0-1)	-0,86	-0,27	0,58	15,37**	0,00
12 (0,1,2)	0,13	0,16	0,02	0,03	0,83
TOPLAM				124,80**	0,00

*p<0,05 ; **p<0,01

Not: parantez içindeki değerler maddenin puanlanma biçimini göstermektedir.

Tablo 3.18’de görüldüğü gibi Leonardo okuma parçasında 12 soru vardır. Bu sorulardan 7 tanesi iki kategorili puanlanan diğer 4 tanesi ise 0-1-2 şeklinde kısmi puanlanan ve 1 tanesi de 0-1-2-3 şeklinde kısmi puanlanma yapılan maddelerdir. Parametre karşılaştırma yöntemi ile ABD ve Türkiye örneklemeleri üzerinden yapılan DMF analizi sonucunda 1,4,5,6,10 ve 11. maddelerin DMF gösteren maddeler olduğu gözlenmiştir. Bunlardan 1,4 ve 11. maddeler 0,01 anlamlılık düzeyinde DMF göstermiştir. Puan kategorilerine göre cevap oranları ülkelere göre karşılaştırıldığında DMF gözlenen maddelerin tamamının ABD örnekleme lehine olduğu gözlenmektedir. Örneğin; 10 madde 0-1-2 şeklinde puanlanan bir maddedir. Bu maddenin puan kategorilerine göre ABD örnekleminde sıralama %46,7 %23,4 ve %29,9 olarak gözlenirken aynı sıralama Türkiye örnekleme için %64,4, %26,9 ve %11,3 olmuştur.

Açıkça görüldüğü gibi bu soruya yanlış cevap verenlerin oranı ABD örnekleminde %46,7 iken Türkiye örnekleminde yanlış cevap oranı %64,4 olarak gözlenmiştir. İki kategorili olarak 0-1 şeklinde puanlanan 11. maddeyi ise ABD örnekleminde % 76,0 doğru cevap alınırken Türkiye örnekleminde doğru cevaplayanların oranı sadece % 14 olarak gözlenmiştir.

DMF içeren tüm maddelerde aynı yetenek düzeyinde fakat farklı örneklem grubunda bulunan öğrencilerin maddeleri doğru cevaplama olasılıklarının farklılaştığı gözlenmektedir. ABD örneklemindeki öğrencilerin daha yüksek θ yetenek seviyesi gerektiren kategorilere cevaplara ulaşmada Türkiye örneklemindeki öğrencilere oranla daha avantajlı oldukları görülmüştür.

Hare okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında parametre karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Hare okuma parçasında 11 soru vardır. Bu sorulardan 7 tanesi iki kategorili puanlanan diğer 3 tanesi 0-1-2 şeklinde kısmi puanlanan ve 1 tanesi de 0-1-2-3 şeklinde kısmi puanlanma yapılan maddelerdir. Analiz sonuçları Tablo 3.19’da verilmiştir.

Parametre karşılaştırma yöntemi ile ABD ve Türkiye örneklemeleri üzerinden yapılan DMF analizi sonucunda 2,3,5,6,7,8 ve 11. maddelerin DMF gösteren maddeler olduğu gözlenmiştir. Bunlardan 2,3,5,7,8 ve 11 maddelerin 0,01 anlamlılık düzeyinde DMF içerdiği gözlenmiştir.

Puan kategorilerine göre cevap oranları ülkelere göre karşılaştırıldığında DMF gözlenen maddelerin tamamının ABD örneklemini lehine avantaj sağladığı gözlenmektedir. Örneğin; 7 madde 0-1-2 şeklinde puanlanan bir maddedir. Bu maddenin puan kategorilerine göre ABD örnekleminde sıralama %24,5 %31,6 ve %43,8 olarak gözlenirken aynı sıralama Türkiye örneklemini için %30,7 , %27,4 ve %41,9 olmuştur. Açıkça görüldüğü gibi bu soruya yanlış cevap verenlerin oranı ABD örnekleminde %46,7 iken Türkiye örnekleminde yanlış cevap oranı %64,4 olarak gözlenmiştir. İki kategorili olarak 0-1 şeklinde puanlanan 2. maddeyi ise ABD örnekleminde % 91 doğru

cevap alınırken Türkiye örnekleminde doğru cevaplayanların oranı sadece % 68 olarak gözlenmiştir.

Tablo 3.19

Hare Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular

Madde No	Madde Güçlük Değerleri		Fark (Contrast)	χ^2 (sd=1)	p
	ABD örnekleme	Türkiye örnekleme			
1 (0-1)	-1,84	-1,98	-0,14	1,39	0,23
2 (0-1)	-1,63	-0,94	0,68	38,06**	0,00
3 (0-1-2)	0,29	0,06	-0,22	8,29**	0,00
4 (0-1)	-1,37	-1,39	-0,01	0,03	0,83
5 (0-1)	-1,56	-1,21	0,35	10,54**	0,00
6 (0-1)	-1,71	-0,89	-0,18	3,95*	0,04
7 (0-1-2)	-0,24	-0,53	-0,28	14,75**	0,00
8 (0-1)	-0,49	-0,13	0,36	16,20**	0,00
9 (0-1-2)	-0,23	-0,30	-0,06	0,81	0,37
10 (0-1,2-3)	0,43	0,33	-0,10	2,28	0,12
11 (0-1)	-0,69	-1,05	-3,36	16,47**	0,00
TOPLAM				112,83**	0,00

*p<0,05 ; **p<0,01

Not: parantez içindeki değerler maddenin puanlanma biçimini göstermektedir.

DMF gözlenen maddelerin tümünde aynı yetenek düzeyinde fakat farklı örneklem grubunda bulunan öğrencilerin maddeleri doğru cevaplama olasılıklarının farklılaştığı gözlenmektedir. ABD örneklemindeki öğrencilerin daha yüksek θ yetenek seviyesi gerektiren kategorilere geçmede Türkiye örneklemindeki öğrencilere oranla daha rahat oldukları görülmüştür.

Pufflings okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında parametre karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Analiz sonuçları Tablo 3.20’de verilmektedir.

Tablo 3.20’de görüldüğü gibi Pufflings okuma parçasında 13 adet değerlendirme sorusu vardır. Bu sorulardan 10 tanesi 0-1 şeklinde puanlanan ve 3 tanesi ise 0-1-2 şeklinde kısmi puanlanma yapılan maddelerdir. Parametre karşılaştırma yöntemi ile ABD ve

Türkiye örneklemi üzerinden yapılan DMF analizi sonucunda 1,2,4,6,7,9,10,11,12 ve 13. maddelerin DMF gösteren maddeler olduğu gözlenmiştir. Bunlardan 1,4,7,10,11,12 ve 13. maddelerin 0,01 anlamlılık düzeyinde kültürler arasında DMF gösterdiği anlaşılmıştır.

Tablo 3.20

Pufflings Okuma Parçası Değerlendirme Soruları İçin ABD ve Türkiye Örneklem Gruplarında Parametre Karşılaştırma Yöntemi İle Yapılan DMF Analizine Ait Bulgular

Madde No	Madde Güçlük Değerleri		Fark (Contrast)	χ^2 (sd=1)	p
	ABD örnekleme	Türkiye örnekleme			
1 (0-1)	-0,73	-1,27	-0,53	40,05**	0,00
2 (0-1)	-0,30	-0,48	-0,17	4,45*	0,03
3 (0-1)	-1,20	-1,30	-0,09	1,15	0,28
4 (0-1)	0,02	0,25	0,22	6,92**	0,00
5 (0-1)	-0,25	-0,37	-0,12	2,24	0,13
6 (0-1)	0,05	0,24	0,19	5,07*	0,02
7 (0-1-2)	0,39	0,10	-0,29	16,33**	0,00
8 (0-1-2)	-0,13	-0,19	-0,05	0,54	0,46
9 (0-1)	-0,91	-0,71	0,19	5,54*	0,01
10 (0-1)	0,91	0,48	0,42	23,00**	0,00
11 (0-1)	0,05	-0,43	-0,49	37,73**	0,00
12 (0,1,2)	-0,35	0,65	1,00	187,54**	0,00
13 (0,1)	-0,50	0,08	0,58	47,81**	0,00
TOPLAM				378,44**	0,00

*p<0,05 ; **p<0,01

Not: parantez içindeki değerler maddenin puanlanma biçimini göstermektedir.

Puan kategorilerine göre cevap oranları ülkelere göre karşılaştırıldığında DMF gözlenen maddelerin tamamının ABD örnekleme lehine olduğu gözlenmiştir. Örneğin; 12 madde 0-1-2 şeklinde kısmi puanlama yapılan bir maddedir. Bu maddenin puan kategorilerine göre cevap sıralaması ABD örnekleminde %26 , %20,6 ve %53,3 olarak gözlenirken aynı sıralama Türkiye örnekleme için %79,7 , %14,1 ve %6,2 olmuştur. Açıkça görüldüğü gibi bu soruya yanlış cevap verenlerin oranı ABD örnekleminde %26 iken Türkiye örnekleminde yanlış cevap oranı %79,7 olarak gözlenmiştir. 12. soru için en doğru cevap kategorisine çıkma oranı ABD örnekleme için %53,3 olarak gerçekleşirken Türkiye örnekleme için bu oran sadece %6,2 olmuştur. İki kategorili olarak 0-1 şeklinde puanlanan 13. maddeyi ise ABD örnekleminde %66,1 doğru cevap alınırken Türkiye örnekleminde doğru cevaplayanların oranı %28,8 olarak gözlenmiştir.

Pufflings okuma parçasına ait değerlendirme testinde DMF gözlenen tüm maddelerde aynı yetenek düzeyinde fakat farklı örneklem grubunda bulunan öğrencilerin maddeleri doğru cevaplama olasılıkları arasındaki farkın anlamlı olduğu gözlenmektedir. ABD örneklemindeki öğrencilerin daha yüksek θ yetenek seviyesi gerektiren kategorilere Türkiye örneklemindeki öğrencilere göre daha kolay geçebildikleri görülmüştür.

PIRLS 2001 projesi kapsamında bulunan sekiz okuma parçasına ait sorular için ABD ve Türkiye örneklem grupları arasında parametre karşılaştırma yöntemi kullanılarak yapılan DMF analizi sonuçları Tablo 3.21’de özetlenmiştir.

Tablo 3.21
Okuma Parçalarına Göre Parametre Karşılaştırma
Yöntemi İle Yapılan DMF Analizi Sonuçları

Testler	Testteki Madde Sayısı	0,05 Düzeyinde DMF Gözlenen Madde Sayısı	0,01 Düzeyinde DMF Gözlenen Madde Sayısı
Flowers	13	-	3
Antarctica	11	2	6
Mice	14	4	4
River	11	2	3
Clay	13	2	4
Leonardo	12	3	3
Hare	11	1	6
Pufflings	13	3	7
Toplam	98	17	36

Tablo 3.21’de görüldüğü gibi PIRLS 2001 projesi kapsamında uygulanan sekiz okuma parçasına ait değerlendirme sorularının toplam sayısı 98 dir. Bu 98 maddenin 17 tanesi (%17,35) 0,05 anlamlılık düzeyinde kültürler arasında değişen madde fonksiyonu göstermiştir. Manidarlık düzeyi 0,01 olarak alındığında ise 36 (%36,73) maddenin kültürler arası uygulama sonucunda DMF içerdiği gözlenmiştir.

3.4. Alt Problem 2b.'ye İlişkin Bulgular ve Yorumlar:

DMF analizi için karşılaştırma yapılan grupların parametrelerini aynı metrik üzerinde eşitlemek üzere iç kriter olarak testin kendi maddelerinden oluşan bağ maddelerini belirlemek için önce tekrarlayıcı (iterative) madde ayıklama işlemleri yapıldı. Birinci iterasyonda örneklem gruplarına ait birleştirilmiş veri setinde önce tüm maddelerin madde güçlük parametrelerinin gruplarda eşit tutulduğu dar model oluşturuldu ve dar modelin $-2\log L$ değeri elde edildi. Daha sonra, sırası ile her bir maddenin yerleşim parametresi tahminlerinin karşılaştırma yapılan gruplarda serbest bırakıldığı geniş modeller kuruldu (madde sayısı kadar yeni model kuruldu) ve kurulan her bir geniş model için $-2\log L$ değerleri elde edildi. Söz konusu işlemlere ait bulgular hazırlanan bir tablo üzerinden verilmektedir. Tablonun en alt satırında verilen $-2\log L$ değeri tüm maddelerin yerleşim (güçlük) parametrelerinin referans ve odak gruplarda eşitlendiği dar modelden elde edilen $-2\log L$ değeridir. Tabloda b_i sütunu iki kategorili puanlanan maddelerin madde yerleşim parametrelerini göstermektedir, c_{i1} ve c_{i2} sütunları ise üç kategorili puanlanan maddelere için sınır tepki fonksiyonlarının yerleşim parametrelerine aittir ve bu parametrelerde basamak güçlük indeksi olarak yorumlanmaktadır. Dört kategorili olarak puanlanan bir maddede üç adet basamak bulunur (puan kategori sayısından bir eksik) ve bireyler en doğru cevaba ulaşmak için öncelikle ilk iki basamağı aşmak zorundadır. Tabloda verilen $-2\log L$ sütunu ise her bir madde için parametrelerinin karşılaştırma yapılan gruplarda serbest bırakıldığı geniş modele ait $-2\log L$ değerini göstermektedir. Testteki madde sayısı kadar geniş model kurulduğundan madde sayısı kadar geniş modele ait $-2\log L$ değeri elde edilmiştir. G_1^2 - G_2^2 sütunu ise her bir maddenin tahmin edilen parametresinin gruplar arasında serbest bırakıldığı geniş modele ait $-2\log L$ değerinin dar modelden elde edilen $-2\log L$ değerinden çıkarılması ile elde edilen G_i^2 fark değerlerini vermektedir. Bu değerler ilgili serbestlik derecesinde χ^2 tablo değeri ile karşılaştırıldı ve modeller arasında uyum farkı gözlenen maddeler 0,05 ve 0,01 anlamlılık düzeyi için belirlendi. Madde ayırıcılığı değeri tüm maddeler için 1'e eşitlenmiştir. Yukarıda anlatılan işlemler her bir okuma parçası için aynı sıra ile tekrarlanmıştır.

Flowers okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında olabilirlik oranı testi model karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Flowers okuma parçası değerlendirme testi değişen madde fonksiyonu analizinde bağ maddelerinin seçimi için yapılan I. tekrarlayıcı madde seçme işlemi sonuçları Tablo 3.22'de verilmiştir.

Tablo 3.22

Flowers Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları

Madde no	Geniş Model							I. İterasyon
	Referans Grup (ABD)			Odak Grup (Türkiye)				
	b_i	c_{i1}	c_{i2}	b_i	c_{i1}	c_{i2}	$-2\log L$	$G_1^2 - G_2^2$
1	-0,74			-0,65			3163,1	0,4
2	-1,08			-0,88			3162,8	0,7
3	-1,46			-0,50			3145,4	18,1**
4	-0,85			-0,96			3163,1	0,4
5	-0,03			-1,96			3101,1	62,4**
6	-0,21			-0,10			3163,1	0,4
7		-1,29	-1,35		-2,04	-1,49	3155,4	8,1*
8	-0,13			-0,23			3163,0	0,1
9		0,18	1,15		0,18	0,20	3142,3	21,2**
10	-1,74			-0,30			3090,4	73,1**
11	-0,42			0,31			3157,0	6,5
12		-1,07	-1,37		-1,07	-2,75	3138,2	25,3**
13	-0,79			0,00			3145,0	18,5**
Dar Model $-2\log L = 3163,5$								

$$\chi^2_{(0,01)}=11,34, \text{sd.}=3, **p<0,01$$

$$\chi^2_{(0,05)}=7,81, \text{sd.}=3, *p<0,05$$

I. tekrarlayıcı madde seçme işlemi sonucunda 3,5,9,10,12 ve 13. maddelerin dar ve geniş modelleri arasındaki fark değerleri 0,01 anlamlılık düzeyinde, 7. maddeye ait fark değeri ise 0,05 düzeyinde manidar bulunmuştur. Bu maddeler değişen madde

fonksiyonu göstermeye aday maddelerdir. En yüksek G_i^2 fark değeri veren 10. madde ($\chi^2_{(3; 0,01)} = 73,1$, $p < 0,01$) testten çıkarılarak kalan 12 madde için II. tekrarlayıcı madde seçme işlemine geçildi. II. tekrarlayıcı madde seçme işlemi sonucunda 7 ve 11. maddelerin dar ve geniş modelleri arasındaki fark değerleri 0,05 anlamlılık düzeyinde 3,5,9,12 ve 13. maddelerin ise 0,01 düzeyinde manidar bulunmuştur.

En yüksek G_i^2 fark değeri veren 5. madde ($\chi^2_{(3; 0,01)} = 62,3$, $p < 0,01$) testten çıkarılarak kalan 11 madde için III. tekrarlayıcı madde seçme işlemine geçildi. III. madde seçme işlemi esnasında teste kalan 11 madde üzerinden dar ve geniş modellerin madde parametre tahminleri ve $-2\log L$ değerleri yeniden tahmin edilmiştir. III. tekrarlayıcı madde seçme işlemi sonucunda 3,9,12 ve 13. maddelerin dar ve geniş modelleri arasındaki fark değerleri 0,01 anlamlılık düzeyinde manidar bulunmuştur. En yüksek G_i^2 fark değeri veren 12. madde ($\chi^2_{(3; 0,01)} = 30,1$, $p < 0,01$) testten çıkarılarak kalan 10 madde için IV. tekrarlayıcı madde seçme işlemine geçildi.

IV. tekrarlayıcı madde seçme işleminde 5, 10 ve 12. madde testten çıkarıldıktan sonra kurulan dar modelin olabilirlik oranı değerinin eksi (negatif) çıkması üzerine analizler tamamlanamadı. Bu nedenle, bağ maddeleri tamamlanan üç iterasyonda da değişen madde fonksiyonu gösterme eğilimi olmayan maddeler arasından 4,6,7 ve 11. maddeler olarak belirlendi. Tekrarlayıcı madde seçme işlemlerinin yapıldığı üç iterasyona ait bilgiler Tablo 3.23'de verilmiştir.

Tablo 3.23'de görüldüğü gibi 3,5,9,10,12 ve 13. maddelerin değişen madde fonksiyonu göstermeye aday maddeler olduğu anlaşılmaktadır. Bu maddeler için her üç iterasyonda da dar modeller ile geniş modelleri arasındaki fark istatistiki olarak anlamlı bulunmuştur. Bu analizler sonucunda 4,6,7 ve 11. maddeler bağ maddesi olarak seçilmiştir.

Tablo 3.23

Flowers Okuma Parçası Değerlendirme Testi I. , II. ve III. Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modellere Göre Karşılaştırma Sonuçları

Madde no	I. İterasyon $G_1^2 - G_2^2$	II. İterasyon $G_1^2 - G_2^2$	III. İterasyon $G_1^2 - G_2^2$
1	0,4	2,6	1,6
2	0,7	1,8	1,4
3	18,1**	23,1**	21,8**
4	0,4	0,0	0,1
5	62,4**	62,3**	Çıkarıldı
6	0,4	1,6	1,1
7	8,1*	10,7*	5,8
8	0,1	0,00	0,1
9	21,2**	29,4**	27,1**
10	73,1**	Çıkarıldı	Çıkarıldı
11	6,5	10,2*	6,1
12	25,3**	31,8**	30,1**
13	18,5**	28**	25**

$$\chi^2_{(0,01)}=11,34, \text{ sd.}=3, **p<0,01$$

$$\chi^2_{(0,05)}=7,81, \text{ sd.}=3, *p<0,05$$

Bağ maddelerini seçme işlemi tamamlandıktan sonra yeniden dar ve geniş modeller için madde parametre tahminleri ve $-2\log L$ değerleri hesaplandı. Bu nedenle oluşturulan geniş modelde sadece 4,6,7, ve 11. maddelerin referans ve odak gruplarda parametreleri eşitlendi ve diğer maddelerin parametre değerleri gruplar arasında eşitlenmedi. Daha sonra bağ maddeleri dışında kalan her bir madde için bağ maddeleri ile birlikte eşitlendikleri dar modeller kuruldu. Örneğin, 1. madde için kurulan dar modelde 4,6,7 ve 11. madde ile birlikte 1. madde de eşitlendi ve bu maddenin de yerleşim parametresi değerinin gruplar arasında değişmemesi sağlandı. Testteki diğer maddeler için de dar modelleri kurulurken aynı işlem yapıldı ve bu dar modellerin $-2\log L$ değerleri elde edildi.

Analizlerin devam eden aşamasında dar ve geniş modellerin olabilirlik oranlarının karşılaştırılması, dar modele sınırlanarak sokulan maddenin ait olduğu dar modelin olabilirlik oranı geniş modelde serbest bırakıldığı durumda elde edilen geniş modelin olabilirlik oranı değerinden çıkarılarak yapılmıştır. Bu şekilde geniş ve dar modeller

karşılaştırıldığında iki modelin uyum değerleri arasındaki farkın istatistiki olarak anlamlı olup olmadığı ortaya çıkarılmıştır. Flowers okuma parçası değerlendirme testine ait dar ve geniş modellerin $-2\log L$ değerleri ile değişen madde fonksiyonu analizi sonuçları Tablo 3.24’de verilmiştir.

Tablo 3.24
Flowers Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin $-2\log L$ Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları.

Madde no	Dar model G_1^2	Geniş model G_2^2	$G_i^2 = G_1^2 - G_2^2$
1	2819,7	2816,9	2,8
2	2820,1	2816,9	3,2
3	2822,1	2816,9	5,2
4	Bağ maddesi	Bağ maddesi	Bağ maddesi
5	2864,1	2816,9	47,2**
6	Bağ maddesi	Bağ maddesi	Bağ maddesi
7	Bağ maddesi	Bağ maddesi	Bağ maddesi
8	2838,4	2816,9	21,5**
9	2891,3	2816,9	74,4**
10	2944,7	2816,9	127,8**
11	Bağ maddesi	Bağ maddesi	Bağ maddesi
12	2882,2	2816,9	65,3**
13	2875,5	2816,9	58,6**

$$\chi^2_{(0,01)}=11,34, \text{ sd.}=3, **p<0,01$$

$$\chi^2_{(0,05)}=7,81, \text{ sd.}=3, *p<0,05$$

Değişen madde fonksiyonu analizi sonuçlarına göre Flowers okuma parçası değerlendirme testinde bulunan bağ maddesi dışındaki 5,8,9,10,12 ve 13. maddelerin dar ve geniş modelleri $-2\log L$ değerleri arasındaki fark 0,01 anlamlılık düzeyinde önemli bulunmuştur. Bu bulgunun anlamı ise söz konusu maddelerin medde yerleşim parametreleri ABD ve Türkiye örneklem gruplarında ölçme değişmezliğine sahip değildir, bu nedenle testin kaynak kültürdeki versiyonu ile hedef kültürdeki versiyonunun aynı bilişsel özellikleri ölçtüklerini söylemek yapılan analiz sonuçlarına göre olası görünmemektedir.

Antarctica okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında olabilirlik oranı testi model karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Antarctica okuma parçası değerlendirme testi değişen madde fonksiyonu analizinde bağ maddelerinin seçimi için yapılan I. tekrarlayıcı madde seçme işlemi sonuçları Tablo 3.25’de verilmiştir. Tabloda görüldüğü gibi 3,5 ve 6. maddelerin dar ve geniş modelleri arasındaki fark değerleri 0,05 anlamlılık düzeyinde 3,5,9,10,12 ve 13. maddelerin ise 0,01 düzeyinde manidar bulunmuştur.

Tablo 3.25
Antarctica Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin
I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları

Madde no	Geniş Model									I. İterasyon
	Referans Grup (ABD)				Odak Grup (Türkiye)					
	b_i	c_{i1}	c_{i2}	c_{i3}	b_i	c_{i1}	c_{i2}	c_{i3}	$-2\log L$	$G_1^2 - G_2^2$
1	-1,08				-1,22				2117,1	0,0
2	-0,31				0,26				2108,2	8,9
3	-0,66				-1,18				2107,1	10*
4		1,34	-0,58			-1,34	-0,87		2062,4	54,7**
5	-1,35				-0,87				2106,8	10,3*
6	-1,86				-1,25				2104,9	12,2*
7		0,26	0,64	0,23		0,78	1,45	1,12	2101,2	15,9**
8		1,18	0,96			1,18	2,07		2070,2	46,9**
9		0,18	-0,42			0,18	-1,20		2091,0	26,1**
10	0,17				-0,04				2101,8	15,3**
11	0,37				0,00				2095,9	21,2**
Dar Model $-2\log L = 2117,1$										

Dar Model $-2\log L = 2117,1$

$\chi^2_{(0,01)} = 13,28$, sd.=4, **p<0,01

$\chi^2_{(0,05)} = 9,49$, sd.=4, *p<0,05

En yüksek G_i^2 fark değeri veren 4. madde ($\chi^2_{(4; 0,01)} = 54,7$, p<0,01) testten çıkarılarak kalan 10 madde için II. tekrarlayıcı madde seçme işlemine geçildi. II. iterasyonda dar

modelin $-2\log L$ değeri 954,4 olarak elde edilmiştir. Bu iterasyonda en yüksek G_i^2 fark değeri veren 11'inci madde testten çıkarılarak kalan 9 madde için III. tekrarlayıcı madde seçme işlemine geçildi. III. İterasyona ait dar modelin $-2\log L$ değeri 161,3 olarak hesaplandı. IV. iterasyonda kurulan dar modelin olabilirlik oranı değerinin eksi (negatif) çıkması üzerine analizler tamamlanamadı Tekrarlayıcı madde seçme işlemlerinin yapıldığı üç iterasyona ait bilgiler Tablo 3.26'da verilmiştir.

Tablo 3.26
Antarctica Okuma Parçası Değerlendirme Testi I. , II. ve III Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modellere Göre Karşılaştırma Sonuçları

Madde no	I. İterasyon $G_1^2 - G_2^2$	II. İterasyon $G_1^2 - G_2^2$	III. İterasyon $G_1^2 - G_2^2$
1	0,0	0,3	0,1
2	8,9	9,1	7,5
3	10*	8,0	7,9
4	54,7**	Çıkarıldı	Çıkarıldı
5	10,3*	13,7*	15,1**
6	12,2*	15,2**	14,9**
7	15,9**	5,3	20,8**
8	46,9**	31,3**	65,2**
9	26,1**	16,4**	71,1**
10	15,3**	2	2,5
11	21,2**	71,2**	Çıkarıldı

$$\chi^2_{(0,01)} = 13,28, \text{sd.}=4, **p<0,01$$

$$\chi^2_{(0,05)} = 9,49, \text{sd.}=4, *p<0,05$$

Tablo 3.26'da görüldüğü gibi üç tekrarlayıcı madde seçme işlemi sonucunda her üç iterasyonda da dar model ile geniş modellerinin uyum farkı istatistiki olarak anlamlı bulunmayan 1,2,3 ve 10. maddeler değişen madde fonksiyonu gösterme eğilimi olmayan maddeler olduklarından bağ maddeleri olarak belirlendiler.

Bağ maddelerini seçme işlemi tamamlandıktan sonra yeniden dar ve geniş modeller için madde parametre tahminleri ve $-2\log L$ değerleri hesaplandı. Bu nedenle oluşturulan geniş modelde sadece 1,2,3 ve 10. maddelerin referans ve odak gruplarda parametreleri eşitlendi ve diğer maddelerin parametre değerleri gruplar arasında eşitlenmedi. Daha sonra bağ maddeleri dışında kalan her bir madde için bağ maddeleri ile birlikte

eşitlendikleri dar modeller kuruldu. Devam eden aşamada dar ve geniş modellerin olabilirlik oranlarının karşılaştırılması, dar modele sınırlanarak sokulan maddenin ait olduğu modelin olabilirlik oranı geniş modelde serbest bırakıldığı durumda elde edilen geniş modelin olabilirlik oranı değerinden çıkarılarak yapılmıştır. Bu şekilde geniş ve dar modeller karşılaştırıldığında iki modelin uyum değerleri arasındaki farkın istatistiki olarak anlamlı olup olmadığı ortaya çıkarılmıştır. Antarctica okuma parçası değerlendirme testine ait dar ve geniş modellerin $-2\log L$ değerleri ile değişen madde fonksiyonu analizi sonuçları Tablo 3.27’de verilmiştir.

Tablo 3.27
Antarctica Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin $-2\log L$ Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları

Madde no	Dar model G_1^2	Geniş model G_2^2	$G_i^2 = G_1^2 - G_2^2$
1	Bağ maddesi	Bağ maddesi	Bağ maddesi
2	Bağ maddesi	Bağ maddesi	Bağ maddesi
3	Bağ maddesi	Bağ maddesi	Bağ maddesi
4	1962,1	1877,3	84,8**
5	1910,2	1877,3	32,9**
6	1906,5	1877,3	29,2**
7	1868,9	1877,3	8,6
8	1902,9	1877,3	25,6**
9	1977,2	1877,3	99,9**
10	Bağ maddesi	Bağ maddesi	Bağ maddesi
11	1902,3	1877,3	25**

$\chi^2_{(0,01)} = 13,28$, sd.=4, **p<0,01

$\chi^2_{(0,05)} = 9,49$, sd.=4, *p<0,05

Değişen madde fonksiyonu analizi için 1,2 ve 10. maddelerin gruplar arasında eşitlendiği geniş modelin $-2\log L$ değeri 1877,3 olarak elde edilmiştir. Analiz sonuçlarına göre Antarctica okuma parçası değerlendirme testinde bulunan bağ maddesi dışındaki 4,5,6,8,9 ve 11. maddelerin dar ve geniş modelleri $-2\log L$ değerleri arasındaki fark 0,01 anlamlılık düzeyinde önemli bulunmuştur. Bu bulgunun anlamı ise çevirisi ve adaptasyonu yapılan Antarctica okuma parçası değerlendirme testinin kültürler arası uygulamalarında değişen madde fonksiyonu gözlenmiştir ve söz konusu maddelerin madde yerleşim parametreleri ABD ve Türkiye örneklem gruplarında ölçme

değişmezliğine sahip değildir. Bu nedenle testin kaynak kültürdeki versiyonu ile hedef kültürdeki versiyonunun aynı bilişsel özellikleri ölçtüklerini ya da aynı değerde ölçtüklerini söylemek mümkün görünmemektedir.

Mice okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında olabilirlik oranı testi model karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Mice okuma parçası değerlendirme testi değişen madde fonksiyonu analizinde bağ maddelerinin seçimi için yapılan I. tekrarlayıcı madde seçme işlemine ait sonuçlar Tablo 3.28’de verilmiştir. Analiz sonucunda 10,11 ve 13. maddelerin dar ve geniş modelleri arasındaki fark değerleri 0,05 anlamlılık düzeyinde 4,6,8 ve 13. maddelerin ise 0,01 düzeyinde manidar bulunmuştur.

Tablo 3.28

Mice Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları

Madde no	Geniş Model									I. İterasyon
	Referans Grup (ABD)				Odak Grup (Türkiye)					
	b _i	c _{i1}	c _{i2}	c _{i3}	b _i	c _{i1}	c _{i2}	c _{i3}	-2logL	
1	-0,85				-1,00				3157,5	4
2	-1,41				-1,13				3155,1	6,4
3	-0,50				-0,22				3153,8	7,7
4	0,48				1,05				3143,9	17,6**
5	-0,98				-0,84				3157,6	3,9
6		1,11	1,52			0,04	1,20		3137,7	23,8**
7	-0,55								3158,6	2,9
8	-0,39				-0,59				3106,3	55,2**
9	-0,87				-0,70				3155,8	5,7
10	-0,76				-0,65				3149,1	12,4*
11	0,39				-1,13				3151,1	10,4*
12		0,36	-0,81	-2,15		0,39	0,08	-0,41	3120,4	41,1
13	-1,43				-0,52				3141,9	19,6*
14	-0,67				0,00				3118,2	43,3**
Dar Model -2logL = 3161,5										

Dar Model -2logL = 3161,5

$\chi^2_{(0,01)} = 13,28$, sd.=4, **p<0,01

$\chi^2_{(0,05)} = 9,49$, sd.=4, *p<0,05

En yüksek G_i^2 fark değeri veren 8. madde ($\chi^2_{(4; 0,01)} = 55,2$, $p < 0,01$) testten çıkarılarak kalan 13 madde için II. tekrarlayıcı madde seçme işlemine geçildi. II. iterasyonda dar modelin $-2\log L$ değeri 2270,2 olarak elde edildi. Bu iterasyonda en yüksek G_i^2 fark değeri veren 14. madde testten çıkarılarak kalan 12 madde için III. tekrarlayıcı madde seçme işlemine geçildi. III. İterasyona ait dar modelin $-2\log L$ değeri 1563,1 olarak hesaplandı. IV iterasyon için kurulan dar modelin olabilirlik oranı değerinin eksi (negatif) çıkması üzerine analizler tamamlanamadı. Tekrarlayıcı madde seçme işlemlerinin yapıldığı üç iterasyona ait bilgiler Tablo 3.29’da verilmiştir.

Tablo 3.29

Mice Okuma Parçası Değerlendirme Testi I., II. ve III. Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modellere Göre Karşılaştırma Sonuçları

Madde no	I. İterasyon $G_1^2 - G_2^2$	II. İterasyon $G_1^2 - G_2^2$	III. İterasyon $G_1^2 - G_2^2$
1	4	0,8	0,2
2	6,4	4,1	6,3
3	7,7	6,2	8,5
4	17,6**	18,6**	23,8**
5	3,9	1,5	3,1
6	23,8**	20,3**	21,4**
7	2,9	0	0,4
8	55,2**	Çıkarıldı	Çıkarıldı
9	5,7	3,9	6,6
10	12,4*	8,9	5,6
11	10,4*	10,1*	13,3**
12	41,1	31,6**	23,2**
13	19,6*	8,6	9,4
14	43,3**	44,8**	Çıkarıldı

$\chi^2_{(0,01)} = 13,28$, sd.=4, ** $p < 0,01$

$\chi^2_{(0,05)} = 9,49$, sd.=4, * $p < 0,05$

Tablo 3.29’da görüldüğü gibi üç tekrarlayıcı madde seçme işlemi sonucunda her üç iterasyonda da dar model ile geniş modellerinin uyum farkı istatistiki olarak anlamlı bulunmayan 3,5,9 ve 13. maddeler değişen madde fonksiyonu gösterme eğilimi olmayan maddeler olduklarından bağ maddeleri olarak belirlendiler.

Bağ maddelerini seçme işlemi tamamlandıktan sonra yeniden dar ve geniş modeller için madde parametre tahminleri ve $-2\log L$ değerleri hesaplandı. Bu nedenle oluşturulan geniş modelde sadece 3,5,9 ve 13. maddelerin referans ve odak gruplarda parametreleri eşitlendi ve diğer maddelerin parametre değerleri gruplar arasında eşitlenmedi. Daha sonra bağ maddeleri dışında kalan her bir madde için bağ maddeleri ile birlikte eşitlendikleri dar modeller kuruldu. Devam eden aşamada dar ve geniş modellerin olabilirlik oranlarının karşılaştırılması, dar modele sınırlanarak sokulan maddenin ait olduğu modelin olabilirlik oranı geniş modelde serbest bırakıldığı durumda elde edilen geniş modelin olabilirlik oranı değerinden çıkarılarak yapılmıştır. Bu şekilde geniş ve dar modeller karşılaştırıldığında iki modelin uyum değerleri arasındaki farkın istatistiki olarak anlamlı olup olmadığı ortaya çıkarılmıştır. Mice okuma parçası değerlendirme testine ait dar ve geniş modellerin $-2\log L$ değerleri ile değişen madde fonksiyonu analizi sonuçları Tablo 3.30'da verilmiştir.

Tablo 3.30

Mice Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin $-2\log L$ Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları

Madde no	Dar model G_1^2	Geniş model G_2^2	$G_i^2 = G_1^2 - G_2^2$
1	2925,0	2924,3	0,7
2	2946,3	2924,3	22**
3	Bağ maddesi	Bağ maddesi	Bağ maddesi
4	2947,8	2924,3	23,5**
5	Bağ maddesi	Bağ maddesi	Bağ maddesi
6	2946,2	2924,3	21,9**
7	2928,8	2924,3	4,5
8	3008,9	2924,3	84,6**
9	Bağ maddesi	Bağ maddesi	Bağ maddesi
10	2927,0	2924,3	2,7
11	2942,2	2924,3	17,9**
12	2937,5	2924,3	13,2*
13	Bağ maddesi	Bağ maddesi	Bağ maddesi
14	2976,0	2924,3	51,7**

$\chi^2_{(0,01)} = 13,28$, sd.=4, **p<0,01

$\chi^2_{(0,05)} = 9,49$, sd.=4, *p<0,05

Değişen madde fonksiyonu analizi için 3,5,9 ve 13. maddelerin gruplar arasında eşitlendiği geniş modelin -2logL değeri 2924,3 olarak elde edilmiştir.

Analiz sonuçlarına göre Mice okuma parçası değerlendirme testinde bulunan bağ maddesi dışındaki 12. maddenin dar ve geniş modelleri -2logL değerleri arasındaki fark 0,05, 2,4,6,8,11 ve 14. maddelerin dar ve geniş modelleri -2logL değerleri arasındaki farklar da 0,01 anlamlılık düzeyinde önemli bulunmuştur. Bu bulgunun anlamı ise çevirisi ve adaptasyonu yapılan Mice okuma parçası değerlendirme testinin kültürler arası uygulamalarında değişen madde fonksiyonu gözlenmiştir ve söz konusu maddelerin madde yerleşim parametreleri ABD ve Türkiye örneklem gruplarında ölçme değişmezliğine sahip değildir. Bu nedenle testin kaynak kültürdeki versiyonu ile hedef kültürdeki versiyonunun uygulama sonuçları arasında doğrudan karşılaştırma yapmanın yanıltıcı bulgulara yol açacağı düşünülmektedir.

River okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında olabilirlik oranı testi model karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Tablo 3.31’de verildiği gibi, River okuma parçası değerlendirme testi değişen madde fonksiyonu analizinde bağ maddelerinin seçimi için yapılan I. tekrarlayıcı madde seçme işlemi sonucunda 5 ve 9. maddelerin dar ve geniş modelleri arasındaki fark değerleri 0,05 anlamlılık düzeyinde 1,4,6,10 ve 11. maddelerin ise 0,01 düzeyinde manidar bulunmuştur.

En yüksek G_i^2 fark değeri veren 10. madde ($\chi^2_{(4; 0,01)} = 145,6$, $p < 0,01$) testten çıkarılarak kalan 10 madde için II. tekrarlayıcı madde seçme işlemine geçildi. II. iterasyonda dar modelin -2logL değeri 1925,0 olarak elde edildi. Bu iterasyonda en yüksek G_i^2 fark değeri veren 4. madde testten çıkarılarak kalan 9 madde için III. tekrarlayıcı madde seçme işlemine geçildi. III. İterasyona ait dar modelin -2logL değeri

1161,3 olarak hesaplandı. III. İterasyondan sonra ise yüksek G_i^2 fark değeri veren 5. madde testten çıkarılarak kalan 8 madde için IV. tekrarlayıcı madde seçme işlemine geçildi. IV. İterasyona ait dar modelin -2logL değeri 535,4 olarak hesaplandı.

Tablo 3.31
River Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin
I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları

Madde no	Geniş Model									I. İterasyon
	Referans Grup (ABD)				Odak Grup (Türkiye)					
	b_i	c_{i1}	c_{i2}	c_{i3}	b_i	c_{i1}	c_{i2}	c_{i3}	$-2\log L$	
1	-1,07				1,96				3766,8	91**
2	0,43				0,90				3854,9	2,9
3	-1,35				-1,13				3856,0	1,8
4	-1,38				-0,48				3827,9	29,9**
5	-0,77				-0,34				3846,8	11*
6		-1,15	-0,56			-1,15	-0,89		3842,3	15,5**
7	0,56				0,40				3856,6	1,2
8		-0,43	-1,40			-0,43	-1,48		3857,6	0,2
9		0,20	0,79			0,20	0,14		3845,3	12,5*
10		0,22	0,44	-1,88		0,22	0,92	0,98	3712,2	145,6**
11		-0,13	0,33	-1,60		-0,13	-0,23	-1,26	3829,6	28,2**
Dar Model $-2\log L = 3857,8$										

Dar Model -2logL = 3857,8

$\chi^2_{(0,01)} = 13,28$, sd.=4, **p<0,01

$\chi^2_{(0,05)} = 9,49$, sd.=4, *p<0,05

Tekrarlayıcı madde seçme işlemlerinin yapıldığı dört iterasyona ait bilgiler Tablo 3.32’de verilmektedir.

Tablo 3.32’de görüldüğü gibi dört tekrarlayıcı madde seçme işlemi sonucunda her dört iterasyonda da dar model ile geniş modellerinin uyum farkı istatistiki olarak anlamlı bulunmayan 1,7,8 ve 9. maddeler değişen madde fonksiyonu gösterme eğilimi olmayan maddeler olduklarından bağ maddeleri olarak belirlendiler.

Tablo 3.32

River Okuma Parçası Değerlendirme Testi I. , II. ,III. ve IV. Tekrarlayıcı Madde Seçme İşlemleri Dar Ve Geniş Modeller Karşılaştırma Sonuçları

Madde no	I. İterasyon $G_1^2 - G_2^2$	II. İterasyon $G_1^2 - G_2^2$	III. İterasyon $G_1^2 - G_2^2$	IV. İterasyon $G_1^2 - G_2^2$
1	91**	0,3	2,2	1,2
2	2,9	1,2	1	1,8
3	1,8	2,4	1,6	3,3
4	29,9**	23,8**	Çıkarıldı	Çıkarıldı
5	11*	6,2	26,6**	Çıkarıldı
6	15,5**	7,7	1	8,9
7	1,2	3,8	2,6	2
8	0,2	0,8	0,5	0,1
9	12,5*	4,8	5,5	8,5
10	145,6**	Çıkarıldı	Çıkarıldı	Çıkarıldı
11	28,2**	18,8**	36**	39,6**

$\chi^2_{(0,01)} = 13,28$, sd.=4, **p<0,01

$\chi^2_{(0,05)} = 9,49$, sd.=4, *p<0,05

Bağ maddelerini seçme işlemi tamamlandıktan sonra yeniden dar ve geniş modeller için madde parametre tahminleri ve $-2\log L$ değerleri hesaplandı. Bu nedenle oluşturulan geniş modelde sadece 1,7,8 ve 9. maddelerin referans ve odak gruplarda parametreleri eşitlendi ve diğer maddelerin parametre değerleri gruplar arasında eşitlenmedi. Daha sonra bağ maddeleri dışında kalan her bir madde için bağ maddeleri ile birlikte eşitlendikleri dar modeller kuruldu. Devam eden aşamada dar ve geniş modellerin olabilirlik oranlarının karşılaştırılması, dar modele sınırlandırılarak sokulan maddenin ait olduğu modelin olabilirlik oranı geniş modelde serbest bırakıldığı durumda elde edilen geniş modelin olabilirlik oranı değerinden çıkarılarak yapılmıştır. Bu şekilde geniş ve dar modeller karşılaştırıldığında iki modelin uyum değerleri arasındaki farkın istatistiki olarak anlamlı olup olmadığı ortaya çıkarıldı. River okuma parçası değerlendirme testine ait dar ve geniş modellerin $-2\log L$ değerleri ile değişen madde fonksiyonu analizi sonuçları Tablo 3.33'de verilmiştir.

Tablo 3.33
River Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş
Modellerin -2logl Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları

Madde no	Dar model G_1^2	Geniş model G_2^2	$G_i^2 = G_1^2 - G_2^2$
1	Bağ maddesi	Bağ maddesi	Bağ maddesi
2	3617,4	3614,2	3,2
3	3620,0	3614,2	5,8
4	3632,9	3614,2	18,7**
5	3629,0	3614,2	14,8**
6	3633,6	3614,2	19,4**
7	Bağ maddesi	Bağ maddesi	Bağ maddesi
8	Bağ maddesi	Bağ maddesi	Bağ maddesi
9	Bağ maddesi	Bağ maddesi	Bağ maddesi
10	3754,6	3614,2	140,4**
11	3665,7	3614,2	51,5**

$$\chi^2_{(0,01)} = 13,28, \text{ sd.}=4, **p<0,01$$

$$\chi^2_{(0,05)} = 9,49, \text{ sd.}=4, *p<0,05$$

Değişen madde fonksiyonu analizi için 1,7,8 ve 9. maddelerin gruplar arasında eşitlendiği geniş modelin -2logL değeri 3614,2 olarak elde edilmiştir. Analiz sonuçlarına göre River okuma parçası değerlendirme testinde bulunan bağ maddesi dışındaki 4,5,6,10 ve 11. maddelerin dar ve geniş modelleri -2logL değerleri arasındaki fark 0,01 anlamlılık düzeyinde önemli bulunmuştur. Bu bulgunun anlamı ise çevirisi ve adaptasyonu yapılan River okuma parçası değerlendirme testinin kültürler arası uygulamalarında değişen madde fonksiyonu gözlenmiştir ve söz konusu maddelerin madde yerleşim parametreleri ABD ve Türkiye örneklem gruplarında ölçme değişmezliğine sahip değildir. Bu nedenle testin kaynak kültürdeki versiyonu ile hedef kültürdeki versiyonunun uygulama sonuçları arasında doğrudan karşılaştırma yapmanın yanıltıcı bulgulara yol açacağı düşünülmektedir.

Clay okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında olabilirlik oranı testi model karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Clay okuma parçası değerlendirme testi değişen madde fonksiyonu analizinde bağ maddelerinin seçimi için yapılan I. tekrarlayıcı madde seçme işlemine ait bulgular Tablo 3.34'de verilmiştir. Analizler sonucunda 1 ve 11. maddelerin dar ve geniş modelleri arasındaki fark değerleri 0,05 anlamlılık düzeyinde 6 ve 10. maddelerin ise 0,01 düzeyinde manidar bulunmuştur.

Tablo 3.34

Clay Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin
I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları

Madde no	Geniş Model									I. İterasyon
	Referans Grup (ABD)				Odak Grup (Türkiye)					
	b_i	c_{i1}	c_{i2}	c_{i3}	b_i	c_{i1}	c_{i2}	c_{i3}	$-2\log L$	
1	0,03				0,33				3830,0	12,1*
2	0,63				0,21				3833,2	8,9
3	-0,29				-0,18				3840,9	1,2
4	-0,08				0,59				3825,1	17
5	-0,70				-0,26				3835,3	6,8
6	0,09				0,65				3820,5	20,6**
7	-0,33				-0,28				3841,9	0,2
8		-1,18	-1,28			-1,02	-0,64		3834,2	7,9
9	0,45								3838,9	3,2
10		-0,13	-1,14	-1,34		-0,13	-0,41	-1,84	3820,7	21,4**
11		0,07	-1,14			0,07	-0,42		3830,5	11,6*
12	-0,56								3833,1	9
13	0,18								3841,4	0,7

Dar Model $-2\log L = 3842,1$

Dar Model -2logL = 3842,1

$\chi^2_{(0,01)} = 13,28$, sd.=4, **p<0,01

$\chi^2_{(0,05)} = 9,49$, sd.=4, *p<0,05

En yüksek G_i^2 fark değeri veren 10. madde ($\chi^2_{(4; 0,01)} = 21,4$, p<0,01) testten çıkarılarak kalan 12 madde için II. tekrarlayıcı madde seçme işlemine geçildi. II. iterasyonda dar modelin -2logL değeri 2202,6 olarak elde edildi. Bu iterasyonda en yüksek G_i^2 fark

değeri veren 6. madde testten çıkarılarak kalan 11 madde için III. tekrarlayıcı madde seçme işlemine geçildi. III. İterasyona ait dar modelin $-2\log L$ değeri 1447,5 olarak hesaplandı. III. İterasyonadan sonra ise yüksek G_i^2 fark değeri veren 11. madde testten çıkarılarak kalan 10 madde için IV. tekrarlayıcı madde seçme işlemine geçildi. IV. İterasyona ait dar modelin $-2\log L$ değeri 11,1 olarak hesaplandı.

Tekrarlayıcı madde seçme işlemlerinin yapıldığı dört iterasyona ait bilgiler Tablo 3.35'de verilmektedir. Tabloda görüldüğü gibi dört tekrarlayıcı madde seçme işlemi sonucunda her dört iterasyonda da dar model ile geniş modellerinin uyum farkı istatistiki olarak anlamlı bulunmayan 1,2,4 ve 8. maddeler değişen madde fonksiyonu gösterme eğilimi olmayan maddeler olduklarından bağ maddeleri olarak belirlendiler.

Tablo 3.35

Clay Okuma Parçası Değerlendirme Testi I., II. , III. ve IV. Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modeller Karşılaştırma Sonuçları

Madde no	I. İterasyon $G_1^2 - G_2^2$	II. İterasyon $G_1^2 - G_2^2$	III. İterasyon $G_1^2 - G_2^2$	IV. İterasyon $G_1^2 - G_2^2$
1	12,1*	4,5	3,8	0,2
2	8,9	11,2*	7,7	7,8
3	1,2	4,9	1	2,2
4	17	21,8**	17,3**	2,4
5	6,8	10,5*	6,2	6,9
6	20,6**	30,3**	Çıkarıldı	Çıkarıldı
7	0,2	3,7	0,4	0,1
8	7,9	7,5	0,2	5,8
9	3,2	7,3	6,4	3,2
10	21,4**	Çıkarıldı	Çıkarıldı	Çıkarıldı
11	11,6*	15,4**	20,8**	Çıkarıldı
12	9	12,5**	5,7	10,2*
13	0,7	2,1	1,3	0,8

$\chi^2_{(0,01)} = 13,28$, sd.=4, **p<0,01

$\chi^2_{(0,05)} = 9,49$, sd.=4, *p<0,05

Bağ maddelerini seçme işlemi tamamlandıktan sonra yeniden dar ve geniş modeller için madde parametre tahminleri ve $-2\log L$ değerleri hesaplandı. Bu nedenle oluşturulan geniş modelde sadece 1,2,4 ve 8. maddelerin referans ve odak gruplarda parametreleri eşitlendi ve diğer maddelerin parametre değerleri gruplar arasında eşitlenmedi. Daha

sonra bağ maddeleri dışında kalan her bir madde için bağ maddeleri ile birlikte eşitlendikleri dar modeller kuruldu. Devam eden aşamada dar ve geniş modellerin olabilirlik oranlarının karşılaştırılması, dar modele sınırlanarak sokulan maddenin ait olduğu modelin olabilirlik oranı geniş modelde serbest bırakıldığı durumda elde edilen geniş modelin olabilirlik oranı değerinden çıkarılarak yapılmıştır. Bu şekilde geniş ve dar modeller karşılaştırıldığında iki modelin uyum değerleri arasındaki farkın istatistiki olarak anlamlı olup olmadığı ortaya çıkarılmıştır. Clay okuma parçası değerlendirme testine ait dar ve geniş modellerin $-2\log L$ değerleri ile değişen madde fonksiyonu analizi sonuçları Tablo 3.36'da verilmiştir.

Tablo 3.36
Clay Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş
Modellerin $-2\log L$ Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları

Madde no	Dar model G_1^2	Geniş model G_2^2	$G_i^2 = G_1^2 - G_2^2$
1	Bağ maddesi	Bağ maddesi	Bağ maddesi
2	Bağ maddesi	Bağ maddesi	Bağ maddesi
3	3685,6	3675,1	10,5*
4	Bağ maddesi	Bağ maddesi	Bağ maddesi
5	3699,5	3675,1	24,4**
6	3721,4	3675,1	46,3**
7	3682	3675,1	6,9
8	Bağ maddesi	Bağ maddesi	Bağ maddesi
9	3686,7	3675,1	11,6*
10	3715,9	3675,1	40,8**
11	3726	3675,1	50,9**
12	3701,8	3675,1	26,7**
13	3680,4	3675,1	5,3

$\chi^2_{(0,01)} = 13,28$, sd.=4, **p<0,01

$\chi^2_{(0,05)} = 9,49$, sd.=4, *p<0,05

Değişen madde fonksiyonu analizi için 1,2,4 ve 8. maddelerin gruplar arasında eşitlendiği geniş modelin $-2\log L$ değeri 3675,1 olarak elde edilmiştir. Analiz sonuçlarına göre Clay okuma parçası değerlendirme testinde bulunan bağ maddesi dışındaki 3 ve 19. maddelerin dar ve geniş modelleri $-2\log L$ değerleri arasındaki fark 0,05 anlamlılık düzeyinde, 5,6,10,11 ve 12. maddelerin dar ve geniş modelleri $-2\log L$ değerleri arasındaki fark ise 0,01 anlamlılık düzeyinde önemli bulunmuştur. Bu

bulgunun anlamı, çevirisi ve adaptasyonu yapılan Clay okuma parçası değerlendirme testinin kültürler arası uygulamalarında değişen madde fonksiyonu gözlenmiştir ve söz konusu maddelerin yerleşim parametrelerinin ABD ve Türkiye örneklem gruplarında ölçme değişmezliğine sahip olmadıkları gözlenmiştir.

Leonardo okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında olabilirlik oranı testi model karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Tablo 3.37’da verildiği gibi, Leonardo okuma parçası değerlendirme testi değişen madde fonksiyonu analizinde bağ maddelerinin seçimi için yapılan I. tekrarlayıcı madde seçme işlemi sonucunda 9 ve 10. maddelerin dar ve geniş modelleri arasındaki fark değerleri 0,05 anlamlılık düzeyinde 4,5,6,7,8 ve 11. maddelerin ise 0,01 düzeyinde manidar bulunmuştur.

Tablo 3.37
Leonardo Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin
I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları

Madde no	Geniş Model									I. İterasyon
	Referans Grup (ABD)				Odak Grup (Türkiye)					
	b_i	c_{i1}	c_{i2}	c_{i3}	b_i	c_{i1}	c_{i2}	c_{i3}	$-2\log L$	
1	-1,89				-2,10				4125,1	0,8
2	0,00				0,38				4124,1	1,8
3	-0,51				-0,28				4125,0	0,9
4		0,97	-1,51	-3,29		0,94	0,59	0,45	1937,5	188,4**
5	-0,99				0,50				4050,2	75,7**
6	0,05				0,75				4101,7	24,2**
7	-0,17				0,60				4111,5	14,4**
8		-0,05	-1,13			-0,05	-2,20		4105,9	20**
9	-0,56				-0,98				4115,2	10,7*
10		-0,71	-1,34			-0,71	-2,13		4114,1	11,8*
11	-1,03				0,11				4082,0	43,9**
12		0,88	-1,10			-0,56	-1,68		4121,9	4
Dar Model $-2\log L = 4125,9$										

Dar Model $-2\log L = 4125,9$

$\chi^2_{(0,01)} = 13,28$, sd.=4, **p<0,01

$\chi^2_{(0,05)} = 9,49$, sd.=4, *p<0,05

En yüksek G_i^2 fark değeri veren 5. madde ($\chi^2_{(4; 0,01)} = 188,4$, $p < 0,01$) testten çıkarılarak kalan 11 madde için II. tekrarlayıcı madde seçme işlemine geçildi. II. iterasyonda dar modelin -2logL değeri 3218,6 olarak elde edildi. Bu iterasyonda en yüksek G_i^2 fark değeri veren 4. madde testten çıkarılarak kalan 10 madde için III. tekrarlayıcı madde seçme işlemine geçildi. III. İterasyona ait dar modelin -2logL değeri 1352,9 olarak hesaplandı. III. İterasyonadan sonra ise yüksek G_i^2 fark değeri veren 8. madde testten çıkarılarak kalan 9 madde için IV. tekrarlayıcı madde seçme işlemine geçildi. IV. İterasyona ait dar modelin -2logL değeri 79,9 olarak hesaplandı.

Tekrarlayıcı madde seçme işlemlerinin yapıldığı dört iterasyona ait bilgiler Tablo 3.38'de verilmektedir. Tablo 3.38'de görüldüğü gibi dört tekrarlayıcı madde seçme işlemi sonucunda her dört iterasyonda da dar model ile geniş modellerinin uyum farkı istatistiki olarak anlamlı bulunmayan 1,2,3 ve 10. maddeler bağ maddeleri olarak belirlendiler ve analizlerin ikinci aşamasına geçildi.

Tablo 3.38

Leonardo Okuma Parçası Değerlendirme Testi I. , II. ,III. ve IV. Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modeller Karşılaştırma Sonuçları

Madde no	I. İterasyon $G_1^2 - G_2^2$	II. İterasyon $G_1^2 - G_2^2$	III. İterasyon $G_1^2 - G_2^2$	IV. İterasyon $G_1^2 - G_2^2$
1	0,8	0,3	1,4	1,1
2	1,8	2,4	0,9	1,8
3	0,9	1,5	0,6	0,6
4	188,4**	167,9**	Çıkarıldı	Çıkarıldı
5	75,7**	Çıkarıldı	Çıkarıldı	Çıkarıldı
6	24,2**	30,5**	18,3**	20,7**
7	14,4**	17,9**	9,3	10,8*
8	20**	25,1**	15**	Çıkarıldı
9	10,7*	7	12,6*	12,7*
10	11,8*	12,5*	4,9	3
11	43,9**	51,3**	36,4**	39,2**
12	4	2,3	11,8*	9,1

$\chi^2_{(0,01)} = 13,28$, sd.=4, **p<0,01

$\chi^2_{(0,05)} = 9,49$, sd.=4, *p<0,05

Bağ maddelerini seçme işlemi tamamlandıktan sonra yeniden dar ve geniş modeller için madde parametre tahminleri ve $-2\log L$ değerleri hesaplandı. Bu nedenle oluşturulan geniş modelde sadece 1,2,3 ve 10. maddelerin referans ve odak gruplarda parametreleri eşitlendi ve diğer maddelerin parametre değerleri gruplar arasında eşitlenmedi. Daha sonra bağ maddeleri dışında kalan her bir madde için bağ maddeleri ile birlikte eşitlendikleri dar modeller kuruldu. Devam eden aşamada dar ve geniş modellerin olabilirlik oranlarının karşılaştırılması, dar modele sınırlanarak sokulan maddenin ait olduğu modelin olabilirlik oranı geniş modelde serbest bırakıldığı durumda elde edilen geniş modelin olabilirlik oranı değerinden çıkarılarak yapılmıştır. Bu şekilde geniş ve dar modeller karşılaştırıldığında iki modelin uyum değerleri arasındaki farkın istatistiki olarak anlamlı olup olmadığı ortaya çıkarıldı. Leonardo okuma parçası değerlendirme testine ait dar ve geniş modellerin $-2\log L$ değerleri ile değişen madde fonksiyonu analizi sonuçları Tablo 3.39'da verilmiştir.

Tablo 3.39
Leonardo Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin $-2\log L$ Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları

Madde no	Dar model G_1^2	Geniş model G_2^2	$G_i^2 = G_1^2 - G_2^2$
1	Bağ maddesi	Bağ maddesi	Bağ maddesi
2	Bağ maddesi	Bağ maddesi	Bağ maddesi
3	Bağ maddesi	Bağ maddesi	Bağ maddesi
4	3781,9	3689,3	92,6**
5	3787,4	3689,3	98,1**
6	3726,3	3689,3	37**
7	3716,8	3689,3	27,5**
8	3717,3	3689,3	28**
9	3691,5	3689,3	2,2
10	Bağ maddesi	Bağ maddesi	Bağ maddesi
11	3752,4	3689,3	63,1**
12	3760,3	3689,3	71**

$\chi^2_{(0,01)} = 13,28$, sd.=4, **p<0,01

$\chi^2_{(0,05)} = 9,49$, sd.=4, *p<0,05

Değişen madde fonksiyonu analizi için 1,2,3 ve 10. maddelerin gruplar arasında eşitlendiği geniş modelin $-2\log L$ değeri 3689,3 olarak elde edilmiştir. Analiz sonuçlarına göre Leonardo okuma parçası değerlendirme testinde bulunan bağ

maddeleri dışındaki 4,5,6,7,8,11 ve 12. maddelerin dar ve geniş modelleri -2logL değerleri arasındaki fark 0,01 anlamlılık düzeyinde önemli bulunmuştur. Bu bulgunun anlamı ise çevirisi ve adaptasyonu yapılan Leonardo okuma parçası değerlendirme testinin kültürler arası uygulamalarında değişen madde fonksiyonu gözlenmiştir ve söz konusu maddelerin madde yerleşim parametrelerinin ABD ve Türkiye örneklem gruplarında ölçme değişmezliğine sahip olmadıkları gözlenmiştir.

Hare okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında olabilirlik oranı testi model karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Hare okuma parçası değerlendirme testi değişen madde fonksiyonu analizinde bağ maddelerinin seçimi için yapılan I. tekrarlayıcı madde seçme işlemine ait sonuçlar Tablo 3.40'da verilmiştir.

Tablo 3.40

Hare Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları

Madde no	Geniş Model									I. İterasyon
	Referans Grup (ABD)				Odak Grup (Türkiye)					
	b _i	c _{i1}	c _{i2}	c _{i3}	b _i	c _{i1}	c _{i2}	c _{i3}	-2logL	
1	-2,03				-1,24				2255,6	0,3
2	-2,15				-2,16				2155,4	100,5**
3		-0,06	-1,10			0,11	-0,66		2244,8	11,1*
4	-1,32				-1,20				2253,9	2
5	-1,54				-1,03				2221,7	34,2**
6	-0,84				-0,90				2255,6	0,3
7		0,29	0,35			0,29	0,67		2244,9	11
8	-0,56				0,24				2183,8	72,1**
9		0,01	0,34			0,03	0,00		2244,2	11,7*
10		-1,30	-1,05	-2,14		-1,30	-1,33	-2,21	2238,9	17*
11	-0,50				-0,78				2196,1	59,8**
Dar Model -2logL = 2255,9										

$$\chi^2_{(0,01)} = 13,28, \text{sd.}=4, **p<0,01$$

$$\chi^2_{(0,05)} = 9,49, \text{sd.}=4, *p<0,05$$

En yüksek G_i^2 fark değeri veren 2. madde ($\chi^2_{(4; 0,01)} = 100,5$, $p < 0,01$) testten çıkarılarak kalan 10 madde için II. tekrarlayıcı madde seçme işlemine geçildi. II. iterasyonda dar modelin $-2\log L$ değeri 243,9 olarak elde edildi. Bu iterasyonda en yüksek G_i^2 fark değeri veren 8. madde testten çıkarılarak kalan 9 madde için III. tekrarlayıcı madde seçme işlemine geçildi. III. İterasyonun dar modeline ait $-2\log L$ değerinin eksi bir değer çıkması üzerine bu iterasyon tamamlanamadı.

Tekrarlayıcı madde seçme işlemlerinin yapıldığı iki iterasyona ait bilgiler Tablo 3.41’de verilmektedir.

Tablo 3.41
Hare Okuma Parçası Değerlendirme Testi I. ve II. Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modeller Karşılaştırma Sonuçları

Madde no	I. İterasyon $G_1^2 - G_2^2$	II. İterasyon $G_1^2 - G_2^2$
1	0,3	0,4
2	100,5**	Çıkarıldı
3	11,1*	8,3
4	2	6,6
5	34,2**	48,9**
6	0,3	0,1
7	11	0
8	72,1**	84,7**
9	11,7*	15,7**
10	17*	6,7
11	59,8**	10,5**

$\chi^2_{(0,01)} = 13,28$, sd.=4, ** $p < 0,01$

$\chi^2_{(0,05)} = 9,49$, sd.=4, * $p < 0,05$

Tablo 3.41’de görüldüğü gibi iki tekrarlayıcı madde seçme işlemi sonucunda her iki iterasyonda da dar model ile geniş modellerinin uyum farkı istatistiki olarak anlamlı bulunmayan 1,6 ve 7. maddeler bağ maddeleri olarak belirlendiler ve analizlerin ikinci aşamasına geçildi..

Bağ maddelerini seçme işlemi tamamlandıktan sonra yeniden dar ve geniş modeller için madde parametre tahminleri ve $-2\log L$ değerleri hesaplandı. Bu nedenle oluşturulan

geniş modelde sadece 1,6 ve 7. maddelerin referans ve odak gruplarda parametreleri eşitlendi ve diğer maddelerin parametre değerleri gruplar arasında eşitlenmedi. Daha sonra bağ maddeleri dışında kalan her bir madde için bağ maddeleri ile birlikte eşitlendikleri dar modeller kuruldu. Devam eden aşamada dar ve geniş modellerin olabilirlik oranlarının karşılaştırılması, dar modele sınırlanarak sokulan maddenin ait olduğu modelin olabilirlik oranı geniş modelde serbest bırakıldığı durumda elde edilen geniş modelin olabilirlik oranı değerinden çıkarılarak yapılmıştır. Bu şekilde geniş ve dar modeller karşılaştırıldığında iki modelin uyum değerleri arasındaki farkın istatistiki olarak anlamlı olup olmadığı ortaya çıkarılmıştır. Hare okuma parçası değerlendirme testine ait dar ve geniş modellerin $-2\log L$ değerleri ile değişen madde fonksiyonu analizi sonuçları Tablo 3.42’de verilmiştir.

Tablo 3.42
Hare Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş
Modellerin $-2\log L$ Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları

Madde no	Dar model G_1^2	Geniş model G_2^2	$G_i^2 = G_1^2 - G_2^2$
1	Bağ maddesi	Bağ maddesi	Bağ maddesi
2	2075	1921,5	153,5**
3	1922,1	1921,5	0,6
4	1953,6	1921,5	32,1**
5	2004,5	1921,5	83**
6	Bağ maddesi	Bağ maddesi	Bağ maddesi
7	Bağ maddesi	Bağ maddesi	Bağ maddesi
8	2048,8	1921,5	127,3**
9	1958,9	1921,5	37,4**
10	1946	1921,5	24,5**
11	1923,5	1921,5	2

$\chi^2_{(0,01)} = 13,28$, sd.=4, **p<0,01

$\chi^2_{(0,05)} = 9,49$, sd.=4, *p<0,05

Değişen madde fonksiyonu analizi için 1,2,3 ve 10. maddelerin gruplar arasında eşitlendiği geniş modelin $-2\log L$ değeri 1921,5 olarak elde edilmiştir. Analiz sonuçlarına göre Hare okuma parçası değerlendirme testinde bulunan bağ maddeleri dışındaki 2,4,5,8,9 ve 10. maddelerin dar ve geniş modelleri $-2\log L$ değerleri arasındaki fark 0,01 anlamlılık düzeyinde önemli bulunmuştur. Bu bulgunun anlamı ise çevirisi ve adaptasyonu yapılan Hare okuma parçası değerlendirme testinin kültürler arası

uygulamalarında değişen madde fonksiyonu gözlenmiştir ve söz konusu maddelerin madde yerleşim parametrelerinin ABD ve Türkiye örneklem gruplarında ölçme değişmezliğine sahip olmadıkları gözlenmiştir.

Pufflings okuma parçası değerlendirme soruları için ABD ve Türkiye örneklem gruplarında olabilirlik oranı testi model karşılaştırma yöntemi ile yapılan DMF analizine ait bulgular:

Pufflings okuma parçası değerlendirme testi değişen madde fonksiyonu analizinde bağ maddelerinin seçimi için yapılan I. tekrarlayıcı madde seçme işlemine ait sonuçlar Tablo 3.43’de verilmiştir. Analizler sonucunda 6 ve 9. maddelerin dar ve geniş modelleri arasındaki fark değerleri 0,05 anlamlılık düzeyinde 1,4,7,8,10,11,12 ve 13. maddelerin ise 0,01 düzeyinde manidar bulunmuştur.

Tablo 3.43
Pufflings Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş Modellerin
I. İterasyon Madde Parametre Tahminleri ve Olabilirlik Oranı Testi Sonuçları

Madde no	Geniş Model							I. İterasyon
	Referans Grup (ABD)			Odak Grup (Türkiye)				
	b_i	c_{i1}	c_{i2}	b_i	c_{i1}	c_{i2}	$-2\log L$	
1	-0,60			-1,23			6950,9	$G_1^2 - G_2^2$ 36,2**
2	-0,16			-0,29			6986,0	1,1
3	-1,14			-1,12			6987,1	0
4	0,43			0,90			6969,7	17,4**
5	0,08			0,00			6986,2	0,9
6	0,49			0,83			6977,9	9,2*
7		-0,34	-2,19		-0,47	-1,29	6935,4	51,7**
8		0,93	-0,26		-0,60	-0,48	6771,7	215,4**
9	-0,66			-0,40			6976,3	10,8*
10	1,52			0,95			6950,2	36,9**
11	0,50						6944,0	43,1**
12		0,07	0,39		-1,67	-3,17	6435,1	552**
13	-0,40			0,59			6868,3	118,8**
Dar Model $-2\log L = 6987,1$								

$$\chi^2_{(0,01)}=11,34, \text{sd.}=3, **p<0,01$$

$$\chi^2_{(0,05)}=7,81, \text{sd.}=3, *p<0,05$$

En yüksek G_i^2 fark değeri veren 12. madde ($\chi^2_{(3; 0,01)} = 552$, $p < 0,01$) testten çıkarılarak kalan 12 madde için II. tekrarlayıcı madde seçme işlemine geçildi. II. iterasyonda dar modelin $-2\log L$ değeri 3489,0 olarak elde edildi. Bu iterasyonda en yüksek G_i^2 fark değeri veren 9. madde testten çıkarılarak kalan 11 madde için III. tekrarlayıcı madde seçme işlemine geçildi. III. İterasyona ait dar modelin $-2\log L$ değeri 1142,6 olarak hesaplandı. Tekrarlayıcı madde seçme işlemlerinin yapıldığı üç iterasyona ait bilgiler Tablo 3.44’de verilmektedir.

Tablo 3.44

Pufflings Okuma Parçası Değerlendirme Testi I. , II. ve III. Tekrarlayıcı Madde Seçme İşlemleri Dar ve Geniş Modeller Karşılaştırma Sonuçları

Madde no	I. İterasyon $G_1^2 - G_2^2$	II. İterasyon $G_1^2 - G_2^2$	III. İterasyon $G_1^2 - G_2^2$
1	36,2**	22,9**	19,5**
2	1,1	0,8	0,2
3	0	2,7	4,9
4	17,4**	35,2**	43,9**
5	0,9	1,4	14,4**
6	9,2*	24,4**	29,4**
7	51,7**	32,2**	17,8**
8	215,4**	19,5**	14,4**
9	10,8*	32,4**	Çıkarıldı
10	36,9**	15,7**	5,7
11	43,1**	25,6**	21,1**
12	552**	Çıkarıldı	Çıkarıldı
13	118,8**	21,9**	28,4**

$$\chi^2_{(0,01)} = 11,34, \text{ sd.} = 3, **p < 0,01$$

$$\chi^2_{(0,05)} = 7,81, \text{ sd.} = 3, *p < 0,05$$

Tablo 3.44’de görüldüğü gibi üç tekrarlayıcı madde seçme işlemi sonucunda her üç iterasyonda da dar model ile geniş modellerinin uyum farkı istatistiki olarak anlamlı bulunmayan 2,3 ve 10. maddeler bağ maddeleri olarak belirlendiler ve analizlerin ikinci aşamasına geçildi.

Bağ maddelerini seçme işlemi tamamlandıktan sonra yeniden dar ve geniş modeller için madde parametre tahminleri ve $-2\log L$ değerleri hesaplandı. Bu nedenle oluşturulan

geniş modelde sadece 2,3 ve 10. maddelerin referans ve odak gruplarda parametreleri eşitlendi ve diğer maddelerin parametre değerleri gruplar arasında eşitlenmedi. Daha sonra bağ maddeleri dışında kalan her bir madde için bağ maddeleri ile birlikte eşitlendikleri dar modeller kuruldu. Devam eden aşamada dar ve geniş modellerin olabilirlik oranlarının karşılaştırılması, dar modele sınırlandırılarak sokulan maddenin ait olduğu modelin olabilirlik oranı geniş modelde serbest bırakıldığı durumda elde edilen geniş modelin olabilirlik oran değerinden çıkarılarak yapılmıştır. Bu şekilde geniş ve dar modeller karşılaştırıldığında iki modelin uyum değerleri arasındaki farkın istatistiki olarak anlamlı olup olmadığı ortaya çıkarılmıştır. Pufflings okuma parçası değerlendirme testine ait dar ve geniş modellerin -2logL değerleri ile değişen madde fonksiyonu analizi sonuçları Tablo 3.45’de verilmiştir.

Tablo 3.45
Pufflings Okuma Parçası Değerlendirme Testine Ait Dar ve Geniş
Modellerin -2logL Değerleri İle Değişen Madde Fonksiyonu Analizi Sonuçları

Madde no	Dar model G_1^2	Geniş model G_2^2	$G_i^2 = G_1^2 - G_2^2$
1	5678,8	5677,9	0,9
2	Bağ maddesi	Bağ maddesi	Bağ maddesi
3	Bağ maddesi	Bağ maddesi	Bağ maddesi
4	5769,8	5677,9	91,9**
5	5710,8	5677,9	32,9**
6	5794	5677,9	116,1**
7	5724,3	5677,9	46,4**
8	5961,7	5677,9	283,8**
9	5792,5	5677,9	114,6**
10	Bağ maddesi	Bağ maddesi	Bağ maddesi
11	5689	5677,9	11,1*
12	6427	5677,9	749,1**
13	5941,3	5677,9	263,4**

$$\chi^2_{(0,01)}=11,34, \text{sd.}=3, **p<0,01$$

$$\chi^2_{(0,05)}=7,81, \text{sd.}=3, *p<0,05$$

Değişen madde fonksiyonu analizi için 2,3 ve 10. maddelerin gruplar arasında eşitlendiği geniş modelin -2logL değeri 5677,9 olarak elde edilmiştir. Analiz sonuçlarına göre Pufflings okuma parçası değerlendirme testinde bulunan bağ maddeleri dışındaki 11. maddenin dar ve geniş modelleri -2logL değerleri arasındaki fark 0,05

anlamlılık düzeyinde 4,5,6,7,8,9,12 ve 13. maddelerin dar ve geniş modelleri -2logL değerleri arasındaki fark ise 0,01 anlamlılık düzeyinde önemli bulunmuştur. Bu bulgunun anlamı ise çevirisi ve adaptasyonu yapılan Pufflings okuma parçası değerlendirme testinin kültürler arası uygulamalarında değişen madde fonksiyonu gözlenmiştir ve söz konusu maddelerin madde yerleşim parametrelerinin ABD ve Türkiye örneklem gruplarında ölçme değişmezliğine sahip olmadıkları gözlenmiştir.

PIRLS 2001 projesi kapsamında bulunan sekiz okuma parçasına ait sorular için ABD ve Türkiye örneklem grupları arasında olabilirlik oranı testi model karşılaştırma yöntemi kullanılarak yapılan DMF analizi sonuçları Tablo 3.46'da özetlenmiştir.

Tablo 3.46
Okuma Parçalarına Göre Olabilirlik Oranı Testi Model
Karşılaştırma Yöntemi İle Yapılan DMF Analizi Sonuçları

Testler	Testteki Madde Sayısı	0,05 Düzeyinde DMF Gözlenen Madde Sayısı	0,01 Düzeyinde DMF Gözlenen Madde Sayısı
Flowers	13	-	6
Antarctica	11	-	6
Mice	14	1	6
River	11	-	5
Clay	13	2	5
Leonardo	12	-	7
Hare	11	-	6
Pufflings	13	1	8
Toplam	98	4	49

Tablo 3.48'de görüldüğü gibi PIRLS 2001 projesi kapsamında uygulanan sekiz okuma parçasına ait değerlendirme sorularının toplam sayısı 98 dir. Bu 98 maddenin 4 tanesi (%4,08) 0,05 anlamlılık düzeyinde kültürler arasında değişen madde fonksiyonu göstermiştir. Manidarlık düzeyi 0,01 olarak alındığında ise 49 (%50,0) maddenin kültürler arası uygulamaları sonucunda DMF içerdiği gözlenmiştir.

BÖLÜM IV

SONUÇ ve ÖNERİLER

Bu araştırma kapsamında, Uluslararası Okuma Becerilerinde Gelişim Projesi PIRLS-2001 testinin ABD ve Türkiye'deki uygulamalarına ait veriler analiz edilerek, testin farklı kültürlerdeki psikometrik özellikleri ortaya konmuş, daha sonra kaynak ve hedef kültürlerde testlerin yapılarının eşdeğer olup olmadığı incelenmiştir. Son aşamada ise testlerdeki maddelerin kültürlere göre değişen madde fonksiyonu içerip içermediği araştırılmıştır.

Bu bölümde, araştırma bulgularına bağlı olarak ulaşılan sonuçlar özetlenmiş ve bu sonuçlara göre bazı önerilerde bulunulmuştur.

Sonuçlar

1. Alt Probleme İlişkin Sonuçlar

Araştırmanın birinci alt probleminin a, b, c ve d maddelerinde belirtilen problemlerin çözümü için PIRLS projesi kapsamında Türkiye ve Amerika Birleşik Devletlerinde uygulanan üç, altı, sekiz ve sıfır kodlu testlerin faktör yapılarının kültürler arası eş değeriğe sahip olup olmadığını anlamak için yapılan çok örneklemli doğrulayıcı faktör analizi sonucunda bahsedilen testlerde faktör yüklerinin ve hata varyanslarına ait değerlerin kaynak kültürden elde edilen ölçme sonuçları ile hedef kültürden elde edilen ölçme sonuçları arasında anlamlı bir farklılık olduğu yani bu testlerin yapılarının kültürler arası ölçme eşdeğerliği göstermedikleri gözlenmiştir. Faktör yüklerinin ve hata varyanslarının kültürlere göre serbest olarak tahmin edilmesi için kurulan modellerin temel model ile karşılaştırıldıklarında daha yüksek uyum iyiliği indeks değeri veren modeller oldukları gözlenmiştir. Diğer önemli bir sonuç ise testlerin tamamında temel model olan ModelA'nın uyum iyiliği kriterlerini sağlamayan bir model olduğu anlaşılmıştır. Bu bulgunun anlamı ise, testlerdeki maddeler kuramsal olarak önerildiği gibi dört ayrı kavrama süreci ile belirtilen boyutlarla iddia edilen ilişkileri vermemiştir.

2. Alt Probleme İlişkin Sonuçlar

PIRLS 2001 projesi kapsamında bulunan sekiz okuma parçasına ait sorular için ABD ve Türkiye örneklem grupları arasında parametre karşılaştırma yöntemi kullanılarak yapılan DMF analizleri sonucunda okuma parçalarının tamamına ait 98 sorunun 53 tanesinin kültürler arasında değişen madde fonksiyonu içerdiği ortaya çıkmıştır. DMF içeren tüm maddelerde aynı yetenek düzeyinde fakat farklı örneklem grubunda bulunan öğrencilerin maddeleri doğru cevaplama olasılıkları arasındaki farkın anlamlı olduğu gözlenmektedir. Maddelerin güçlük düzeylerinin kültürlere göre farklılık gösterdiği, genel olarak ABD örneklemindeki öğrencilerin daha yüksek θ yetenek seviyesi gerektiren kategorilere Türkiye örneklemindeki öğrencilere göre daha kolay geçebildikleri sonucuna ulaşılmıştır.

Parametre karşılaştırma yöntemi kullanılarak yapılan DMF analizlerine kavrama süreçlerine göre bakıldığında Tablo 4.1’de verilen sonuçlara ulaşılmıştır.

Tablo 4.1. Kavrama Süreçlerine Göre DMF İçeren Maddelerin Dağılımı

KAVRAMA BOYUTLAR	SÜREÇLERİ/	Madde Sayısı	DMF İçeren Madde Sayısı
Açık bir şekilde ifade edilmiş bilgi ve fikirlerin üzerinde durma ve çıkarımlar yapma		26	12
Doğrudan çıkarımlar yapma		26	13
Fikir ve bilgileri birleştirme ve yorumlama		31	18
Metnin öğelerini, içeriğini ve dilini inceleme ve değerlendirme		15	10
Toplam		98	53

Birinci ve ikinci kavrama düzeyinde olan ve DMF gösteren maddelerin ilgili oldukları kavrama düzeyindeki toplam madde sayısına oranları %50 iken, bu oranın üçüncü ve dördüncü kavrama düzeylerinde arttığı gözlenmiştir. Diğer bir anlatımla, maddelerin yokladığı davranışın kavrama süreçleri açısından düzeyi yükseldikçe DMF içirme olasılıkları da artmaktadır.

Parametre karşılaştırma yöntemi kullanılarak yapılan DMF analizlerinde ortaya çıkan bir diğer sonuç ise, madde kategorileri dikkate alındığında iki kategorili olarak puanlanan 70 sorudan 37 tanesinin (%53) DMF içerdiği, çok kategorili olarak puanlanan 28 sorudan ise 16 tanesinin (%57) DMF içerdiğidir. PIRLS okuma parçalarına ait toplam 98 sorudan 70 (%71) tanesinin iki kategorili puanlanan sorular, diğer 28 (%29) tanesinin ise çok kategorili olarak puanlanan açık uçlu sorular olduğu göz önünde bulundurulduğunda çok kategorili olarak puanlanan soruların DMF içerme eğiliminin daha yüksek olduğu ortaya çıkmaktadır.

PIRLS 2001 projesi kapsamında bulunan sekiz okuma parçasına ait sorular için ABD ve Türkiye örneklem grupları arasında olabilirlik oranı testi model karşılaştırma yöntemi kullanılarak yapılan DMF analizi sonucunda parametre karşılaştırma yönteminde olduğu gibi 98 sorunun 53 tanesinin kültürler arasında değişen madde fonksiyonu gösterdiği ortaya çıkmıştır. Dikkate değer diğer bir sonuç ise, parametre karşılaştırma yöntemi ile yapılan DMF analizi sonucunda 36 madde 0,01 manidarlık düzeyinde değişen madde fonksiyonu gösterirken olabilirlik oranı testi model karşılaştırma yöntemine göre 49 maddenin 0,01 anlamlılık düzeyinde kültürler arasında ölçme değişmezliğine sahip olmadıkları görülmüştür. Olabilirlik oranı testi model karşılaştırma yönteminin maddelerin gruplardan elde edilen parametre değerleri arasındaki farkı parametre karşılaştırma yöntemine oranla daha yüksek tahminlediği gözlenmiştir. Diğer taraftan bu sonuç beklenmeyen bir sonuç değildir, çünkü Childs ve Chen (1999) tarafından açıklandığı gibi Multilog ve Parscale ile elde edilen madde parametre tahminleri benzer olmamakta ve her iki programın parametre tahminlerini birbirine dönüştüren işlemler yapıldıktan sonra ancak kıyaslama yapmak mümkün olabilmektedir.

PIRLS 2001 projesi kapsamında bulunan sekiz okuma parçasına ait 98 sorunun 53 (%54) tanesinin DMF analizi sonucunda hem parametre karşılaştırma yönteminde hem de olabilirlik oranı testi model karşılaştırma yönteminde kültürler arasında değişen madde fonksiyonu içerdiği ortaya çıkmıştır. Bu oranın çok yüksek çıkması şaşırtıcı olmasına rağmen Bertrand ve Boiteau'nun 2003 yılında tamamladıkları çalışma

sonuçları ile uyushmaktadır. Araştırmacılar 1995 ve 1999 TIMSS uygulamalarında matematik sorularının yaklaşık %80'inin ABD ve Japonya karşılaştırmalarında kültürler arasında DMF içerdığını bildirmişlerdir.

Hambleton (2005) tarafından belirtildiği gibi, kültürlerarası eşdeğerliği sağlanamayan testlerin bu durumu göz ardı edilerek kullanılması hatalı karşılaştırmalar yapılmasına yol açmaktadır. Bu tür testlerden elde edilen ölçme sonuçlarına dayalı olarak yapılan başarı sıralamalarının da hata içermesi mümkün olabilmektedir. PIRLS 2001 uygulama sonuçlarına göre ülkemizin 35 katılımcı ülke arasında başarı sıralaması açısından 28 sırada yer almış olmasının gerçek durumu ne derece yansıttığı, bu araştırmadan elde edilen sonuçlara göre tartışılır görünmektedir.

Öneriler

PIRLS-2001 projesinin başlangıcında kültürel adaptasyon ve çeviri işlemlerinden sonra her ülkede, pilot uygulama (alan uygulaması) öncesi ve sonrası maddelerde ve okuma parçalarında gerekli görülen düzeltmeler yapılmış ölçeklerinin son formları oluşturulmuş ve asıl uygulama için hazırlanmış olduğu belirtilmektedir. Bundan sonra yapılacak olan benzer projelerde, son formlar oluşturulmadan önce pilot uygulamalardan elde edilen veriler üzerinden referans kültür ile her bir odak kültür arasında DMF analizlerinin yapılması test ve madde istatistiklerinin değerlendirilmesi açısından faydalı olabilir.

Testlerdeki maddelerin önerildiği gibi dört ayrı kavrama süreci ile belirtilen boyutlarla iddia edilen ilişkileri vermediği anlaşılmış olduğundan, testlerin yapılarını doğru tanımlamak amacı ile hem referans grubun verileri için hemde odak grubun verileri için temel bileşenler analizi yapılarak her iki kültür için benzer olan boyutlar ortaya çıkarılabilir. Bu işlem yapıldıktan sonra çok örneklemlili doğrulayıcı faktör analizi yapılarak testlerin kültürler arasında yapı geçerliği açısından denklik gösterip göstermediğinin incelenmesinin daha doğru bir yaklaşım olacağı düşünülmektedir.

Bu arařtırmada yapılan karřılařtırmaya benzer bir karřılařtırma kaynak kltr ile PIRLS uygulamasına katılan diğerk hedef kltrler arasında tekrar edilebilir. Bylece hem kullanılan yntemlerin diğerk karřılařtırma sonuları ile tutarlı olup olmadıėı hem de PIRLS kapsamındaki testler iin bu arařtırma sonularına gre ortaya ıkan sonuların kesinliėi hakkında bilgilere ulařılabilir.

Bu arařtırma kapsamında DMF analizleri madde dzeyinde yapılmıřtır. Testlerin deėiřen test fonksiyonu ierip iermediėi de ayrıca arařtırılabilir. DMF ieren maddelerin aynı zamanda deėiřen test fonksiyonuna ne kadar etki yaptıėı uygun analiz teknikleri kullanılarak ortaya ıkarılabilir.

Olabilirlik oranı testi model karřılařtırma ynteminde baė maddelerinin seimi iin yapılan tekrarlayıcı madde iřlemleri sırasında her bir iterasyonda madde sayısı kadar geniř model kurulmakta, her yeni model iin veri dosyası ve kalan maddeler yeniden tanımlanmakta, program iin gerekli olan komutlar yeniden yazılmak zorunda kalındıėından sz konusu iřlemler ok zaman ve emek gerektirmekte ayrıca modeller ve iterasyonlar arası geiřlerde hata yapma riskine yol amaktadır. Diğerk taraftan PARSCALE programı ile yapılan parametre karřılařtırma ynteminde sadece bir komut dosyası kullanılarak ok kısa srede sonulara ulařılabilmektedir. Ayrıca bu program, deėiřik řekillerde puanlanmış ok kategorili maddeler iin uygun olup, arařtırmacıya istediėi sayıda madde parametresi ile alıřma olanaėı tanımakta, istediėi parametreyi karřılařtırma yapılan gruplar arasında sınırlama kolaylıėı saėlamakta, χ^2 testi ile deėiřen madde parametreleri farkının test edilmesine olanak saėlamakta ve sonuların kolay yorumlanabilir olması gibi avantajlar saėlamaktadır. Bu nedenle zellikle madde sayısının fazla olduėu testler iin yapılacak olan DMF analizlerinde PARSCALE programı ile yapılan parametre karřılařtırma ynteminin kullanılması arařtırmacıya kolaylık saėlayabilir.

KAYNAKÇA

- ADAMS, R. J. and Rowe, K. J., (1988). *Item Bias* in Keeves, J.P.(ed.) Educational research, methodology, and measurement: An international handbook. Oxford: Pergamon Press.
- ALKAN, M (2005). *Uluslararası Okuma Becerilerinde Gelişim Projesi (PIRLS)'nde Türk Öğrencilerin Başarılarını etkileyen Faktörler*.Yayınlanmamış Yüksek Lisans Tezi. Eğitim Bilimleri Bölümü, Hacettepe üniversitesi.
- ANDRICH, D. (1978). A Rating Formulation for Ordered Response categories. *Psychometrika*, 43, 561-573.
- ANGOFF, W. H., (1972). A Technique For The Investigation Of Cultural Differences. *Paper Presented at The American Psychological Association Convention, Honolulu, HI.*
- BAYKUL, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara. ÖSYM Yayınları.
- BERTRAND, R. And Boiteau, N. (2003). Comparing the Stability of IRT-Based and Non IRT-Based DIF Methods in Different Cultural Context Using TIMSS Data. *ERIC Report-Research (143). EDRS. ED 476 924. 20p.*
- BOCK, R. D. (1972). Estimating Item Parameters and Latent Ability When Responses Are Scored in Two or More Nominal Categories. *Psychometrika*, 37, 29-51.
- BOLLEN, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.
- BRISLIN, R. (Ed.) (1976). *Translation: Application and research*. New York: John Wiley.
- BRISLIN, R. (1980). Translation and Content Analysis of Oral and Written Materials. *In H. C. Triandis and J. W. Berry (Eds.), Handbook of Cross-cultural Psychology: Methodology (Vol. 2)*. Boston: Allyn and Bacon.

- BRISLIN, R. (1970). Back-Translation for Cross-cultural Research. *Journal of Cross-Cultural Psychology*. 1, 185-216.
- BRISLIN, R. (1976). Translation Research and Its Applications: An Introduction. In R. Brislin (Ed.), *Translation: Applications and Research*. New York: Wiley/Halstead.
- BYRNE, B. M. (1994) *Structural Equation Modeling with EQS and EQS/Windows: Basic Concepts, Applications and Programming*. California: SAGE Publications, Inc.
- CAMILLI, G., Dhepard, L. A., (1994). *Methods for Identifying Biased test Items*. Vol4. California: Sage Publications, Inc.
- CANDELL, G. L., and Roznowski, M. (1984) Using IRT to Establish Equivalence Across U.S. and Canadian Subpopulations. *Paper Presented at American Psychological Association Annual Meeting, Toronto, Canada*.
- CATTELL, R. B. (1957) *Personality And Motivation Structure And Measurement*. New York: Harcourt Brace Jovanovich.
- CHILDS, R.A. and Chen, W. (1999). Obtaining Comparable Item Parameter Estimates in MULTILOG and PARSCALE for Two Polythomous IRT Models. *Applied Psychological Measurement*, Vol. 23 No. 4, 371-379.
- CROCKER, L. And Algina, J., (1986) *Introduction to Classical and Modern Test Theory*. Orlando: Harcourt Brace Jovanovich, Inc.
- DEVINE, P. J. and Raju N. S., (1982) Extent of Overlap Among Four Item Bias Methods. *Educational and Psychological Measurement*, 42, pp. 1049-1066.

- ELLIS, Barbara. B., (1989) Differential Item Functioning: Implications for Test Translations. *Journal of Applied Psychology*. 74, 912-921.
- GEORGE, D. And Mallery, P. (2001). *SPSS for Windows A simple Guide and Reference*. Third Edition. Boston, Allyn and Bacon.
- GONZALEZ, E.J. (Ed.). and Kennedy, A.M. (Ed.), (2003) "*PIRLS 2001 User Guide for the International Database*". (IEA) International Study Center, Boston College.
- GORDON, L. V., and Kikuchi, A. (1966) American Personality Tests in Cross-cultural Research-A caution. *Journal of Cross-Cultural Psychology*, 69, 179-183.
- HAMBLETON, R K., Swaminathan, H. and Rogers, H. J., (1991). *Fundamentals of Item Response Theory*. London: Sage Publication.
- HAMBLETON, R. K. (1994) Guidelines For Adapting Educational and Psychological Tests: A Progress Report. *European Journal of Psychological Assessment*. 10, 229-244.
- HAMBLETON, R. K., Merenda, P.F. and Spielberger, C.D., (2005) *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. New Jersey: Lawrence Erlbaum Associates, Publishers.
- HOLLAND, P. W., Wainer, H., (1993). *Differential Item Functioning*. Hillside,. New Jersey: Lawrence Erlbaum Associates, Publishers.
- HUANG, C.D., Church, A.T. and Katigbak, M.S. (1997). Identifying Cultural Differences in Items and Traits: Differential Item Functioning in the NEO Personality Inventory. *Journal of Cross-Cultural Psychology*, Vol. 28, No. 2, 192-248.

- HUI, C., H. And Triandis, H., C., (1985) Measurement in Cross-Cultural Psychology. *Journal of Cross-Cultural Psychology*. 16, 131-152.
- HUI, C., H. And Triandis, H., C., (1983). Multistrategy Approach to Cross-cultural Research: The Case of Locus of Control. *Journal of Cross-Cultural Psychology*, 14, 65-83.
- HULIN, C. L., Drasgow, F., and Parsons, C. K. (1983) Psychometric Equivalence of a Translation of the JDI into Hebrew. *Journal of Applied Psychology*, 71, 83-94.
- HULIN, C. L., and Mayer, L. M., (1986) Psychometric Equivalence of a Translation of the JDI into Hebrew. *Journal of Applied Psychology*, 71, 83-94.
- HULIN, C., L. (1987) A Psychometric Theory of Evaluations of Item and Scale Translations. *Journal of Cross-Cultural Psychology*. 18, 115-142.
- IEA (International Association For the Evaluation of Educational Achievement). (n.d.). TIMSS 1999 Publications. Retrieved November 15, 2001 from <http://isc.bc.edu/timss1999i/database.html>
- IRVINE, S. H., and Carroll, W. K., (1980) Testing and Assessment Across Cultures: Issues in Methodology and Theory. In H. C. Triandis and J. W. Berry (Eds.), *Handbook of Cross-cultural Psychology: Methodology (Vol. 2)*. Boston: Allyn and Bacon.
- JENNRICH, R. I., (1970) An Asymptotic Chi-square Test for the Equality of Two Correlation Matrices. *Journal of American Statistical Association*., 65,-904-912.
- JÖRESKOG, K. G. (1971) Simultaneous Factor Analysis in Several Populations. *Psychometrika*, 36, 409-426.

- JÖRESKOG, K. G. and Sörbom, D., (1993) *Lisrel 8: Structural Equation Modeling With The Simplis Command Language*, Hillsdale, NJ, Lawrence Erlbaum Associates Publishers.
- KAPLAN, D., (2000) *Structural Equation Modeling: Foundations and Extensions*. California: Sage Publications Inc.
- KAZELSKIS, R., Thames, D. And Reeves, C. (2004). The Elementary Reading Attitude Survey: Factor Invariance , Across Gender and Race. *Reading Psychology*, 25: 111-120.
- KLINE, R. B., (1998) *Principles and Practice of Structural Equation Modeling*. New York: The Guilford Press.
- KIM, S.-H. and Cohen, A. S. (1995). A Comparison of Lord's Chi-square, Raju's Area Measures, and the Likelihood ratio Test on Detection of Differential Item Functioning. *Applied Measurement in Education*, 8 (4), 291-312.
- KIM, S.-H., Cohen, A. S., DiStefano, C.A. and Kim, S. (1998). An Investigation of the Likelihood Ratio Test for Detection of Differential Item Functioning Under the Partial Credit Model. *ERIC Reports- Evaluative. EDRS. ED 442 837*. 23 p.
- KORKMAZ, M. (2005). *Madde Cevap Kuramına Dayalı Olarak Çok kategorili Maddelerde Madde ve Test Yanlılığının (İşlevsel Farklılığın) İncelenmesi*. Yayınlanmamış Doktora Tezi. Psikoloji Bölümü, Ege Üniversitesi.
- LAWLEY, D. N. (1943). On Problems Connected with Item Selection and Test Construction. *Proceedings of the Royal Society of Edinburgh.*, 61, 273-287.
- LINN, R. L., Levine, M. V., Hastings, C. N., and Wardrop, J. L. (1981) Item Bias in a Test of Reading Comprehension. *Applied Psychological Measurement*, 5, 159-173.

- LORD, F. M., (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum.
- LORD, F.M., and Novick, M.R. (1968). *Statistical Theories of Mental test Scores*. Reading, MA: Addison-Wesley.
- LUBKE, G. H., Muthén, B. O. (2004) Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons. *Structural Equation Modeling*, 11 (4), p. 514-534.
- MASTERS, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika*, 47, 149-174.
- MEADE, A.W. and Lautenschlager, G.J. (2004). A Comparison of Item Response Theory and Confirmatory Factor analytic Methodologies for Establishing Measurement Equivalence/Invariance. *Organizational Research Methods*, Vol. 7, No. 4, 361-388.
- MELLENBERG, G. J., (1983). Conditional Item Bias Methods. In S. H. Irvine and W. J. Barry (Eds), *Human Assessment and Cultural Factors* (pp. 293-302), Newyork: Plenum Pres.
- MELLENBERG, G. J., (1989). Item Bias and Item Response Theory. *International Journal of Educational Research: Applications of Item Response Theory*. 13,2, pp.123-144.
- MEREDITH, W. (1993). Measurement Invariance, Factor Analysis, and Factorial Invariance. *Psychometrika*, 58, 521-543.
- MULLIS, I.V.S., Martin, M.O., Gonzalez, E.J. and Kennedy, A.M., (2003) "*PIRLS 2001 International Report*." (IEA) International Study Center, Boston College.
- MURAKI, E. (1992). A Generalized Partial Credit Model: Applications of an EM Algoritm. *Applied Psychological Measurement*, Vol. 16, 159-176.
- MURAKI, E. And Bock, R. D. (1996). *PARSCALE (V4.1). Parameter Scaling of Rating Data*. Chicago, IL: Scientific Software, Inc.

- OSGOOD, C. E., May, W. H., and Miron, M. S., (1975) *Cross-cultural Universals of Affective Meaning*. Urbana, IL: University of Illinois Pres.
- OSGOOD, C. E. (1967) Cross-Cultural Comparability in Attitude Measurement via Multilingual Semantic Differentials. In M. Fishbein (Ed.), *Readings in Attitude Theory And Measurement*. New York: John Wiley.
- PISA Web Site. www.pisa.oecd.org
- POORTINGA, Y. H. (1975) Some Implications of Three Different Approaches to Intercultural Comparison. In J. W. Berry and W. J. Lonner (Eds.), *Applied Cross-cultural Psychology* (pp. 329-332). Amsterdam: Swets and Zeitlinger.
- RAJU, N. S. (1988) The Area Between Two Item Characteristic Curves. *Psychometrika*, 53,495-502.
- RECKASE, M. D. (1979). Unifactor Latent Trait Models Applied to Multifactor Tests: Results and Implications. *Journal of Educational Statistics*, 4, 207-230.
- REISE, S.P., Wideman, K.F. and Pugh, R.H. (1993) Confirmatory Factor Analysis and Item Response Theory: Two Approaches for Exploring Measurement Invariance. *Psychological Bulletin*. Vol. 114, No. 3, 552-566.
- REISE, S.P., Smith, L. And Furr, R.M. (2001). Invariance on the PI-R Neuroticism Scale. *Multivariate Behavioral Research*, 36 (1), 83-110.
- RODNEY. G. L. and Drasgow, F., (1990). Evaluation of Two Methods for Estimating Item Response Theory Parameters When Assessing Differential Item Functioning. *Journal of Applied Psychology*. 75,2,pp.164-174.
- RUDNER, L., Getson, P. R. and Knight, D. L. (1980). Biased Item Detection Techniques. *Journal of Educational Statistics*. 5, pp. 213-233.
- SALZBERGER, T., Sinkovics, R.R. and Schlegelmilch, B.B (2000). Data Equivalence in Cross-Cultural Research: A Comparison of Classical Test theory and Latent trait Theory Based approaches. *Australasian Marketing Journal*, Vol. 7. No. 2, 23-38.

- SCHUMACKER, R. E., Lomax, R. G. (2004) *A Beginner's Guide to Structural Equation Modeling*. Second Edition. New Jersey. Lawrence Erlbaum Associates,
- SHEPARD, L. A., Camilli, G. and Williams, D. M. (1984). Validity of Approximation Techniques for Detecting Item Bias. *Journal of Educational Measurement*. 22, 2, pp77-105.
- SIRECI, S., G., (1997) Problems and Issues in Linking Assessment Across Languages. *Educational Measurement : Issues and Practice*. Spring, 12-20.
- SIRECI, S., G., Berberoğlu, G., (2000) Using Bilungial Respondents to Evaluate Translated-Adapted Items. *Applied Measurement in Education*. 13 (3), 229-248.
- SMITH, L.L. (2002). On the Usefulness of Item Bias Analysis to Personality Psychology. *Personality and Social Psychology Bulletin*. Vol. 28, No. 6, 754-763.
- SOMER, O. (2004). Gruplararası Karşılaştırmalarda Ölçek Eşdeğerliğinin İncelenmesi: Madde ve Test Fonksiyonlarının Farklılaşması. *Türk Psikoloji Dergisi*, 19 (53), 69-82.
- SÜMER, N. (2000). Yapısal eşitlik Modelleri: Temel Kavramlar ve Örnek Uygulamalar. *Türk Psikoloji Yazıları*, 3, (6) 49-74.
- STEVENS, J. P. (2002) *Applied Multivariate Statistics for the Social Sciences*. New Jersey. Lawrence Erlbaum Associates, Inc. Fourth Edition.
- ŞENCAN, H. (2005) *Sosyal ve Davranışsal Ölçmelerde Güvenilirlik ve Geçerlilik*. Ankara. Seçkin Yayıncılık.
- TABACNICK, B. G. And Fidel, L. S., (2001) *Using Multivariate Statistics*. Boston: Allyn and Bacon IV. Ed.

- TERESI, J.A., Kleinman, M., and Welikson, O-K. (2000). Modern Psychometric Methods for Detection of Differential Item Functioning: Application to Cognitive Assessment Measures. *Statistics in Medicine.*, 19, p: 1651-1683.
- THISSEN, D. (1992). *MULTILOG: Multiple Category Item Analysis and Test Scoring Using Item response Theory (V7.03)*. Chiago: Scientific Software International, Inc.
- THISSEN, D., Steinberg, L. And Wainer, H. (1993). *Detection of Differential Item Functioning Using the Parameters of Item Response Model*. İç. P.W. Holland and H. Wainer (ed). *Differential Item Functioning* (67-113). Hillside, NJ: Erlbaum.
- TIMSS (2003). *1999 Türkiye Raporu*. Ankara, MEB.
- Du TOIT, M. (2003). *IRT from SSI: Bilog-MG, Multilog, Parscale, Testfact*. Scientific Software International, Inc.
- TRIANDIS, H. C. (1972). *The Analysis of Subjective Culture*. New York: John Wiley.

ÖZGEÇMİŞ**Kişisel Bilgiler**

Adı Soyadı : Tuncay Öğretmen
Doğum Yeri ve Tarihi : Hekimhan - 20.12.1961

Eğitim Durumu

Lisans Öğrenimi : Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme Anabilim Dalı

Yüksek Lisans Öğrenimi : Orta Doğu Teknik Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü

Bildiği Yabancı Diller : İngilizce

Bilimsel Faaliyetleri :

İş deneyimi

Stajlar :
Projeler :
Çalıştığı Kurumlar : İçişleri Bakanlığı
Danıştay Başkanlığı
Ege Üniversitesi Eğitim Fakültesi
Hacettepe Üniversitesi Eğitim Fakültesi

İletişim

E-Posta Adresi :ogretmen@hacettepe.edu.tr
togretmen@gmail.com

Tarih : 31.05.2006

EKLER

Ek 1

LISREL 8.54 Programında Sıfır Kodlu PIRLS Testi İçin Oluşturulan Çok Örneklemli Doğrulayıcı Faktör Analizi Komut Dosyaları

MODEL A

Group 1: Testing Equality Of Factor Structures

Model A: Factor Loadings, Factor Correlation, Error Variances Invariant

Observed Variables: H1 H2 H3 H4 H5 H6 H7 H8 H9 H10 H11 P1 P2 P3 P4 P5
P6 P7 P8 P9 P10 P11 P12 P13

Covariance Matrix from File USA.COV

Sample Size = 308

Latent Variables: Acikbil Cikarim Yorum Degerlen

Relationships:

H1 H2 H4 P1 P2 P3 P5 P9 = Acikbil

H5 H7 P10 = Cikarim

H6 H8 H9 H10 H11 P6 P7 P8 P11 P12 P4 = Yorum

H3 P13 = Degerlen

Group 2: Testing Equality Of Factor Correlations

Covariance Matrix from File TURK.COV

Sample Size = 422

Path Diagram

MODEL B

Group 1: Testing Equality Of Factor Structures

Model B : Factor Correlation and Error Variances Invariant

Observed Variables: H1 H2 H3 H4 H5 H6 H7 H8 H9 H10 H11 P1 P2 P3 P4 P5
P6 P7 P8 P9 P10 P11 P12 P13

Covariance Matrix from File USA.COV

Sample Size = 308

Latent Variables: Acikbil Cikarim Yorum Degerlen

Relationships:

H1 H2 H4 P1 P2 P3 P5 P9 = Acikbil

H5 H7 P10 = Cikarim

H6 H8 H9 H10 H11 P6 P7 P8 P11 P12 P4 = Yorum

H3 P13 = Degerlen

Group 2: Testing Equality Of Factor Correlations

Covariance Matrix from File TURK.COV

Relationships:

H1 H2 H4 P1 P2 P3 P5 P9 = Acikbil

H5 H7 P10 = Cikarim

H6 H8 H9 H10 H11 P6 P7 P8 P11 P12 P4 = Yorum

H3 P13 = Degerlen

Sample Size = 422

Path Diagram

End of Problem

MODEL C

Group 1: Testing Equality Of Factor Structures

Model C: Factor Correlation Invariant

Observed Variables: H1 H2 H3 H4 H5 H6 H7 H8 H9 H10 H11 P1 P2 P3 P4 P5
P6 P7 P8 P9 P10 P11 P12 P13

Covariance Matrix from File USA.COV

Sample Size = 308

Latent Variables: Acikbil Cikarim Yorum Degerlen

Relationships:

H1 H2 H4 P1 P2 P3 P5 P9 = Acikbil

H5 H7 P10 = Cikarim

H6 H8 H9 H10 H11 P6 P7 P8 P11 P12 P4 = Yorum

H3 P13 = Degerlen

Group 2: Testing Equality Of Factor Correlations

Covariance Matrix from File TURK.COV

Relationships:

H1 H2 H4 P1 P2 P3 P5 P9 = Acikbil

H5 H7 P10 = Cikarim

H6 H8 H9 H10 H11 P6 P7 P8 P11 P12 P4 = Yorum

H3 P13 = Degerlen

Sample Size = 422

Set the Error Variances of H1 - P13 free

Path Diagram

End of Problem

MODEL D

Group 1: Testing Equality Of Factor Structures

Model D: Factor Loadings and Factor Correlation Invariant

Observed Variables: H1 H2 H3 H4 H5 H6 H7 H8 H9 H10 H11 P1 P2 P3 P4 P5
P6 P7 P8 P9 P10 P11 P12 P13

Covariance Matrix from File USA.COV

Sample Size = 308

Latent Variables: Acikbil Cikarim Yorum Degerlen

Relationships:

H1 H2 H4 P1 P2 P3 P5 P9 = Acikbil

H5 H7 P10 = Cikarim

H6 H8 H9 H10 H11 P6 P7 P8 P11 P12 P4 = Yorum

H3 P13 = Degerlen

Group 2: Testing Equality Of Factor Correlations

Covariance Matrix from File TURK.COV

Sample Size = 422

Set the Error Variances of H1 - P13 free

Path Diagram

End of Problem

Ek 2

**PARSCALE Programında Parametre Karşılaştırma Yöntemi ile Yapılan Değişen
Madde Fonksiyonu Analizi İçin Yazılan Komut Dosyası ve Kısaltılmış
Analiz Sonuçları**

HARE PARTIAL CREDIT ITEMS - DIF ANALYSIS USING PARTIAL CREDIT MODEL

>COMMENTS

Forced the common slope closer to 1.

```
>FILES DFNAME='hare.DAT';
>INPUT NID=4, MGROUP=2, NFMT=1, NTOTAL=11;
(4A1,1X,1A1,1X,11A1)
>TEST TNAME=HARE, NBLOCK=11, INAMES=(M1,M2,M3,M4,M5,M6,M7,M8,M9,M10,M11);
>BLOCK1 BNAME=MC, NITEMS=1, NCAT=2, ORIGINAL=(0,1), MOD=(1,2);
>BLOCK2 BNAME=MC, NITEMS=1, NCAT=2, ORIGINAL=(0,1), MOD=(1,2);
>BLOCK3 BNAME=OE, NITEMS=1, NCAT=3, ORIGINAL=(0,1,2), MOD=(1,2,3);
>BLOCK4 BNAME=OE, NITEMS=1, NCAT=2, ORIGINAL=(0,1), MOD=(1,2);
>BLOCK5 BNAME=MC, NITEMS=1, NCAT=2, ORIGINAL=(0,1), MOD=(1,2);
>BLOCK6 BNAME=MC, NITEMS=1, NCAT=2, ORIGINAL=(0,1), MOD=(1,2);
>BLOCK7 BNAME=OE, NITEMS=1, NCAT=3, ORIGINAL=(0,1,2), MOD=(1,2,3);
>BLOCK8 BNAME=OE, NITEMS=1, NCAT=2, ORIGINAL=(0,1), MOD=(1,2);
>BLOCK9 BNAME=OE, NITEMS=1, NCAT=3, ORIGINAL=(0,1,2), MOD=(1,2,3);
>BLOCK10 BNAME=OE, NITEMS=1, NCAT=4, ORIGINAL=(0,1,2,3), MOD=(1,2,3,4);
>BLOCK11 BNAME=MC, NITEMS=1, NCAT=2, ORIGINAL=(0,1), MOD=(1,2);
>MGROUP GNAME=('ABD','TRK'), GCODE=(A,B), DIF=(0,1,0,0);
>CALIB LOGISTIC, PARTIAL, ITEMFIT=1, NQPT=25, CYCLES=100, NEWTON=2, CRIT=0.001,
ESTORDER, SPRIOR, PRIORREAD;
>PRIORS SMU=(1.0(0)11), SSIGMA=(0.0001(0)11);
>SCORE ;
```

Kısaltılmış Program Çıktısı

PARSCALE V4.1

MAXIMUM LIKELIHOOD ITEM ANALYSIS AND TEST SCORING: POLYTOMOUS MODEL

[PHASE 2]

CURRENT DATE: 1-30-2006
CURRENT TIME: 00:35:47

*** POLYTOMOUS MODEL ITEM ANALYSER ***

*** PHASE 2 ***

HARE PARTIAL CREDIT ITEMS - DIF ANALYSIS USING PARTIAL CREDIT MODEL

MAINTTEST: HARE

```
*****
CALIBRATION OF MAINTTEST
HARE
*****
```

[E-M CYCLES] PARTIAL CREDIT MODEL

STEP AND ITEM PARAMETERS AFTER CYCLE 0

LARGEST CHANGE= 0.000
-2 LOG LIKELIHOOD = 33934.246

LARGEST CHANGE= 0.086 (-1.309-> -1.395) at Location of Item: 4 M4
-2 LOG LIKELIHOOD = 33882.193

STEP AND ITEM PARAMETERS AFTER CYCLE 2

LARGEST CHANGE= 0.069 (-0.006-> 0.063) at Location of Item: 3 M3

MULTIPLE GROUP MODEL [DIF(TREND) MODEL]

[GROUP: 1 ABD]
ITEM BLOCK 1 MC

SCORING FUNCTION : 1.000 2.000
STEP PARAMTER : 0.000 0.000
S.E. : 0.000 0.000
ITEM BLOCK 2 MC

SCORING FUNCTION : 1.000 2.000
STEP PARAMTER : 0.000 0.000
S.E. : 0.000 0.000
ITEM BLOCK 3 OE

SCORING FUNCTION : 1.000 2.000 3.000 4.000
STEP PARAMTER : 0.000 -0.113 0.480 -0.367
S.E. : 0.000 0.040 0.046 0.048
ITEM BLOCK 11 MC

SCORING FUNCTION : 1.000 2.000
STEP PARAMTER : 0.000 0.000
S.E. : 0.000 0.000

ITEM	BLOCK	SLOPE	S.E.	LOCATION	S.E.	GUESSING	S.E.
M1	1	1.000	0.000	-1.841	0.101	0.000	0.000
M2	2	1.000	0.000	-1.630	0.097	0.000	0.000
M3	3	1.000	0.000	0.291	0.065	0.000	0.000
M4	4	1.000	0.000	-1.372	0.088	0.000	0.000
M5	5	1.000	0.000	-1.568	0.094	0.000	0.000
M6	6	1.000	0.000	-0.717	0.076	0.000	0.000
M7	7	1.000	0.000	-0.249	0.062	0.000	0.000
M8	8	1.000	0.000	-0.494	0.074	0.000	0.000
M9	9	1.000	0.000	-0.239	0.061	0.000	0.000
M10	10	1.000	0.000	0.439	0.060	0.000	0.000
M11	11	1.000	0.000	-0.696	0.073	0.000	0.000

[GROUP: 2 TRK]
ITEM BLOCK 1 MC

SCORING FUNCTION : 1.000 2.000
STEP PARAMTER : 0.000 0.000

S.E. : 0.000 0.000
ITEM BLOCK 2 MC

SCORING FUNCTION : 1.000 2.000 3.000 4.000
STEP PARAMTER : 0.000 -0.113 0.480 -0.367
S.E. : 0.000 0.040 0.046 0.048
ITEM BLOCK 11 MC

SCORING FUNCTION : 1.000 2.000
STEP PARAMTER : 0.000 0.000
S.E. : 0.000 0.000

ITEM	BLOCK	SLOPE	S.E.	LOCATION	S.E.	GUESSING	S.E.
M1	1	1.000	0.000	-1.983	0.065	0.000	0.000
M2	2	1.000	0.000	-0.944	0.054	0.000	0.000
M3	3	1.000	0.000	0.063	0.045	0.000	0.000
M4	4	1.000	0.000	-1.390	0.054	0.000	0.000
M5	5	1.000	0.000	-1.216	0.054	0.000	0.000
M6	6	1.000	0.000	-0.899	0.052	0.000	0.000
M7	7	1.000	0.000	-0.537	0.042	0.000	0.000
M8	8	1.000	0.000	-0.134	0.051	0.000	0.000
M9	9	1.000	0.000	-0.306	0.043	0.000	0.000
M10	10	1.000	0.000	0.330	0.041	0.000	0.000
M11	11	1.000	0.000	-1.059	0.052	0.000	0.000

MULTIPLE GROUP MODEL [DIF(TREND) MODEL]

CONTRAST OF ITEM LOCATIONS:

GROUP 2: TRK MINUS REFERENCE GROUP 1: ABD

ITEM	BLOCK	CONTRAST (S.E.)	STD (PROB.)
M1	1	-0.142 (0.120)	-1.181 (0.119)
M2	2	0.686 (0.111)	6.171 (0.000)
M3	3	-0.228 (0.079)	-2.880 (0.002)
M4	4	-0.018 (0.103)	-0.176 (0.430)
M5	5	0.352 (0.108)	3.247 (0.001)
M6	6	-0.183 (0.092)	-1.989 (0.023)
M7	7	-0.289 (0.075)	-3.841 (0.000)
M8	8	0.360 (0.089)	4.025 (0.000)

M9	9	-0.067	-0.901
		(0.075)	(0.184)
M10	10	-0.109	-1.512
		(0.072)	(0.065)
M11	11	-0.363	-4.059
		(0.089)	(0.000)

CHI-SQUARES OF ITEM LOCATION CONTRASTS:

ITEM	BLOCK	CHI-SQRS	D.F. PROB.
M1	1	1.396	1. 0.236
M2	2	38.083	1. 0.000
M3	3	8.296	1. 0.004
M4	4	0.031	1. 0.837
M5	5	10.543	1. 0.001
M6	6	3.956	1. 0.044
M7	7	14.753	1. 0.000
M8	8	16.200	1. 0.000
M9	9	0.812	1. 0.371
M10	10	2.287	1. 0.126
M11	11	16.479	1. 0.000
TOTAL		112.836	11. 0.000

SUMMARY STATISTICS OF PARAMETER ESTIMATES

1 GROUP NAME: ABD

PARAMETER	MEAN	STN DEV	N
SLOPE	1.000	0.000	11
LOG(SLOPE)	0.000	0.000	11
THRESHOLD	-0.734	0.781	11
GUESSING	0.000	0.000	0

2 GROUP NAME: TRK

PARAMETER	MEAN	STN DEV	N
SLOPE	1.000	0.000	11
LOG(SLOPE)	0.000	0.000	11
THRESHOLD	-0.734	0.689	11
GUESSING	0.000	0.000	0

187616 BYTES OF NUMERICAL WORKSPACE USED OF 8192000 AVAILABLE IN PHASE 2
 556 BYTES OF CHARACTER WORKSPACE USED OF 2048000 AVAILABLE IN PHASE 2
 NORMAL END

MULTILOG Programında Olabilirlik Oranı Testi Model Karşılaştırma Yöntemi ile Yapılan Değişen Madde Fonksiyonu Analizi İçin Yazılan Komut Dosyası ve Kısaltılmış Analiz Sonuçları

```

MULTILOG for Windows 7.00.2327.2 STEP 4
Created on: 3 June 2005, 11:28:18
>PROBLEM      RANDOM,
                INDIVIDUAL,
                DATA = 'hare.dat',
                NITEMS = 22,
                NGROUPS = 2,
                NEXAMINEES = 2209,
                NCHARS = 4;
>TEST         ITEMS=1, L1;
>TEST         ITEMS=2, L1;
>TEST         ITEMS=3, NOMINAL, NC=3, HIGH=3;
>TEST         ITEMS=4, L1;
>TEST         ITEMS=5, L1;
>TEST         ITEMS=6, L1;
>TEST         ITEMS=7, NOMINAL, NC=3, HIGH=3;
>TEST         ITEMS=8, L1;
>TEST         ITEMS=9, NOMINAL, NC=3, HIGH=3;
>TEST         ITEMS=10, NOMINAL, NC=4, HIGH=4;
>TEST         ITEMS=11, L1;
>TEST         ITEMS=12, L1;
>TEST         ITEMS=13, L1;
>TEST         ITEMS=14, NOMINAL, NC=3, HIGH=3;
>TEST         ITEMS=15, L1;
>TEST         ITEMS=16, L1;
>TEST         ITEMS=17, L1;
>TEST         ITEMS=18, NOMINAL, NC=3, HIGH=3;
>TEST         ITEMS=19, L1;
>TEST         ITEMS=20, NOMINAL, NC=3, HIGH=3;
>TEST         ITEMS=21, NOMINAL, NC=4, HIGH=4;
>TEST         ITEMS=22, L1;
>TMATRIX      ITEMS= (3,7,9,10,14,18,20,21), AK, POLYNOMIAL;
>EQUAL        ITEMS= (3,7,9,10,14,18,20,21), AK=1;
>FIX          ITEMS= (3,7,9,14,18,20), AK=2, VALUE=0.0;
>FIX          ITEMS= (10,21), AK=(2,3), VALUE=0.0;
>TMATRIX      ITEMS= (3,7,9,10,14,18,20,21), CK, TRIANGLE;
>EQUAL        AJ, ITEMS = (12,17),
                WITH = (1,6);
>EQUAL        AK=(1,2,3), ITEMS = (18), WITH= (7);
>EQUAL        BJ, ITEMS = (12,17),
                WITH = (1,6);
>EQUAL        CK=(1,2,3), ITEMS= (18), WITH= (7);
>FIX          MU, GROUPS= 1, VALUE=0.0;
>FIX          SD, GROUPS= 1, VALUE=1.0;
>FIX          SD, GROUPS=2, VALUE=1.0;
>ESTIMATE     NCYCLES=100;
>SAVE ;
>END ;
4
1234
11111111111111111111
22222222222222222222
0030003033000300030330
0000000004000000000040
(4A1, 1X, I1 ,1X, 22A1)

```

DATA FILE NAME IS
HARE.DAT

```

1234
  VECTOR OF CATEGORIES FOR CODE=1

11111111111111111111111111111111
  VECTOR OF CATEGORIES FOR CODE=2

22222222222222222222222222222222
  VECTOR OF CATEGORIES FOR CODE=3

0030003033000300030030330
  VECTOR OF CATEGORIES FOR CODE=4

00000000004000000000000040
(4A1, 1X, I1, 1X, 22A1)

```

MULTILOG--FOR MULTIPLE CATEGORICAL ITEM RESPONSE DATA--VERSION 7.0.3
MULTILOG for Windows 7.00.2327.2 STEP 4
Created on: 3 June 2005, 11:28:18

NUMBER OF LINES IN THE DATA FILE: 2209
NUMBER OF CATEGORICAL-RESPONSE ITEMS: 22
NUMBER OF CONTINUOUS-RESPONSE ITEMS, AND/OR GROUPS: 2
TOTAL NUMBER OF "ITEMS" (INCLUDING GROUPS): 24
NUMBER OF CHARACTERS IN ID FIELDS: 4
MAXIMUM NUMBER OF RESPONSE-CODES FOR ANY ITEM: 4
THE MISSING VALUE CODE FOR CONTINUOUS DATA: 9.0000
THE DATA WILL BE STORED IN MEMORY

THE ITEMS WILL BE CALIBRATED--
BY MARGINAL MAXIMUM LIKELIHOOD ESTIMATION
MAXIMUM NUMBER OF EM CYCLES PERMITTED: 100
NUMBER OF PARAMETER-SEGMENTS USED IS: 32
NUMBER OF FREE PARAMETERS IS: 41
MAXIMUM NUMBER OF M-STEP ITERATIONS IS 4 TIMES
THE NUMBER OF PARAMETERS IN THE SEGMENT
THE M-STEP CONVERGENCE CRITERION IS: 0.000100

THE EM-CYCLE CONVERGENCE CRITERION IS: 0.001000
 THE RK CONTROL PARAMETER (FOR THE M-STEPS) IS: 0.9000
 THE RM CONTROL PARAMETER (FOR THE M-STEPS) IS: 1.0000
 THE MAXIMUM ACCELERATION PERMITTED IS: 0.0000
 THETA-GROUP LOCATIONS WILL REMAIN UNCHANGED

OUTPUT-CONTROL PARAMETERS

THE ITEM PARAMETERS WILL BE SAVED ON FILE <jobname>.PAR

QUADRATURE POINTS FOR MML,

AT THETA:

-4.500
 -4.000
 -3.500
 -3.000
 -2.500
 -2.000
 -1.500
 -1.000
 -0.500
 0.000
 0.500
 1.000
 1.500
 2.000
 2.500
 3.000
 3.500
 4.000
 4.500

MULTILOG for Windows 7.00.2327.2 STEP 4

READING DATA...

KEY-

CODE CATEGORY

1	11111111111111111111
2	22222222222222222222
3	0030003033000300030330
4	0000000004000000000040

FORMAT FOR DATA-

(4A1, 1X, I1 ,1X, 22A1)

FIRST OBSERVATION AS READ-

ID	1	
ITEMS	21311111111	
NORML	0.000	9.000

FINISHED CYCLE 81
 MAXIMUM INTERCYCLE PARAMETER CHANGE= 0.00095 P(3)

ITEM SUMMARY

MULTILOG for Windows 7.00.2327.2 STEP 4

ITEM 6: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 13 1.03 (0.07)

B(1) 12 -0.92 (0.07)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6	0.100	0.117	0.135	0.156	0.176	0.197	0.217	0.235
-1.4 - 0.0	0.249	0.259	0.264	0.264	0.258	0.247	0.232	0.214
0.2 - 1.6	0.193	0.173	0.152	0.132	0.113	0.097	0.082	0.069
1.8 - 3.0	0.057	0.048	0.040	0.033	0.027	0.022	0.018	

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 249 677

OBS. PROP. 0.2689 0.7311

GROUP 1:

EXP. PROP. 0.3145 0.6855

GROUP 2:

EXP. PROP. 0.3145 0.6855

ITEM 7: 3 NOMINAL CATEGORIES, 3 HIGH

CATEGORY(K): 1 2 3

A(K) -0.75 0.00 0.75

C(K) 0.00 0.32 0.54

CONTRAST-COEFFICIENTS (STANDARD ERRORS)

FOR: A C

CONTRAST P(#) COEFF.[POLY.] P(#) COEFF.[TRI.]

1 5 0.75 (0.02) 14 -0.32 (0.08)

2 45 0.00 (0.00) 15 -0.23 (0.07)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6	0.093	0.109	0.127	0.148	0.171	0.197	0.224	0.253
-1.4 - 0.0	0.282	0.310	0.334	0.354	0.368	0.373	0.371	0.360
0.2 - 1.6	0.343	0.320	0.293	0.265	0.236	0.208	0.181	0.157
1.8 - 3.0	0.135	0.116	0.099	0.084	0.072	0.061	0.052	

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3

OBS. FREQ. 227 293 406

OBS. PROP. 0.2451 0.3164 0.4384

GROUP 1:

EXP. PROP. 0.2816 0.2920 0.4264

GROUP 2:

EXP. PROP. 0.2816 0.2920 0.4264

ITEM 17: 2 GRADED CATEGORIES

P(#) ESTIMATE (S.E.)

A 13 1.03 (0.07)

B(1) 12 -0.92 (0.07)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6	0.100	0.117	0.135	0.156	0.176	0.197	0.217	0.235
-1.4 - 0.0	0.249	0.259	0.264	0.264	0.258	0.247	0.232	0.214
0.2 - 1.6	0.193	0.173	0.152	0.132	0.113	0.097	0.082	0.069
1.8 - 3.0	0.057	0.048	0.040	0.033	0.027	0.022	0.018	

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2
 OBS. FREQ. 445 838
 OBS. PROP. 0.3468 0.6532
 GROUP 1:
 EXP. PROP. 0.3145 0.6855
 GROUP 2:
 EXP. PROP. 0.3145 0.6855

ITEM 18: 3 NOMINAL CATEGORIES, 3 HIGH

CATEGORY(K): 1 2 3
 A(K) -0.75 0.00 0.75
 C(K) 0.00 0.32 0.54

CONTRAST-COEFFICIENTS (STANDARD ERRORS)

FOR: A C
 CONTRAST P(#) COEFF.[POLY.] P(#) COEFF.[TRI.]
 1 5 0.75 (0.02) 14 -0.32 (0.08)
 2 59 0.00 (0.00) 15 -0.23 (0.07)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)
 -3.0 - -1.6 0.093 0.109 0.127 0.148 0.171 0.197 0.224 0.253
 -1.4 - 0.0 0.282 0.310 0.334 0.354 0.368 0.373 0.371 0.360
 0.2 - 1.6 0.343 0.320 0.293 0.265 0.236 0.208 0.181 0.157
 1.8 - 3.0 0.135 0.116 0.099 0.084 0.072 0.061 0.052

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2 3
 OBS. FREQ. 394 351 538
 OBS. PROP. 0.3071 0.2736 0.4193
 GROUP 1:
 EXP. PROP. 0.2816 0.2920 0.4264
 GROUP 2:
 EXP. PROP. 0.2816 0.2920 0.4264

ITEM 23: GRP1, N[MU: 0.00 SIGMA: 1.00]

P(#);(S.E.): 71; (0.00) 72; (0.00)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)
 -3.0 - -1.6 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000
 -1.4 - 0.0 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000
 0.2 - 1.6 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000
 1.8 - 3.0 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

ITEM 24: GRP2, N[MU: 0.00 SIGMA: 1.00]

P(#);(S.E.): 74; (0.00) 75; (0.00)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)
 -3.0 - -1.6 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000
 -1.4 - 0.0 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000
 0.2 - 1.6 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000
 1.8 - 3.0 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

TOTAL TEST INFORMATION

FOR GROUP 1:

@THETA:	INFORMATION:
-3.0 - -1.6	3.726 4.173 4.649 5.143 5.642 6.132 6.599 7.024
-1.4 - 0.0	7.382 7.648 7.804 7.840 7.767 7.605 7.382 7.117
0.2 - 1.6	6.823 6.500 6.146 5.757 5.338 4.896 4.448 4.012
1.8 - 3.0	3.603 3.232 2.903 2.617 2.372 2.164 1.988

@THETA:	POSTERIOR STANDARD DEVIATION:
-3.0 - -1.6	0.518 0.490 0.464 0.441 0.421 0.404 0.389 0.377
-1.4 - 0.0	0.368 0.362 0.358 0.357 0.359 0.363 0.368 0.375
0.2 - 1.6	0.383 0.392 0.403 0.417 0.433 0.452 0.474 0.499
1.8 - 3.0	0.527 0.556 0.587 0.618 0.649 0.680 0.709

MARGINAL RELIABILITY: 0.8391

TOTAL TEST INFORMATION

FOR GROUP 2:

@THETA:	INFORMATION:
-3.0 - -1.6	3.726 4.173 4.649 5.143 5.642 6.132 6.599 7.024
-1.4 - 0.0	7.382 7.648 7.804 7.840 7.767 7.605 7.382 7.117
0.2 - 1.6	6.823 6.500 6.146 5.757 5.338 4.896 4.448 4.012
1.8 - 3.0	3.603 3.232 2.903 2.617 2.372 2.164 1.988

@THETA:	POSTERIOR STANDARD DEVIATION:
-3.0 - -1.6	0.518 0.490 0.464 0.441 0.421 0.404 0.389 0.377
-1.4 - 0.0	0.368 0.362 0.358 0.357 0.359 0.363 0.368 0.375
0.2 - 1.6	0.383 0.392 0.403 0.417 0.433 0.452 0.474 0.499
1.8 - 3.0	0.527 0.556 0.587 0.618 0.649 0.680 0.709

MARGINAL RELIABILITY: 0.8391

GROUP 1

NEGATIVE TWICE THE LOGLIKELIHOOD= 682.4
 (CHI-SQUARE FOR SEVERAL TIMES MORE EXAMINEES THAN CELLS)

GROUP 2

NEGATIVE TWICE THE LOGLIKELIHOOD= 1239.1
 (CHI-SQUARE FOR SEVERAL TIMES MORE EXAMINEES THAN CELLS)

TOTAL, NEGATIVE TWICE THE LOGLIKELIHOOD, ALL GROUPS= 1921.5

NORMAL PROGRAM TERMINATION

START DATE: 01-30-2006
 START TIME: 01:18:48
 END TIME: 01:18:51