



Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü  
Eđitim Bilimleri Anabilim Dalı  
Eđitimde Ölçme ve Deđerlendirme Bilim Dalı

**ULUSLARARASI ÖĐRENCİ DEĐERLENDİRME PROGRAMI  
(PISA) 2006 ÖĐRENCİ ANKETİNİN KÜLTÜRLER ARASI  
EŐDEĐERLİĐİNİN İNCELENMESİ**

Mustafa ASİL

Doktora Tezi

Ankara, 2010



**ULUSLARARASI ÖĐRENCİ DEĐERLENDİRME PROGRAMI (PISA) 2006  
ÖĐRENCİ ANKETİNİN KÜLTÜRLER ARASI EŐDEĐERLİĐİNİN  
İNCELENMESİ**

Mustafa ASİL

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü  
Eđitim Bilimleri Anabilim Dalı  
Eđitimde Ölçme ve Deđerlendirme Bilim Dalı

Doktora Tezi

Ankara, 2010

## KABUL VE ONAY

Mustafa ASİL tarafından hazırlanan “Uluslararası Öğrenci Değerlendirme Programı (PISA) 2006 Öğrenci Anketinin Kültürler Arası Eşdeğerliğinin İncelenmesi” başlıklı bu çalışma, 25.01.2010 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından doktora tezi olarak kabul edilmiştir.



Prof. Dr. Petek AŞKAR (Başkan)



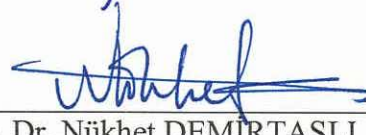
Prof. Dr. Giray BERBEROĞLU



Doç. Dr. Selahattin GELBAL (Danışman)



Doç. Dr. Hülya KELECİOĞLU



Doç. Dr. Nükhet DEMİRTAŞLI

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.

25.01.2010



Mustafa ASİL

Gölce Bahar ve Annesi'ne

## TEŐEKKÜR

Çalıőmalarımnda desteęini esirgemeyen danıőmanım Doç. Dr. Selahattin Gelbal'a, deęerli katkıları için Prof. Dr. Giray Berberoęlu'na ve bu süreçte beni yalnız bırakmayan Zafer Çepni, Dilara Bakan Kalaycıoęlu ve Ümran Çekinmez'e teşekkürlerimi sunarım. Çalıőmalarımın sonuçlandırılmasında görüşlerine başvurduğum uzmanlar ile jüri üyelerine ayrıca teşekkür ederim.

Her zaman yanımda olduğunu bildiğim eşime desteęinden ve fedakarlığından dolayı müteőekkirim.

## ÖZET

ASİL, Mustafa. Uluslararası Öğrenci Değerlendirme Programı (PISA) 2006 Öğrenci Anketinin Kültürler Arası Eşdeğerliğinin İncelenmesi, Doktora Tezi, Ankara, 2010.

Bu çalışmada, Uluslararası Öğrenci Değerlendirme Programı (PISA) 2006 kapsamında uygulanan öğrenci anketinin kültürler ve diller arası eşdeğerliği, Avustralya, Yeni Zelanda, Amerika Birleşik Devletleri ve Türkiye örneklemi üzerinde karşılaştırmalı olarak incelenmiştir.

Araştırmanın birinci aşamasında, öğrenci anketinin faktör yapısı doğrulayıcı faktör analizi (DFA) yöntemiyle incelenmiştir. İkinci aşamada ise, ortalama ve kovaryans modeline (OKM) (Sörbom, 1974) dayanan çoklu grup doğrulayıcı faktör analizi (ÇG-DFA) kullanılarak anket maddelerinin farklı kültürler ve diller arasında değişen madde fonksiyonu (DMF) gösterip göstermediği araştırılmıştır. Son aşamada, uzman görüşlerine başvurularak, maddelerin DMF göstermesinin nedenleri belirlenmeye çalışılmıştır.

DFA sonuçları ölçme modelinin tüm kültürlerde aynı faktör yapısına sahip olduğunu göstermiştir. ÇG-DFA bulguları ülkeler arasında farklı fonksiyon gösteren maddelerin olduğunu ortaya koymuştur. Ülkeler arasında dilsel ve kültürel farklılıklar arttıkça DMF gösteren maddelerin sayısının da arttığı gözlenmiştir. Maddelerin DMF göstermesinin nedenlerinin çeviri problemleri ve kültürel farklılıklar olduğu sonucuna varılmıştır.

### **Anahtar Sözcükler**

PISA, Değişen Madde Fonksiyonu (DMF), Ölçme Eşdeğerliği, Çoklu Grup Doğrulayıcı Faktör Analizi (ÇG-DFA), Ortalama ve Kovaryans Modeli (OKM).

## **ABSTRACT**

ASİL, Mustafa. Assessing Cross Cultural Equivalence of The Programme For International Student Assessment (PISA) 2006 Student Questionnaire, Ph.D. Dissertation, Ankara, 2010.

In this study, cross lingual and cross cultural equivalence of the Programme For International Student Assessment (PISA) 2006 Student Questionnaire were assessed comparatively across Turkey, the United States of America, Australia, and New Zealand samples.

At the first stage of this study, factor structure of the questionnaire constructs were examined using confirmatory factor analysis (CFA). At the second stage, multi-group confirmatory factor analyses (MG-CFA) based on mean and covariance structure (MACS) (Sörbom, 1974) were conducted in order to assess whether questionnaire items exhibit differential item functioning (DIF) across different cultures or languages. At the last stage, causes of DIF were identified by expert judgments.

CFA results showed that the measurement model had the same factor structure across cultures. MG-CFA results indicated that some items exhibited DIF between countries. It is observed that as the linguistic and cultural differences increased between countries, the number of DIF items increased. The main reasons for DIF were found to be due to poor translation and differences in cultural relevance.

### **Key Words**

PISA, Differential Item Functioning (DIF), Measurement Equivalence, Multi-Group Confirmatory Factor Analysis (MG-CFA), Mean and Covariance Structure (MACS).

## İÇİNDEKİLER

	Sayfa
<b>KABUL VE ONAY</b> .....	<b>i</b>
<b>BİLDİRİM</b> .....	<b>ii</b>
<b>ADAMA</b> .....	<b>iii</b>
<b>TEŞEKKÜR</b> .....	<b>iv</b>
<b>ÖZET</b> .....	<b>v</b>
<b>ABSTRACT</b> .....	<b>vi</b>
<b>İÇİNDEKİLER</b> .....	<b>vii</b>
<b>KISALTMALAR</b> .....	<b>x</b>
<b>ÇİZELGELER</b> .....	<b>xi</b>
<b>ŞEKİLLER</b> .....	<b>xii</b>
<b>1. GİRİŞ</b> .....	<b>1</b>
1.1 PROBLEM DURUMU .....	1
1.2 DOĞRULAYICI FAKTÖR ANALİZİ.....	6
1.3 ÇOKLU GRUP DOĞRULAYICI FAKTÖR ANALİZİYLE DMF ANALİZİ....	13
1.4 REFERANS DEĞİŞKENİN BELİRLENMESİ .....	20
1.5 DEĞİŞEN MADDE FONKSİYONUNUN OLASI SEBEPLERİ.....	22
1.6 PISA VE PISA ÖĞRENCİ ANKETİ .....	27
1.7 PROBLEM CÜMLESİ .....	32
1.8 ARAŞTIRMANIN AMACI.....	32
1.9 ARAŞTIRMANIN ÖNEMİ .....	33
1.10 İLGİLİ ARAŞTIRMALAR .....	33
<b>2. YÖNTEM</b> .....	<b>37</b>
2.1 ARAŞTIRMANIN TÜRÜ .....	37
2.2 ÇALIŞMA GRUBU .....	37
2.3 VERİLERİN ELDE EDİLMESİ.....	38
2.4 VERİLERİN ÇÖZÜMLENMESİ.....	38
<b>3. BULGULAR VE YORUMLAR</b> .....	<b>48</b>
3.1 ALT PROBLEM 1'E İLİŞKİN BULGULAR VE YORUMLAR.....	48

3.2 ALT PROBLEM 2A'YA İLİŞKİN BULGULAR VE YORUMLAR.....	55
3.3 ALT PROBLEM 2B'YE İLİŞKİN BULGULAR VE YORUMLAR .....	56
3.4 ALT PROBLEM 2C'YE İLİŞKİN BULGULAR VE YORUMLAR .....	57
3.5 ALT PROBLEM 3'E İLİŞKİN BULGULAR VE YORUMLAR.....	58
<b>4. SONUÇ VE ÖNERİLER.....</b>	<b>64</b>
4.1 SONUÇLAR .....	64
4.1.1 Alt Problem 1'e İlişkin Sonuçlar.....	64
4.1.2 Alt Problem 2'ye İlişkin Sonuçlar.....	65
4.1.3 Alt Problem 3'e İlişkin Sonuçlar.....	66
4.2 SINIRLILIKLAR.....	67
4.3 ÖNERİLER.....	67
<b>KAYNAKÇA .....</b>	<b>69</b>
Ek 1 Avustralya, Yeni Zelanda ve ABD Öğrenci Anketlerinde Bilimsel Sorgulamaya Verilen Destek Tutumunu Ölçmek için Kullanılan Maddeler .....	78
Ek 2 Türkiye Öğrenci Anketinde Bilimsel Sorgulamaya Verilen Destek Tutumunu Ölçmek için Kullanılan Maddeler .....	79
Ek 3 Kayıp Veri Oranları .....	80
Ek 4 Avustralya Örnekleme için Aykırı Değer Analizi Grafikleri.....	81
Ek 5 Yeni Zelanda Örnekleme için Aykırı Değer Analizi Grafikleri.....	82
Ek 6 ABD Örnekleme için Aykırı Değer Analizi Grafikleri .....	83
Ek 7 Türkiye Örnekleme için Aykırı Değer Analizi Grafikleri.....	84
Ek 8 Avustralya Örnekleme için Çoklu Bağlantılılık Analizi Sonuçları.....	85
Ek 9 Yeni Zelanda Örnekleme için Çoklu Bağlantılılık Analizi Sonuçları.....	86
Ek 10 ABD Örnekleme için Çoklu Bağlantılılık Analizi Sonuçları.....	87
Ek 11 Türkiye Örnekleme için Çoklu Bağlantılılık Analizi Sonuçları.....	88
Ek 12 Birleştirilmiş Veriden Ortak Eşik Değerlerin Tespit Edilmesinde Kullanılan PRELIS Komut Dosyası .....	89
Ek 13 Eşik Değerlerin Sabitletmesinde Kullanılan PRELIS Komut Dosyası .....	90
Ek 14 Doğrulayıcı Faktör Analizi LISREL-SIMPLIS Komut Dosyası.....	91
Ek 15 Avustralya Örnekleme için Standartlaştırılmış Faktör Yükleri ve Hata Varyansları .....	92

Ek 16 Yeni Zelanda Örnekleme için Standartlaştırılmış Faktör Yükleri ve Hata Varyansları .....	93
Ek 17 ABD Örnekleme için Standartlaştırılmış Faktör Yükleri ve Hata Varyansları .....	94
Ek 18 Türkiye Örnekleme için Standartlaştırılmış Faktör Yükleri ve Hata Varyansları .....	95
Ek 19 Referans Değişkeni Belirlemede Kullanılan Sınırlandırılmış Temel Model LISREL-SIMPLIS Komut dosyası .....	96
Ek 20 Avustralya ve Yeni Zelanda Karşılaştırması için Referans Değişkeni Belirleme Sonuçları.....	98
Ek 21 Avustralya ve ABD Karşılaştırması için Referans Değişkeni Belirleme Sonuçları .....	99
Ek 22 Avustralya ve Türkiye Karşılaştırması için Referans Değişkeni Belirleme Sonuçları .....	100
Ek 23 DMF Gösteren Maddeleri Belirlemede Kullanılan Serbest Temel Model LISREL-SIMPLIS Komut dosyası .....	101
Ek 24 Madde İnceleme Formu .....	103
ÖZGEÇMİŞ .....	104

## KISALTMALAR

<b>PISA</b>	: Uluslararası Öğrenci Değerlendirme Programı
<b>TIMSS</b>	: Üçüncü Uluslararası Fen Bilgisi ve Matematik Çalışması
<b>PIRLS</b>	: Uluslararası Okuma Becerilerinde Gelişim Projesi
<b>OECD</b>	: Ekonomik İşbirliği ve Kalkınma Örgütü
<b>YEM</b>	: Yapısal Eşitlik Modellemesi
<b>DFA</b>	: Doğrulayıcı Faktör Analizi
<b>ÇG-DFA</b>	: Çoklu Grup Doğrulayıcı Faktör Analizi
<b>OKM</b>	: Ortalama ve Kovaryans Modeli
<b>MTK</b>	: Madde Tepki Kuramı
<b>DMF</b>	: Değişen Madde Fonksiyonu
<b>AUS</b>	: Avustralya
<b>ABD</b>	: Amerika Birleşik Devletleri
<b>NZL</b>	: Yeni Zelanda
<b>TUR</b>	: Türkiye

## ÇİZELGELER

<b>Çizelge 1.1</b> Ölçme Modeli Parametreleri.....	8
<b>Çizelge 1.2</b> PISA 2006’da Ölçülen Öğrenci Tutumları.....	30
<b>Çizelge 2.1</b> Çalışma Grubu Frekans ve Yüzdeleri.....	38
<b>Çizelge 3.1</b> Fen Bilimlerinin Genel Değeri Faktörünü Oluşturan Maddelere Ait Betimsel İstatistikler .....	49
<b>Çizelge 3.2</b> Fen Bilimlerinin Kişisel Değeri Faktörünü Oluşturan Maddelere Ait Betimsel İstatistikler .....	50
<b>Çizelge 3.3</b> Bilimsel Sorgulamaya Verilen Destek Ölçme Modeli Uyum İstatistikleri .	51
<b>Çizelge 3.4</b> Fen Bilimlerinin Genel Değeri ve Fen Bilimlerinin Kişisel Değeri Faktörlerinin Güvenirlikleri .....	52
<b>Çizelge 3.5</b> Fen Bilimlerinin Genel Değeri Faktörünü Oluşturan Maddelerin Güvenirlikleri .....	53
<b>Çizelge 3.6</b> Fen Bilimlerinin Kişisel Değeri Faktörünü Oluşturan Maddelerin Güvenirlikleri .....	54
<b>Çizelge 3.7</b> Bilimsel Sorgulamaya Verilen Destek Ölçme Modeli Maddeleri için Avustralya ve Yeni Zelanda Örneklemelerinde Yapılan DMF Analizi Sonuçları ...	55
<b>Çizelge 3.8</b> Bilimsel Sorgulamaya Verilen Destek Ölçme Modeli Maddeleri için Avustralya ve ABD Örneklemelerinde Yapılan DMF Analizi Sonuçları .....	56
<b>Çizelge 3.9</b> Bilimsel Sorgulamaya Verilen Destek Ölçme Modeli Maddeleri için Avustralya ve Türkiye Örneklemelerinde Yapılan DMF Analizi Sonuçları .....	57
<b>Çizelge 3.10</b> Maddelerin DMF Göstermesinin Olası Nedenlerine İlişkin Uzman Görüşleri.....	63

## ŞEKİLLER

<b>Şekil 1.1</b> İki Faktörlü Ölçme Modeli .....	7
<b>Şekil 1.2</b> Ölçme Değişmezliğinin Sağlandığı İdeal Durum.....	15
<b>Şekil 1.3</b> Faktör Yükleri Eşitsizliğinin Ölçme Değişmezliği Üzerine Etkisi.....	16
<b>Şekil 1.4</b> Regresyon Sabitleri Eşitsizliğinin Ölçme Değişmezliği Üzerine Etkisi.....	17
<b>Şekil 2.1</b> Bilimsel Sorgulamaya Verilen Destek Ölçme Modeli .....	41

## 1. GİRİŞ

Bu bölümde problem durumu, doğrulayıcı faktör analizi, çoklu grup doğrulayıcı faktör analiziyle değişen madde fonksiyonu analizi, referans değişkenin belirlenmesi, değişen madde fonksiyonunun olası sebepleri, PISA ve PISA öğrenci anketi, problem cümlesi, araştırmanın amacı, önemi ve ilgili araştırmalar yer almaktadır.

### 1.1 PROBLEM DURUMU

Eğitim, sağlık, ekonomi, sanat ve benzeri alanlardaki artan uluslararası işbirliği, göç, küreselleşme gibi birçok nedenlerden dolayı psikometri uzmanları, eğitimciler ve diğer alanlardaki araştırmacılar uzun zamandır farklı dillerdeki bireyleri değerlendirme durumuyla karşı karşıya kalmaktadırlar. Böyle durumlarda tek bir ölçme ve değerlendirme aracının kullanılması neredeyse imkansızdır. Bu yüzden, ölçme araçları genelde farklı dillere ve kültürlere uyarlanmaktadır (de Klerk, 2008; Sireci, 2005). Günümüzde başarı, yetenek, kişilik testleri ve anketlerin diğer dillere uyarlanmasına olan ihtiyaç da giderek artmaktadır (Hambleton, 2005; Hambleton ve Patsula, 1998; de Klerk, 2008).

Testlerin bir dil veya kültürden diğerlerine uyarlanmasının birçok sebebi bulunmaktadır. Okul başarısının kültür ve dil grupları arası karşılaştırılması, yeni ölçme araçları geliştirmede harcanan zaman ve paradan tasarruf etme ve değerlendirmelerde tarafsızlığın sağlanması, bireyleri kendi dillerinde değerlendirme isteği bu sebeplerden bazılarıdır. Uluslararası başarı karşılaştırmaları eğitimde politika geliştiricilere ülkelerindeki eğitimin kalitesi hakkında veriler, elde edilen bulgular için olası açıklamalar ve başarının nasıl artırılacağına yönelik tavsiyeler sunabilir (Hambleton, 2002, 2005; Hambleton ve Patsula, 1998, 1999; Sireci ve Berberoğlu, 2000; Yayan ve Berberoğlu, 2004; Zumbo, 2003).

Farklı diller ve kültürler arası karşılaştırmalarda uyarlama teriminin kullanımı çeviri teriminin yerine tercih edilmektedir. Test uyarlama süreci daha esnek, farklı diller arasında istenilen anlamın verilebilmesi için daha karmaşık kelimelerin değiştirilmesine

imkan sunar (Sireci, Patsula ve Hambleton, 2005). Uygun bir uyarlama psikolojik, dilsel ve kültürel özellikleri dengeli bir şekilde ele almayı gerektirir (Van de Vijver ve Hambleton, 1996). Test uyarlama, bir kültür ve dilde geliştirilmiş bir aracın ikinci bir kültür ve dilde kullanılması ve geçerliğinin sağlanması çalışmalarının tüm aktivitelerini yansıtmaktadır. Çeviri yapılması ise bu aktivitelerden sadece bir tanesidir (Hambleton, 2002, 2005).

Milletler ve kültürler arası karşılaştırma çalışmalarının bazı yöntem problemlerinden dolayı iyi bir şekilde uygulanması oldukça zordur. Bu çalışmalardaki problem alanları (a) ölçülecek özelliklerin, yapıların tanımlanması ve değişkenlerin belirlenmesi, (b) ulusları temsil edecek örneklemelerin seçilmesi, (c) test uygulama durumlarının standartlaştırılması ve (d) ölçme araçlarının uyarlanmasıdır (Hambleton, 2002).

Testler veya anketler bir dilden başka bir dile uyarlandığında farklı dillerdeki versiyonların eşdeğerliğinin sağlandığından emin olunmalıdır. Uyarlanan veya çevirisi yapılan her bir form farklı bir yapıyı ölçebilir, zorluk ya da aşinalık dereceleri bakımından farklılık gösterebilir (Hambleton, 1994; Sireci ve Bastari, 1998; Sireci ve Berberoğlu, 2000). Testlerin uyarlanmasında amaç, farklılıkları anlamsal, psikometrik, dilsel ve psikolojik açılarından kabul edilebilir seviyelere indirmektir (Hambleton ve de Jong, 2003). Kullanılan her bir madde bir kültürdeki bireyler için ne anlam ifade ediyorsa diğer kültürlerdeki bireyler için de aynı şeyi ifade etmelidir (Hui ve Triandis, 1985).

Farklı kültür veya dilden bireyleri karşılaştırmak amacıyla yapılan her bir çalışmada, karşılaştırmanın anlamlı olabilmesi için testlerin ölçtüğü yapıların eşdeğer olması gerekir. Bu tür çalışmalarda ölçme eşdeğerliğinin sağlanması temel bir sayılıdır (Gierl, 2000). Bu nedenle Eğitimde ve Psikolojide Ölçme Standartları (AERA, APA ve NCME, 1999) ve Test Uyarlama Yönergeleri (Hambleton, 1994; ITC, 2001) kültürler arası çalışmalar yapan araştırmacılardan farklı dillerdeki ölçme araçlarından elde edilen puanların karşılaştırılabilirliği için kanıtlar sunmalarını istemektedir. Test Uyarlama Yönergelerindeki bu duruma ilişkin bazı maddeler şunlardır (ITC, 2001):

- Çalışmanın esas amacı için önemli olmayan kültürel farklılıkların etkileri mümkün olduğu kadar minimize edilmelidir (Yönerge 1).
- Test geliştiriciler, a) testlerin farklı dil versiyonlarının eşdeğerliğini belirlemek ve (b) testlerin bazı gruplar için olan olumsuz yönlerini tespit etmek için uygun istatistiksel teknikleri uygulamalıdır (Yönerge 9).
- Test geliştiriciler maddelerin bütün gruplar arasında eşdeğer olduğuna dair istatistiksel kanıtlar sunmalıdır (Yönerge 11).
- Farklı gruplar üzerinde karşılaştırma, ancak ölçme değişmezliğinin sağlandığı ölçekler için yapılabilir (Yönerge 21).

Eşdeğerlik (equivalence) terimi kültürler arası ve milletler arası karşılaştırmalarda sıklıkla kullanılan bir terimdir. Buna rağmen kültürler arası çalışmalar yapan araştırmacılar eşdeğerliğe alternatif anlamlar yüklemiş ve farklı şekillerde sınıflamışlardır. Kültürler arası araştırmalar literatürü incelendiğinde elliden fazla eşdeğerlik teriminin kullanıldığı görülmektedir (Johnson, 1998). Bu çalışmada “ölçme eşdeğerliği” (measurement equivalence) ve “ölçme değişmezliği” (measurement invariance) terimlerinin kullanılması tercih edilmiştir.

Gruplar arası karşılaştırmaların yapılabilmesi için ölçme eşdeğerliğinin ya da değişmezliğinin gösterilmesi gereklidir (Reise, Widaman ve Pugh, 1993; Van de Vijver ve Tanzer, 2004; Vandenberg, 2002; Vandenberg ve Lance, 2000). Ölçmelerin eşdeğer olması için, gözlenen değişkenlerle örtük değişkenler arasındaki ilişkilerin gruplar üzerinde aynı olması gerekmektedir (Drasgow ve Kanfer, 1985). Diğer bir ifadeyle, ölçülen özellik üzerinde eşit yetenekte olan farklı gruplara ait bireyler maddeyi farklı yanıtlıyorlarsa o madde eşdeğer değildir (farklı fonksiyon gösteriyordur) denilir (Chan, 2000).

Ölçme eşdeğerliğinin sağlanamaması madde tepki kuramı literatüründe değişen (farklı) fonksiyon gösterme olarak adlandırılır. Ölçme değişmezliği madde bazında

sağlanamıyorsa ‘Değişen Madde Fonksiyonu’ (DMF), test bazında sağlanamıyorsa ‘Değişen Test Fonksiyonu’ (DTF) olarak isimlendirilir. Bir testin DMF gösterip DTF göstermemesi mümkündür; çünkü maddeler farklı yönlerde eşit büyüklükte DMF gösterebilir (Stark, Chernyshenko ve Drasgow, 2006). Farklı gruplara ait test puanlarının karşılaştırılabilirliği, yani eşdeğerliği genellikle DMF analizleriyle değerlendirilmektedir (Ercikan ve diğerleri, 2004).

Ölçme araçlarının kültürler arası eşdeğerliği iki yöntemle incelenmektedir: uzman görüşleri ve psikometrik analizler. Psikometrik yöntemlerden en çok kullanılanları, madde tepki kuramına (MTK) (Lord, 1980) ve doğrulayıcı faktör analizine (DFA) (Jöreskog, 1971) dayalı yöntemlerdir. DMF belirleme yöntemlerinden hiçbiri maddedeki hangi unsurun DMF’ye sebep olduğuna dair bilgi vermemektir. Dolayısıyla, DMF’nin kaynağına yönelik araştırmalarda uzman görüşlerine başvurulmaktadır.

Gruplar arası ölçme eşdeğerliğinin sağlanmasında ve yapısal eşdeğerliğin incelenmesinde çoklu grup doğrulayıcı faktör analizi (ÇG-DFA) uygun ve genellikle önerilen bir metot olup çok boyutlu ölçeklere uygulanabilmesi, çok kategorili verilerin analizinde kullanılabilmesi ve hem tek biçimli (uniform) hem de tek biçimli olmayan (non-uniform) DMF’yi belirleyebilmesi gibi birçok yönden diğer metotlara tercih edilmektedir. Bu yöntemle ortak ve kültüre özgü maddeler tespit edilebilmektedir. (Byrne ve Stewart, 2006; Cheung ve Rensvold, 1999; Drasgow ve Kanfer, 1985; Meade ve Lautenschlager, 2004; Raju, Laffitte ve Byrne, 2002; Van de Vijver, 1998; Zumbo, Sireci ve Hambleton, 2003). DMF analizlerinde ÇG-DFA’nın MTK ile aynı etkililiğe sahip olduğu hatta verinin çok kategorili olduğu durumlarda daha iyi performans gösterdiği simülasyon çalışmalarıyla kanıtlanmıştır (Stark, Chernyshenko ve Drasgow, 2006).

Test puanlarının karşılaştırılabilirliğini değerlendirmede kullanılan psikometrik teknikler ölçek seviyesinde analizler ve madde seviyesinde analizler olmak üzere iki sınıfa ayrılabilir. Ölçek seviyesindeki analizlerde genellikle faktör yapısının uyarlanmış versiyonlarda da aynı kalıp kalmadığı incelenmektedir. Eğer faktör yapısı tüm versiyonlarda aynı ise test uyarlamasının başarılı olduğu kabul edilmektedir. Hambleton

ve de Jong (2003) makalelerinde bu yaklaşımı eleştirmişlerdir. Yazarlar bu konuda, her iki versiyondaki faktör yapıları aynı bulunsa bile testlerde madde seviyesinde yanlılığın hala olabileceğine dair kanıtlar sunan Zumbo'nun makalesine gönderme yapmaktadırlar.

Çalışmasında Zumbo (2003, 2005) ölçek seviyesinde yapılan incelemelerde yanlı çalışan maddelerin belirlenemeyebileceğini göstermiştir ve testlerin çevirisi ve uyarlanması çalışmalarında ölçme eşdeğerliğini sağlamak için madde bazında (DMF) analizlerin yapılması gerektiği sonucuna varmıştır. Ellis ve Raju (2003) da bazı maddelerin bir grup aleyhinde diğer maddelerin de aynı grup lehinde fonksiyon gösterdiği durumlarda yanlı maddelerin etkilerinin kaybolabileceğini ve madde bazında yanlılıklar olduğu halde test bazındaki yanlılığın belirlenemeyebileceğini iddia etmişlerdir. Hambleton ve Patsula (1999), eğer kültürler arası karşılaştırma yapılacaksa, testin farklı dil versiyonlarının eşdeğerliğinden emin olabilmek için bir veya birden fazla yöntemle DMF analizlerinin yapılmasını önermektedirler.

Ölçme eşdeğerliği sağlandığında, (a) ölçülen yapılar her bir kültüre genellenebilir, (b) hata ve yanlılık kaynakları en azdır, (c) kültürel farklılıklar ölçülen yapıyı gruplar üzerinde farklı olarak etkilememiştir ve (d) kültürler veya gruplar arası farklılıklar karşılaştırılabilir (Little, 1997).

Uluslararası eğitim araştırmalarının en önemli amaçlarından birisi de öğrenci, öğretmen, veli ve okul anketleri kullanarak öğrenci ve öğretmen özellikleri, özgeçmişleri, etkinlikleri ve okuldaki öğrenme ortamları hakkında bilgi toplamaktır. Özellikle öğrenci tutum, davranış ve değer yargıları önemli öğrenme çıktıları olarak görülmektedir. Öğrenci performanslarındaki farklılıklar elde edilen bu bilgiler yardımıyla açıklanmakta ve eğitim sistemlerinin geliştirilmesi için politikalar üretilmektedir. TIMSS, PISA ve PIRLS gibi çalışmalarda bu amaçla öğrenci, öğretmen ya da veli anketi gibi ölçme araçları uygulanmaktadır. Uluslararası karşılaştırmalarda daha çok bilişsel testlerin eşdeğerliğinin gösterilmesine emek harcanmakta fakat anketlerin eşdeğerliğiyle ilgili araştırmalara yeterince önem verilmediği görülmektedir (Schulz, 2003, 2005, 2008).

Bu çalışmada PISA 2006 kapsamında uygulanan öğrenci anketinin maddelerinin Avustralya, Yeni Zelanda, ABD ve Türkiye örneklerinde eşdeğer olup olmadığı ÇG-DFA yöntemiyle incelenmiştir. Maddelerin değişen fonksiyon göstermesinin kültürel farklılıklardan mı yoksa uyarlamadan kaynaklanan problemlerden mi kaynaklandığı irdelenmiştir. Maddelerin kültürler arasında DMF göstermesinin nedenleri uzman görüşlerine başvuru olarak değerlendirilmiştir. Maddelerin DMF gösterip göstermediğinin incelenmesinden önce her bir kültürde aynı yapının ölçüldüğünün doğrulanması gerekmektedir. Ölçülen yapıların benzerliğini göstermek amacıyla bu çalışmada her bir örnekte DFA uygulanmış daha sonra madde bazında ölçme eşdeğerliği analizleri yapılmıştır.

## 1.2 DOĞRULAYICI FAKTÖR ANALİZİ

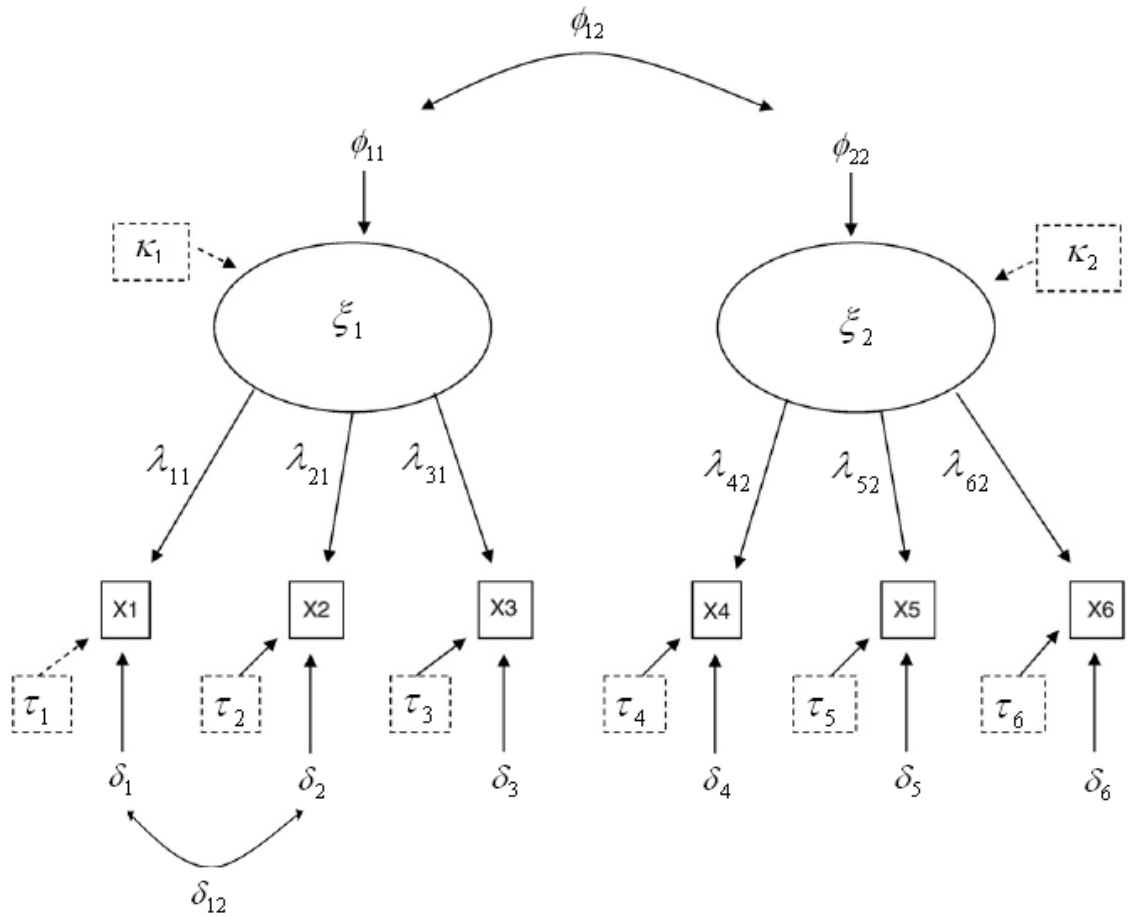
Yapısal eşitlik modellemesi (YEM), gözlenen değişkenlerle örtük değişkenler arası ilişkileri inceleyen ölçme modellerinin ve örtük değişkenler arası ilişkileri inceleyen yapısal modellerin test edilmesinde kullanılan analizlerin birleşimidir (Raykov ve Marcoulides, 2006).

DFA, özellikle ölçme modellerinin (measurement models) incelendiği bir tür yapısal eşitlik modellemesi (YEM) türüdür. Ölçme modellerinde gözlenen değişkenlerle (gösterge) örtük değişkenler (faktör) arasındaki ilişkiler incelenir. DFA'nın en temel özelliği hipoteze ve teoriye dayalı olmasıdır. Analizlerde teorik ilişkilerin ampirik gözlemlerle elde edilmiş veri setlerinde de var olup olmadığı araştırılır. Açımlayıcı faktör analizinin tersine, DFA'da araştırmacı, ölçme modelini önceden belirlemelidir.

DFA, uygulamalı araştırmalarda en çok kullanılan istatistiksel yöntemlerden biri haline gelmiştir. Bunun nedeni ise DFA'nın araştırmacıların en çok yanıt aradıkları sorulara yanıtlar vermesidir. DFA'nın en yaygın kullanım biçimleri şu şekildedir: (a) ölçme araçlarının psikometrik özelliklerinin incelenmesi, (b) metod etkisinin incelenmesi, (c) yapısal geçerlik çalışmaları ve (d) ölçme değişmezliğinin incelenmesidir (Brown, 2006).

YEM terminolojisinde gözlenemeyen ve ancak bazı göstergeler (indicator) aracılığıyla ölçülebildikleri varsayılan teorik yapılara örtük değişken (latent variable) veya faktör

adı verilir (Şimşek, 2007). Örtük değişkenler DFA modellerinde oval veya dairelerle gösterilir. Gözlenen değişkenler (observed variables) ise örtük değişkenleri ölçmede kullanılan göstergelerdir ve genellikle bunlar ölçme araçlarındaki maddelerdir. Gözlenen değişkenler DFA modellerinde dikdörtgenlerle veya karelerle gösterilir. İki çeşit örtük değişken vardır: dışsal ve içsel. Dışsal değişkenler modelde başka değişkenler tarafından yordanan değişkenlerdir. İçsel değişkenler ise modelde başka değişkenler tarafından yordanan değişkenlerdir. Kompleks modellerde bazı değişkenler hem içsel hem de dışsal olabilmektedir (Harrington, 2009).



Şekil 1.1 İki Faktörlü Ölçme Modeli

Şekil 1.1'de Lisrel notasyonu ile ifade edilmiş iki faktörlü (örtük değişken) bir ölçme modeli verilmiştir. Modeldeki parametreler ve tanımları Çizelge 1.1'de verilmiştir (Brown, 2006, s. 55).

Çizelge 1.1 Ölçme Modeli Parametreleri

İsim	Parametre	Matrix	Tanım
Lambda-X	$\lambda_x$	$\Lambda_x$	Faktör yükleri
Theta delta	$\delta$	$\Theta_\delta$	Hata varyansları ve kovaryansları
Phi	$\phi$	$\Phi$	Faktör varyansları ve kovaryansları
Tau-X	$\tau_x$		Regresyon sabitleri veya madde ortalamaları
Kappa	$\kappa$		Örtük değişkenlerin ortalamaları
Xi (Ksi)	$\xi$		Dışsal değişkenlerin isimleri

Model parametreleri, DFA’da tahmin ve test edilen popülasyon özellikleridir. Gözlenen değişkenlerle örtük değişkenler arasındaki ilişkiler DFA modellerinde tek yönlü oklarla gösterilir ve okun yönü örtük değişkenden gözlenen değişkene doğrudur. Örtük değişkenlerin gözlenen değişkenleri yordadığı biçiminde bir hipotez kurulur. Faktör yükleri, gözlenen değişkenlerin örtük değişkenlerden kestirilmesindeki regresyon katsayılarıdır. Bu değerlerin oldukça yüksek olması istenilir. Genel olarak 0,71 üzeri faktör yükleri mükemmel, 0,63-0,70 arası çok iyi, 0,55-0,62 arası iyi, 0,45-0,54 arası normal ve 0,32-0,44 arası zayıf olarak değerlendirilir (Tabacknick ve Fidell, 2007).

Gözlenen değişkenlerde örtük değişken tarafından açıklanamayan varyans ölçme hatası veya hata varyansı olarak adlandırılır. YEM çalışmalarının en önemli avantajı yapılan ölçmelerdeki hatanın modellenemesine olanak sunmasıdır. Diğer DFA parametreleri ise her bir faktörün örneklemdaki varyansını gösteren faktör varyansı ve gözlenen değişkenler arasında örtük değişkenin etkisi haricinde bir nedenden dolayı oluşan hata kovaryanslarıdır. Faktörler arasındaki ilişkileri ise faktör kovaryansları/korelasyonları göstermektedir (Harrington, 2009).

DFA, gözlenen değişkenlerle örtük değişkenlerin arasındaki ilişkiyi ifade eden bir regresyon modelidir. Örneğin,  $i$  bireyli,  $j$  gözlenen değişkenli ve  $p$  faktörlü bir ölçme modelinde,  $i$  bireyinin  $j$  gözlenen değişkenindeki puanı kovaryans modeliyle şöyle gösterilir:

$$x_{ij} = \sum_{p=1}^P \lambda_{jp} \xi_{pi} + \delta_{ij} \quad (1)$$

Denklemden  $(x_{ij})$ ,  $i$  bireyinin  $X_j$  değişkenindeki puanını,  $(\lambda_{jp}) X_j$ 'nin  $\xi_p$  faktör puanındaki regresyon doğrusunun eğimini (faktör yükü) ve  $(\delta_{ij})$   $i$  bireyinin  $X_j$  değişkenindeki hata varyansını göstermektedir.

DFA'dan önce kullanılacak veri seti, aykırı değerler, kayıp veri, tek değişkenli ve çok değişkenli normallik, doğrusallık (linearity), tekli bağlantılılık (singularity) ve çoklu bağlantılılık (multicollinearity) sayılıtları açısından incelenmelidir (Harrington, 2009). Ayrıca örtük değişkenlerin psikometrik açıdan ne kadar iyi olduklarının belirlenmesi yani geçerlik ve güvenilirlik kanıtlarının sunulması da gerekir (Hair, Anderson, Tatham ve Black, 1998; Şimşek, 2007).

DFA modellerinin test edilmesinde beş aşamadan oluşan bir yol takip edilmektedir (Schumacker ve Lomax, 2004):

- Modelin Oluşturulması (Model Specification)
- Modelin Tanımlanması (Model Identification)
- Model Parametrelerinin Kestirilmesi (Model Estimation)
- Modelin Test Edilmesi (Model Testing)
- Modelin Düzeltilmesi (Model Modification)

DFA modellerinin oluşturulmasında teori ve yapılmış araştırmalar oldukça önemlidir. DFA, modelin oluşturulduğu yapı hakkında güçlü bir temele sahip olunmadığı durumlarda uygun bir analiz yöntemi olmayabilir (Harrington, 2009). Model oluşturma aşamasında araştırmacı incelemek istediği değişkenler ve aralarındaki ilişkilerle ilgili literatürü detaylı bir şekilde taramalı ve değişkenlerin doğası hakkında kapsamlı bir bilgiye sahip olmalıdır. Eldeki tüm bilgiler kullanılarak hangi değişkenlerin teorik modele alınıp alınmayacağına ve bunların nasıl bir ilişki içinde olduklarına karar verilir. Model kurulurken teorik çerçevenin ve gerekçelerin sağlam bir şekilde oluşturulması gerekir (Şimşek, 2007). Yanlış oluşturulmuş bir model hatalı parametre kestirimleri

üretecek diğer bir ifadeyle parametre tahminleri gerçek modelde olduğundan farklı olacaktır. Yapısal eşitlik modellemesinin en zor kısmı burasıdır (Schumacker ve Lomax, 2004).

DFA'da parametre kestirimlerini elde edebilmek için ölçme modelinin tanımlanmış olması gerekmektedir. Veri setindeki bilgiler yardımıyla değerleri bilinmeyen her bir parametre için sadece bir tane parametre tahmini elde edilebiliyorsa ölçme modeli tanımlanmıştır (Brown, 2006).

Model tanımlandığında değeri bilinmeyen her bir parametre için tek bir (tekil) kestirim elde edilebilir. Örneğin,  $a + b = 44$  eşitliğinin sonsuz çözümü vardır. Bu durumda model (eşitlik) eksik tanımlanmıştır çünkü tek bir çözüm elde etmek için yeterli sayıda bilinen parametre yoktur. Diğer bir ifadeyle bilinmeyenler (a ve b) bilinenlerden (44) fazladır. DFA modelleri eksik (underidentified), tam (just identified) veya fazla tanımlanmış (over identified) olabilir (Harrington, 2009).

Serbestçe kestirilen parametre sayısı (bilinmeyen), bilinenlerden fazla ise model eksik tanımlanmıştır. Böyle durumlarda serbestlik derecesi negatiftir ve model tek bir çözüme ulaşamaz. Bazı parametreler sabitlenerek bilinmeyen sayısı azaltılabilir. Bilinen sayısı bilinmeyen sayısına eşit olduğunda model tam tanımlanmıştır ve serbestlik derecesi 0'dır. Bu durumda tek bir çözüm vardır ve mükemmel uyum verir. İlk bakışta iyi bir durum gibi gelse de pratikte mükemmel uyum veren modeller istenilmez; çünkü model, test etmeye olanak vermezler. Bilinmeyen sayısı bilinen sayısından az ise modeller fazla tanımlanmıştır ve serbestlik derecesi 0'dan büyüktür. Pratikte modellerin fazla tanımlanmış olması istenilir (Harrington, 2009).

Modelin tanımlanmış olması için gerekli koşullardan birisi de örtük değişkenlerin ölçeklenmesidir. Örtük değişkenler gözlenemediğinden daha önceden belirlenmiş bir ölçme birimleri yoktur ve bu yüzden araştırmacının bir ölçme birimi belirlemesi gerekir. Ölçekleme işlemi genellikle iki yöntemle yapılmaktadır. Birincisi örtük değişkenin ölçme birimini gözlenen değişkenlerden birininkiyle (referans değişken) aynı yapmaktır. İkinci yöntem ise örtük değişkenin varyansını 1'e eşitlemektir. Genelde

birinci yöntem daha popülerdir (Brown, 2006; Kline, 2005). Her iki yöntem de benzer model uyumu verse de bu her zaman geçerli değildir ve seçilen ölçekleme yöntemi DFA sonuçlarını ve standart hataları etkileyebilir (Brown, 2006; Kline, 2005; Vandenberg ve Lance, 2000).

DFA'nın amacı ölçme modelinin parametre kestirimlerini (faktör yükleri, regresyon sabiti, faktör varyansları, kovaryansları ve ortalamaları, hata varyansları ve kovaryansları) elde etmektir. Elde edilen bu parametreler örneklem varyans-kovaryans matrisini (S) en yakın temsil eden tahmini varyans-kovaryans matrisini ( $\Sigma$ ) oluşturmaktadır (Brown, 2006). Genel model uyumunun değerlendirilmesinde  $\Sigma$  ve S matrisleri karşılaştırılır. Matrisler arasındaki fark sıfıra yakınsa model veriye uyum gösteriyor şeklinde yorumlanır.

Parametre kestirimlerinde sıklıkla kullanılan en çok olabilirlik (Maximum Likelihood, ML) yöntemidir. Bununla birlikte DFA ve YEM analizlerinde diğer birçok yöntem de vardır. ML yönteminin kullanılmasının uygun olmadığı durumlarda (kategorik, sıralı veri, çok değişkenli normallik sayılıtısının karşılanamaması vb.) diğer yöntemlere başvurulmaktadır. Bir veya daha fazla değişkenin kategorik olması durumunda, ağırlıklandırılmış en küçük kareler (Weighted Least Squares, WLS), robust ağırlıklandırılmış en küçük kareler (Robust Weighted Least Squares, WLSMV) veya ağırlıklandırılmamış en küçük kareler yöntemleri kullanılabilir (Unweighted Least Squares, ULS) (Brown, 2006).

DFA'da birçok parametre kestirim yöntemi verinin sürekli olduğunu varsayar. Analizlerde kullanılan veri setleri çoğu zaman kategorik veya sıralı değişkenler içermektedir. Kullanılan yanıt kategorilerinin sürekliymiş gibi ele alınıp alınmayacağı her zaman net değildir. Bazı veriler net olarak kategorik (ırk gibi) bazıları ise sürekli (yaş gibi). Birçok ölçme aracı ise Likert tipi yanıt kategorilerine sahiptir. Az sayıda yanıt kategorisinin olduğu durumlarda verinin sürekliymiş gibi ele alınması ve kategorik yapısının göz ardı edilmesi bazı parametre kestirim yöntemlerinde hatalı sonuçların elde edilmesine neden olacaktır (Flora ve Curran, 2004; Raykov ve Marcoulides, 2006). Lubke ve Muthen (2004), yaptıkları simülasyon çalışmasında, sıralı

değişkenlerin sürekli değişkenler gibi kabul edilerek analiz edilmesinin, ölçme değişmezliği analizi sonuçlarını da etkileyeceğini göstermişler ve gruplar arası ortalama karşılaştırmalarının hatalı olabileceğini belirtmişlerdir. Bu yüzden verilerin normal olmadığı veya kategorik olduğu durumlarda ML yönteminden başka bir yöntemin seçilmesi gerekmektedir (Brown, 2006; Flora ve Curran, 2004).

Kategorik veya sıralı değişkenlerle kullanılacak parametre kestirim metotları WLS, WLSMV veya ULS yöntemleridir. Bunlar içinde en çok kullanılanı WLS metodudur ve çok değişkenli normallik sayıltısı gerektirmez. Bu yöntemin dezavantajı ise çok büyük örneklere ihtiyaç duymasıdır (Brown, 2006). Jöreskog ve Sörbom (1999, 2001, 2002) kategorik veriler için WLS veya WLSMV yöntemlerinin polikorik korelasyon matrisinden üretilen asimptotik kovaryans matrisiyle (AKM) beraber kullanılmasını tavsiye etmektedir.

Modelin test edilmesinde YEM programlarından birisi kullanılır. Analiz sonuçlarında öncelikle ilişkilerin beklendiği yönde ve istatistiksel olarak anlamlı olup olmadıklarına bakılır. Ayrıca modelin uyum değerleri bakımından istenilen değerleri üretilip üretilmediği incelenir. Modelin yeterli uyum vermediği veya araştırmacının daha iyi uyum değerleri elde etmek istediği durumlarda, analiz sonrasında üretilen düzeltme (modification) indisleri incelenir. Düzeltme indisleri modele eklenecek muhtemel yeni ilişkileri ve bunların modele katkısını gösterir. Her bir düzeltme indisinin teorik çerçeve temelinde değerlendirilmesi ve ele alınması gerekir. Düzeltmeler yapıldıktan sonra model tekrar analiz edilir ve model yorumlanarak raporlaştırılır (Şimşek, 2007).

Model veri uyumunun değerlendirilmesinde kullanılan birçok indeks vardır. Her bir indeks model uyumu hakkında farklı bilgiler vermektedir bu yüzden araştırmacılar genellikle birden fazla indeksi rapor etmektedir. Brown (2006) indekslerin Monte Carlo çalışmalarındaki performanslarına bakarak Ki-kare, RMSEA, SRMR, CFI ve NNFI'nın rapor edilmesini önermektedir. Bu çalışmada model uyumunun değerlendirilmesinde Ki-kare ile örneklem büyüklüğünden ve model karmaşıklığından diğer indeks değerlerine göre daha az etkilendikleri için RMSEA, SRMR, CFI ve NNFI değerleri kullanılmıştır.

Kabul edilebilir bir uyum için istatistiksel olarak anlamlı olmayan bir ki-kare,  $RMSEA < 0,08$ ,  $SRMR < 0,10$ ,  $CFI > 0,90$  ve  $NNFI > 0,90$  olmalıdır (Hair, Anderson, Tatham ve Black, 1998; Kline, 2005; Raykov ve Marcoulides, 2006).

### 1.3 ÇOKLU GRUP DOĞRULAYICI FAKTÖR ANALİZİYLE DMF ANALİZİ

DFA'nın kullanım amaçlarından biri de, ölçme ve yapısal modellerin gruplar arasında eşdeğerliğini inceleyebilmektir. Ölçme modeli, gözlenen değişkenlerin özelliklerini (faktör yükleri, regresyon sabiti ve hata varyansları) içerir. Yapısal model ise faktör varyanslarını, kovaryanslarını ve ortalamalarını içerir. Çoklu grup doğrulayıcı faktör analizi (ÇG-DFA), iki veya daha fazla grup üzerinde, DFA'nın her bir grubun varyans-kovaryans matrislerini (veya ham datalarını) kullanarak aynı anda çalıştırılmasıdır (Harrington, 2009).

Maddelerin ve faktör yapılarının gruplar arasındaki ölçme değişmezliği genellikle ÇG-DFA ile test edilmektedir. Bu noktada kovaryans modelini (Jöreskog, 1971) (covariance structure) ortalama ve kovaryans modelinden (mean and covariance structure) (Sörbom, 1974) ayırt etmek önemlidir; çünkü ÇG-DFA hem kovaryans modeli hem de ortalama ve kovaryans modeli datalarına uygulanabilir. Kovaryans modeli ve ortalama ve kovaryans modeli arasındaki en önemli farklılık ortalama ve kovaryans modelinin sadece varyans ve kovaryansları değil aynı zamanda gözlenen değişkenlerin ortalamalarını da modelleyebilmesidir (Wu, Li ve Zumbo, 2007). Gruplar üzerinde ölçme değişmezliğinin test edilebilmesi için analizlerin kovaryans modelinin yerine ortalama ve kovaryans modeliyle (OKM) yapılması gerekmektedir (Byrne, 2008; Little, 1997; Meredith, 1993).

Ölçülen yapıların eşdeğerliğinin ortaya konmasında ve gruplar arası farkların belirlenmesinde OKM'ye dayalı ÇG-DFA yöntemi oldukça uygundur; çünkü (a) faktör yapısının model uyumunun iki veya daha fazla grupta simültane olarak test edilmesine, (b) ölçme parametrelerinin eşdeğerliğinin (değişmezliğinin) gruplarda test edilmesine, (c) ölçmelerin hatadan arındırılmasına ve (d) ölçülen yapılar üzerindeki olası sosyo-kültürel etkilerin test edilmesine imkan verir (Little, 1997).

Doğrulayıcı faktör analizi bağlamında, ölçme değişmezliğinin teorik olarak tanımlanması, Meredith'in (1964, 1993) çalışmaları sayesinde olmuştur. Ölçme değişmezliği, belli bir örtük yetenek düzeyindeki bir bireyin test puanlarının beklenen değerinin bireyin içinde bulunduğu grup üyeliğinden bağımsız olmasını gerektirir (Meredith, 1993). Ölçme değişmezliği daha genel olarak Mellenbergh (1989) tarafından şu şekilde ifade edilmiştir.

$$f(Y|\eta, s) = f(Y|\eta) \quad (2)$$

Yukarıdaki eşitlik, verilen bir faktör puanı ( $\eta$ ) için, gözlenen puanların ( $Y$ ) olasılık dağılımının grup üyeliğinden ( $s$ ) bağımsız olduğunu göstermektedir. OKM ise şu şekilde gösterilebilir:

$$x_{ij} = \tau_j + \lambda_{j1}\xi_{1i} + \lambda_{j2}\xi_{2i} + \dots + \lambda_{jp}\xi_{pi} + \delta_{ij} \quad (3)$$

Bu denklemde  $x_{ij}$ ,  $i$ 'nci bireyin,  $j$ 'nci gözlenen değişkendeki puanını ifade etmektedir. Her bir yanıtın regresyon sabiti,  $\tau_j$ , bir veya daha fazla faktör,  $\xi_{pi}$ , ve normal dağılmış tesadüfi hata terimlerinin,  $\delta_{ij}$ , lineer bir kombinasyonu olduğu varsayılır. Regresyon katsayıları,  $\lambda_{jp}$ ,  $j$  maddesinin  $p$  faktöründeki faktör yükleri ve regresyon sabiti,  $\tau_j$ , faktör puanının 0 olduğu durumdaki  $x_{ij}$  puanıdır. Çoklu grup uygulamalarında (ÇG-DFA) bu modelin birden fazla grupta geçerli olup olmadığı test edilir.

Eşitliğin sağ tarafı yedi elementten oluşur (Wu, Li ve Zumbo, 2007):

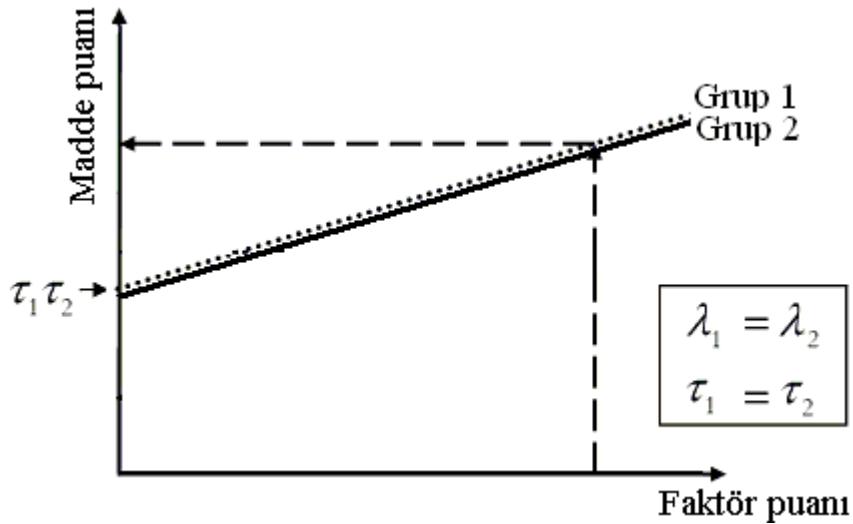
1. Model oluşturma (faktör sayısı ve faktör yüklerinin örüntüsü)
2. Regresyon katsayısı
3. Regresyon sabiti
4. Regresyon hata varyansı
5. Faktörlerin ortalamaları
6. Faktörlerin varyansları
7. Faktörlerin kovaryansları

İlk dört element gözlenen değişkenlerle örtük değişkenlerin ilişkisini inceleyen ölçme modeline aittir. Son üç tanesi ise örtük değişkenlerin dağılımı ve aralarındaki ilişkiyi inceleyen yapısal modele aittir (Vandenberg ve Lance, 2000; Wu, Li ve Zumbo, 2007).

Ölçme değişmezliğinin test edilmesinde, son üç element için gruplar arası eşitlik sınırlandırması yapmanın gerekli bir şart olmadığı konusunda literatürde bir fikir birliği vardır. Bu elementlerin eşitliği maddelerle faktörler arası ilişkilerin tanımlanmasında gerekli değildir (Little, 1997; Meredith ve Millsap, 1992; Widaman ve Reise, 1997).

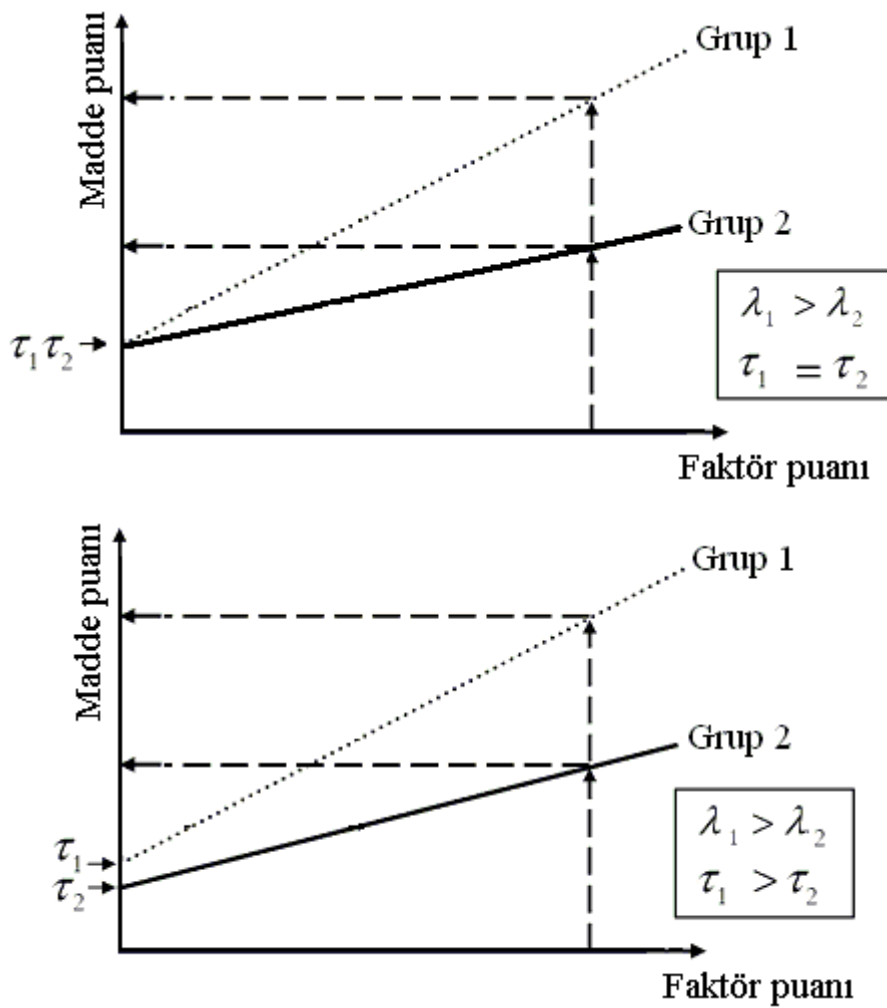
Maddeler veya ölçekler, gruplar arasında, temsil ettikleri yapıyla benzer ilişkiler gösteriyorsa ölçme eşdeğerliği (ölçme değişmezliği) vardır. MTK literatüründe ölçme değişmezliği DMF olarak adlandırılmaktadır (Stark, Chernyshenko ve Drasgow, 2006).

OKM'de, regresyon sabiti ( $\tau$ ) MTK'daki madde zorluğu parametresini (b), faktör yükü ( $\lambda$ ) ise madde ayırt ediciliği parametresini (a) temsil etmektedir. Faktör yükü ve/veya regresyon sabitlerinde değişmezliğin sağlanamaması DMF'nin varlığını göstermektedir. Regresyon sabitlerindeki farklılıklar tek biçimli (uniform), faktör yüklerindeki farklılıklar ise tek biçimli olmayan (non-uniform) DMF olarak adlandırılmaktadır (Chan, 2000).



Şekil 1.2 Ölçme Değişmezliğinin Sağlandığı İdeal Durum

Şekil 1.2’de iki farklı grup için madde puanlarıyla faktör puanlarının ilişkisini gösteren regresyon doğruları gösterilmiştir. Her iki grup için de faktör yüklerinin ( $\lambda$ ) ve regresyon sabitlerinin ( $\tau$ ) eşit olduğu görülmektedir. Bu şekil ölçme değişmezliğinin sağlandığı ideal bir durumu ifade etmektedir. Örtük yetenek (faktör puanı) üzerindeki herhangi bir değer için gözlenen (madde) puanların her iki grup için de aynı olduğu görülmektedir.

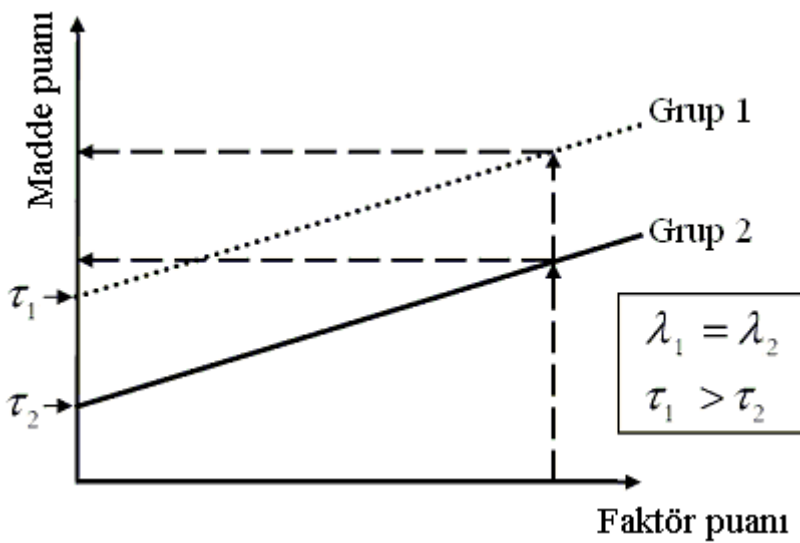


Şekil 1.3 Faktör Yükleri Eşitsizliğinin Ölçme Değişmezliği Üzerine Etkisi

Şekil 1.3’te faktör yükleri eşitsizliğinin ölçme değişmezliği üzerindeki etkisi gösterilmiştir. Faktör yüklerinin gruplar arasında eşitsizliği, maddelerin faktör puanlarına farklı ölçme birimleriyle kalibre edildiğini gösterir. Madde puanındaki bir birimlik değişme grupların faktör puanlarındaki aynı miktar değişmeye karşılık

gelmezse, eğimler aynı olmayacağından gruplar için regresyon doğruları farklı olur. Faktör yüklerinde değişmezliğin sağlanması maddeler için gruplarda aynı ölçme biriminin kullanıldığı ve gruplardaki bireylerin maddeleri benzer şekilde yorumladıkları ve yanıtladıkları anlamına gelir (Wu, Li ve Zumbo, 2007).

Şekil 1.3'te yeteneğin ne seviyede eksik kestirildiği yetenek seviyesine göre değişmektedir. Yüksek yetenek seviyeleri için daha fazla tahmin hatası yapılmaktadır. Bu durum tek biçimli olmayan yanlılık olarak adlandırılmaktadır (Mellenbergh, 1982).



Şekil 1.4 Regresyon Sabitleri Eşitsizliğinin Ölçme Değişmezliği Üzerine Etkisi

Şekil 1.4'te regresyon sabitlerindeki eşitsizliğin ölçme değişmezliği üzerindeki etkisi gösterilmiştir. Regresyon sabitlerinin farklılığının ölçme değişmezliğini ihlal ettiği açıkça görülmektedir. Şekil 1.4'te faktör yükleri her iki grupta da aynı iken, grup 2'nin regresyon sabiti grup 1'inkinden daha küçüktür. Yetenek seviyesinden bağımsız olarak, aynı yetenek düzeyindeki bireylerden grup 2 de olanlar grup 1'de olanlardan daha düşük puan elde edeceklerdir. Grup 2 deki eksik yetenek tahmini, tüm yetenek düzeyleri için aynı olacağından bu durum tek biçimli yanlılık olarak adlandırılır (Mellenbergh, 1982).

Ölçme değişmezliği testlerinde hata varyanslarının eşitliğinin test edilmesinin gerekliliği net değildir. Vandenberg ve Lance (2000) hata varyanslarının eşitliği testini araştırmacıların tercihinin bırakmıştır. Bu düşüncenin ardında yatan neden, hata

varyanslarındaki farklılıkların gözlenen puanların güvenilirliklerinin bir göstergesi olması düşüncesidir. Meredith (1993), Meredith ve Teresi (2006) ve Zumbo (2007) gibi araştırmacılar ise hata varyansının sistematik hata ve tesadüfi hatadan oluştuğunu ve sistematik hatanın grup üyeliğinden kaynaklanması durumunda regresyon doğrularının gruplar için farklı olacağını belirtmişlerdir.

Faktör yüklerinin ve regresyon sabitlerinin değişmezliği sağlandığında test puanları farklı kültürler arasında problemsiz ve geçerli bir şekilde karşılaştırılabilir (Schmitt ve Kuljanin, 2008; Van de Vijver ve Poortinga, 2005). Bunlara ek olarak hata varyanslarının değişmezliğinin de test edilmesi ölçme değişmezliği için ideal olsa da, hata varyanslarında değişmezliğin sağlanması gerçekte oldukça zordur ve yapılan birçok araştırma için gerekli değildir.

ÇG-DFA ile DMF analizi, temel model ile kademeli olarak model parametreleri gruplar üzerinde eşit olacak şekilde kısıtlanarak veya serbest bırakılarak oluşturulan iç içe modellerin test edilmesini içerir. İç içe modellerin karşılaştırılmasında ki-kare değerleri kullanılır. Kısıtlanan veya serbest bırakılan parametrelerin model uyumuna anlamlı bir katkısının olup olmadığına bakılır. Ki-kare değeri anlamsız ise ölçme değişmezliği sağlanmıştır denilir (Wu, Li ve Zumbo, 2007). Ölçme parametrelerinin değişmezliği gruplar üzerinde aynı yapının ölçüldüğü anlamına gelir. Ölçme değişmezliği testlerinden önce ölçme modelinin her bir grupta test edilmesi ve uyumunun incelenmesi gerekmektedir.

İç içe modellerle DMF belirlenirken, oluşturulan temel modellerin türlerinde literatürde iki farklı uygulama bulunmaktadır. Bazı araştırmacılar (Chan, 2000; Meade ve Lautenschlager, 2004) DMF analizlerinde sınırlandırılmış temel (constrained baseline) model kullanırken, bazı araştırmacılar ise (Reise, Widaman ve Pugh, 1993; Stark, Chernyshenko ve Drasgow, 2006) serbest temel (free baseline) modeli yaklaşımı tercih etmektedirler.

Sınırlandırılmış temel modelli yaklaşımda ölçekteki tüm maddelerin parametrelerinin  $(\lambda, \tau)$  gruplarda eşit olacak (değişmez) şekilde sınırlandırıldığı bir temel model oluşturulur. Diğer bir ifadeyle, başlangıçta tüm maddelerin DMF'siz olduğu varsayılır.

Daha sonra ardı ardına her seferde birer maddenin serbest bırakıldığı modellerle temel model arasındaki ki-kare farklılıkları incelenir. Anlamli olmayan ki-kare farklılıkları maddenin ölçme değışmezliđi sağladığını gösterir.

Serbest temel modellenli yaklaşımda ise sadece referans değışkenin parametrelerinin sınırlandırılıp diđer tüm maddelerin parametrelerinin  $(\lambda, \tau)$  serbest bırakıldığı bir temel model oluşturulur. Daha sonra ardı ardına her seferde birer maddenin sınırlandırıldığı modellerle temel model arasındaki ki-kare farklılıkları incelenir. Anlamli olmayan ki-kare farklılıkları maddenin ölçme değışmezliđi sağladığını ifade eder.

Stark, Chernyshenko ve Drasgow (2006) sınırlandırılmış temel modellenli yaklaşım ile serbest temel modellenli yaklaşımı yaptıkları simülasyon çalışmasında karşılaştırmış ve DMF belirlemede serbest temel modellenli yaklaşımın örneklem büyüklüğü, etki miktarı, DMF'nin türü ve miktarından bağımsız olarak sınırlandırılmış temel modellenli yaklaşımdan çok daha üstün performans gösterdiğini tespit etmişlerdir. Sınırlandırılmış temel modellenli yaklaşım sadece temel modelin DMF içermediđi durumlarda iyi çalışmış, maddeler DMF içerdiğinde ise çok yüksek birinci tip hatalar üretmiştir. Temel modelin DMF gösteren maddeler içermesi birinci tip hataları artırmasına rağmen gücü etkilememiştir. Bunun tersine serbest temel modellenli yaklaşım ise tüm durumlarda iyi çalışmıştır. DMF tespitindeki gücü oldukça yüksek olup birinci tip hatalar da 0,05'in altında çıkmıştır. Bonferroni düzeltmesi uygulandığında ise özellikle büyük örneklerde birinci tip hatalar neredeyse tamamen elimine olmuştur. Yazarlar DMF analizlerinde Bonferroni düzeltmeli serbest temel modellenli yaklaşımın (free baseline approach with banferroni correction) kullanılmasını önermişlerdir. Hernandez ve Gonzalez-Roma (2003) da yaptıkları simülasyon çalışmasında, sınırlandırılmış temel modellenli yaklaşımın, maddelerin DMF içerdiğı durumlarda yüksek birinci tip hatalar ürettiğini ortaya koymuşlardır.

Literatürdeki ÇG-DFA ile yapılan DMF analizlerinde genelde öncelikle faktör yüklerinin değışmezliđi test edilmekte daha sonra ise faktör yükleri değışmez olan maddelerin regresyon sabitlerinin değışmezliđi incelenmektedir (Oort, 1998).

Stark, Chernyshenko ve Drasgow (2006) DMF analizlerinde hem faktör yüklerinin hem de regresyon sabitlerinin aynı anda (simültane) incelenmesi gerektiğini belirtmekte ve bu iki basamaklı uygulama yöntemini şu nedenlerden dolayı eleştirmektedirler. Birincisi, geleneksel yaklaşım yani önce faktör yüklerinin incelenmesi daha sonra ise buna göre DMF göstermeyen maddeler için regresyon sabitlerinin incelenmesi gereksiz yere külfetli olabilir çünkü OKM, faktör yükleri ve/veya regresyon sabitlerinden dolayı olan DMF'ye aynı derecede hassastır. İkincisi, faktör yüklerinin ve regresyon sabitlerinin ayrı ayrı incelenmesi karşılaştırma sayısını ve dolayısıyla birinci tip hata olasılığı artmakta ve daha katı anlamlılık düzeyleri gerektirmektedir. Üçüncüsü ve belki de en önemlisi, regresyon sabitleri testi faktör yüklerinin kısmi değişmezlik sonuçlarına dayanıyorsa, birinci adımdaki DMF tespitindeki başarısızlık veya hata, sonraki seviyedeki DMF tespitinde hataları artıracaktır. Son olarak, ölçek revizyonu veya dil çevirisi gibi uygulamalı amaçlar için DMF'nin faktör yükleri ve/veya regresyon sabitleri üzerinde olmasının (tek biçimli veya tek biçimli olmayan DMF) bir önemi yoktur; çünkü her iki durumda da DMF'li madde revize edilmeli ve gerekiyorsa testten çıkarılmalıdır. Bu çalışmadaki DMF analizlerinde yazarların önerdiği Bonferroni düzeltmeli serbest temel modelli yaklaşım benimsenmiştir.

#### **1.4 REFERANS DEĞİŞKENİN BELİRLENMESİ**

Daha önce de ifade edildiği gibi, ölçme modellerinin ölçeklenmesinde her bir örtük değişken (faktör) için bir referans değişkenin atanması gerekmektedir. Tekli grup uygulamalarında referans değişken seçimi herhangi bir problem teşkil etmemektedir. Ancak çoklu grup uygulamalarında örtük değişkenlerin her birine her bir grupta da ortak bir referans değişkeni tanımlamak zorunludur. Bu durumda ise bu değişkenin her bir grupta ölçme değişmezliğine sahip olup olmadığını test etmek olanaksızdır. Dolayısıyla gözlenen değişkenlerden hangisinin gerçekten tüm gruplarda aynı parametre değerlerine sahip olduğunu belirlemek gerekmektedir (Johnson ve Meade, 2007; Kline, 2005; Şimşek, 2007).

Referans değişkenin belirlenmesi hem ölçek hem de madde bazındaki ölçme değişmezliği sonuçlarını oldukça etkileyebilmektedir. Çoklu grup uygulamalarında araştırmacı gerçekte değişmez olmayan bir değişkeni referans değişken olarak

belirlediğinde problemler ortaya çıkmaktadır. Birincisi araştırmacı bu maddenin değişmez olmadığını belirleyemeyecektir; çünkü referans değişkenin standardize edilmemiş faktör yükü gruplar üzerinde 1 olacak şekilde sabitlenmiştir. İkincisi, diğer maddelerin standardize edilmemiş parametreleri etkileneceğinden, sonraki kısmi değişmezlik testleri iyi uyum vermeyecektir. Dolayısıyla diğer maddelerdeki farklılıklar gruplar arasındaki gerçek farklılıkları yansıtmayacaktır (Bollen, 1989; Brown, 2006; Cheung ve Rensvold, 1999).

Referans değişken sistematik bir seçme prosedürü olmadan belirlendiğinde farklı seçimler oldukça farklı model uyum değerleri üretecektir (Millsap, 2001). Referans değişkenin DMF içermesi birinci tip hataları oldukça artırmaktadır (Stark, Chernyshenko ve Drasgow, 2006). Uygun bir referans değişken seçildiğinde yani referans değişken DMF içermediğinde, uygulanan ölçme değişmezliği testleri DMF gösteren maddeleri doğru bir şekilde tespit etmektedir (Johnson ve Meade, 2007).

Cheung ve Rensvold (1999) referans değişkenin belirlenmesi için bir prosedür geliştirmişlerdir. Referans değişkenle ilgili bugüne kadar yapılmış en kapsamlı araştırma niteliğindeki bu çalışmada, araştırmacılar faktör oran testinin kullanılmasını önermektedirler. Bu yöntemde her bir değişken sırayla referans değişkeni olarak alınmakta ve diğer her bir değişken değişmez olacak şekilde sınırlandırılarak iteratif bir şekilde referans değişkeni tespit edilmektedir. Bu yöntemin dezavantajı değişken sayısı arttıkça daha fazla karşılaştırma gerektirmesidir. Örneğin 10 değişken için bu prosedür 45 karşılaştırma gerektirmektedir.

Vandenberg (2002) referans değişkeni belirlemek için, oblimin veya promax eğik döndürme yöntemlerini uygulayarak açımlayıcı faktör analizinin temel eksenler (principal axis) veya ML prosedürleriyle birlikte kullanılmasını önermiştir. Açımlayıcı faktör analizi uygulandıktan sonra değişkenler sadece ilişkili oldukları faktör üzerindeki yükleri bakımından değil, diğer faktörler üzerindeki yükleri de göz önüne alınarak incelenmesi gerekmektedir. Gruplar arasında faktör yükleri bakımından en az değişkenlik gösteren madde referans değişkeni olarak kabul edilir.

Yoon ve Millsap (2007) tek faktörlü modelde düzeltme indislerini kullanarak gruplar arasında değişmez olan maddeleri belirlemişlerdir. Bu yöntem sadece maddelerin çoğunun gruplarda değişmez, faktör yükleri farkının büyük ve örneklemin de geniş olduğu durumlarda etkili sonuçlar vermiş diğer durumlarda ise kullanışlı olmamıştır. Diğer bir yöntem ise değişik maddeleri referans değişkeni olarak seçip ÇG-DFA uygulamalarını tekrar ve tekrar çalıştırmaktır (Brown, 2006).

Stark, Chernyshenko ve Drasgow (2006) çalışmalarında, sınırlandırılmış temel modelli yaklaşımın, maddelerin çoğununun DMF'li olduğu durumlarda bile, DMF tespitindeki gücünün yüksek çıktığını ortaya koymuşlardır. Araştırmacılara göre referans değişkeni belirlemek için, sınırlandırılmış temel modelli yaklaşım uygulanıp ortak faktör üzerinde faktör yükü en yüksek olan (muhtemelen) en yansız madde tespit edilebilir. Bu madde daha sonra serbest temel modelli yaklaşımda referans değişkeni olarak kullanılabilir. Bu çalışmada referans değişken (madde) belirlenirken Stark, Chernyshenko ve Drasgow'un (2006) önerdiği yöntem uygulanmıştır.

### **1.5 DEĞİŞEN MADDE FONKSİYONUNUN OLASI SEBEPLERİ**

DMF'nin sebepleri ve kaynakları bilinebilirse test geliştiriciler farklı test formlarının değişik diller arasında karşılaştırılabilirliğini garanti edebilmek için test geliştirme, çeviri ve uyarlama süreçlerini daha dikkatli denetleyebilirler. DMF analizleri ayrıca yeni politika ve uygulamaların (madde yazım yönergeleri vb.) gelişmesine rehberlik edebilir ve bu şekilde dil ve kültürler arası değişen fonksiyon gösteren maddeler azaltılabilir (Gierl ve Khaliq, 2001).

Yanıtlama stillerindeki sistematik farklılıklar veya tepki yanlılığı (response bias) kültürler arası çalışmaların geçerliğini ciddi anlamda etkileyebilmektedir. Katılımcılar maddeleri yanıtlarken bazen toplumsal açıdan arzu edilir veya kendisini toplumsal açıdan istenilir kılacak yönde cevaplar verme eğiliminde (social desirability) bazen içerikten bağımsız olarak olumsuz cevap yerine olumlu cevap verme eğiliminde (acquiescence) bazen de yine içerikten bağımsız olarak ölçeğin uç noktalarında yanıtlar verme eğiliminde olabilmektedirler (Van Herk, Poortinga ve Verhallen, 2004).

Allalouf, Hambleton ve Sireci (1999)'nin Angoff ve Cook (1988, s. 2)'tan aktardığına göre, dikkatli ve detaylı kontroller olmadan, çevrilmiş maddelerin çeviriden önceki orijinal grup için sahip olduğu aynı anlam ve göreceli zorluğa ikinci grup için de sahip olmasını beklemek çok olası değildir. Bir testin referans dilden odak dile çevrilmesi, muhakkak psikometrik olarak eşdeğer testler elde edileceği sonucunu vermez. Çeviri işlemi farklı diller arasındaki uzun test uyarlaması sürecinin ilk basamaklarından biridir. Test uyarlamasının nihai amacı her iki dilde de yapı eşdeğerliğini ve kapsamın temsil edilebilirliğini korumaktır. DMF analizi test geliştirme sürecinde önemli bir role sahiptir. Eğer farklı gruplardaki aynı yetenek düzeyindeki bireylerin maddeyi doğru yanıtlama olasılıkları eşit değilse, madde değişen bir fonksiyon gösteriyordur. Çevrilmiş maddelerdeki DMF'yi etkileyen faktörler bilinebilirse bunlar test geliştirme sürecinin ilk basamaklarında dikkate alınır ve test geliştirme, puanlama ve eşitleme hakkında daha iyi ve gelişmiş kararlar verilebilmesini sağlar (Allalouf, Hambleton ve Sireci, 1999).

Madde yanlılığı çoğunlukla yetersiz çeviriler, orijinal maddedeki anlam belirsizliği, madde içeriğinin belirli kültürlerdeki aşinalığı/uygunluğu veya madde yazılış tarzı ile ilgili kültüre özgü imalardan dolayı ortaya çıkmaktadır (Van de Vijver ve Tanzer, 2004).

Zayıf çeviriden kaynaklı madde yanlılığına güzel bir örnek Hambleton (1994) tarafından verilmiştir. İngilizce ve İsveççe dillerinde yapılan bir başarı karşılaştırmasında İngilizce konuşan adaylara şu soru sorulmuştur:

Where is a bird with webbed feet most likely to live?

- a) in the mountains
- b) in the woods
- c) in the sea
- d) in the desert

Maddenin İsveççe çevirisinde “webbed feet” ifadesi “swimming feet” şeklini alarak İsveççe konuşan öğrencilere doğru yanıt hakkında açık bir ipucu vermiştir. Dolayısıyla madde İsveççe konuşan öğrencilere haksız bir avantaj sağlamıştır.

Çevrilmiş sözel maddelerdeki DMF'nin sebeplerini bulmak amacıyla Allalouf, Hambleton ve Sireci (1999) İsrail'de üniversite başvurularında kullanılan Psikometrik Giriş Testi'nin İbranice ve Rusça versiyonlarının sözel maddelerini analiz etmişlerdir. Çalışmaları sonucunda DMF'nin sebeplerini (1) kelime veya cümlelerin zorluğundaki değişimler, (2) yetersiz çeviriden kaynaklı içerikteki değişimler, (3) madde formatındaki değişimler ve (4) madde içeriğinin kültürel uygunluğu olarak sıralamışlardır. Sireci ve Allalouf (2003) da benzer sonuçları rapor etmiştir.

Kanada'da uygulanan Okul Başarısı Göstergeleri Programı (SAIP) testlerinin İngilizce ve Fransızca versiyonlarının karşılaştırılabilirliğini değerlendirdikleri çalışmalarında Ercikan ve diğerleri (2004), maddelerin değişen fonksiyon göstermesinin kaynağının eğitim programları farklılıkları ile uyarılma etkileri olduğu hipotezlerini test etmişlerdir. Değişen fonksiyon gösteren maddelerin tespitinde 1997, 1998 ve 1999 yıllarında uygulanan okuma, matematik ve fen bilimleri testleri kullanılmıştır. Yazarların elde ettiği sonuçlar test versiyonlarının madde seviyesinde önemli psikometrik farklılıklar gösterdiğini ortaya koymuştur. Maddelerin yaklaşık %18 ile %36'sının DMF gösterdiği belirlenmiş ve maddelerin farklı fonksiyon göstermesinin çoğunlukla uyarılma kaynaklı farklılıklardan, az bir kısmının ise eğitim programları farklılıklarından kaynaklandığı tespit edilmiştir.

Kelime kelime çeviri bazen dereceleme ölçeğindeki noktalar arasında küçük veya büyük psikolojik boşluklar oluşturmaktadır. Örneğin, bir dilde izin vermek (allow) ve izin vermemek (not allow) arasında puanlanan bir ölçek, başka bir dilde izin vermek (allow) ve yasaklamak (forbid) arasında puanlanacak şekilde ifade edilmiştir. Yasaklamak ifadesi izin vermemek ifadesinden daha olumsuz bir anlam içerdiğinden ölçeğin ikinci dildeki kullanımını anlamlı bir şekilde etkilemiştir. Dereceleme ölçeklerinin çevirisi açıkça kelime kelime çeviriden daha öte bir şeydir. Sözcüklerin ifade tarzındaki en ufak bir farklılık bile madde yanıtlarında farklılığa neden olabilmektedir (Harkness, 1998).

Ercikan (1998) uluslararası değerlendirmelerde çeviri etkilerini incelediği çalışmasında, tespit ettiği çeviri kaynaklı olası DMF sebeplerini üç başlık altında toplamıştır: (a) sözcüklerin frekans, zorluk ve yaygınlığındaki, (b) cümlelerin uzunluk veya

karmaşıklığındaki ve (c) sözcüklerin içeriksel anlamındaki farklılıklar. Ercikan'ın, Brislin, Lonner ve Thorndike (1973)'tan aktardığı çeviri önerileri şunlardır:

1. Kısa ve 16 kelimedenden az olacak basit cümleler kullanılmalı,
2. Edilgen yerine etken cümle yapısı kullanılmalı,
3. Zamir kullanmak yerine isimler tekrar edilmeli,
4. Mecaz ve günlük konuşma dili kullanılmamalı,
5. İstek ya da dilek kipinde cümlelerden kaçınılmalı,
6. “Ne zaman”, “nerede”, “sık” veya “öte” gibi zarf veya edatların kullanımından kaçınılmalı,
7. Sahiplik bildiren yapıların kullanımından olduğunca kaçınılmalı,
8. Genel terimler yerine spesifik terimler kullanılmalı,
9. Belirsizlik bildiren “muhtemelen, sıklıkla vb.” kelimelerin kullanımından kaçınılmalı ve
10. Fiilleri farklı eylem bildiren çift fiilli cümleler kullanılmamalıdır.

Ercikan (2002) TIMSS 1995 verisini kullanarak DMF kaynaklarını çözümlenmeye çalışmıştır. Araştırmasında, DMF uyarlama etkilerinden dolayıdır ve DMF eğitim programlarındaki farklılıklarından kaynaklanmaktadır hipotezlerini test etmiştir. Bu amaçla Kanada, ABD, İngiltere ve Fransa’da uygulanan TIMSS matematik ve fen bilimleri testlerinin İngilizce ve Fransızca versiyonlarını incelemiştir. Çalışma sonucunda DMF kaynaklarının test uyarlaması ve eğitim programındaki farklılıklar olabileceğine dair kanıtlar bulmuştur. Son bir not olarak, DMF’nin birçok kaynağının olabileceğini ve karşılaştırılan gruplara ve yapılan değerlendirmenin tipi ve amacına bağlı olarak DMF’nin kaynaklarının ve anlamının değişebileceğini belirtmiştir.

Elosua ve Jauregúi (2007), test uyarlamasında karşılaşılan olası DMF kaynaklarını çeviri problemleri, gramer farklılıkları, semantik farklılıklar ve kültürel uygunluk olarak sınıflamışlardır. Araştırmacılara göre bu kaynakların testlerin uyarlanması üzerindeki etkisi farklı dil ailelerine ait dillerde ve kültürel açıdan daha fazla farklılık gösteren gruplarda daha fazla olmaktadır. Schulz (2005)’a göre eğitim sistemlerindeki farklılıklar da öğrencilerin soruları nasıl algıladığına ve yorumladığına etki edebilir.

Sireci ve Berberoğlu (2000) Orta Doğu Teknik Üniversitesi'nde kullanılan kurs değerlendirme formunun orijinal olan İngilizce ve uyarlanmış Türkçe versiyonlarının eşdeğerliğini her iki dilde de oldukça yeterli öğrencilere uygulayarak incelemişlerdir. Araştırmacılar dile özgü cinsiyet belirten ibarelerin kullanımının ve Likert tipindeki yanıt kategorilerinin (tamamen katılıyorum, katılıyorum vb.) diller arasında farklı algılanmasının DMF kaynağı olabileceğini saptamışlardır.

Kanada'da uygulanan matematik ve sosyal bilimler başarı testlerinin İngilizce ve Fransızca versiyonlarını karşılaştırdıkları çalışmalarında (Gierl ve Khaliq, 2001) DMF kaynakları olarak (1) kelime ekleme veya çıkarmaların anlamı etkilemesini, (2) dile veya kültüre özgü kelime, ifade veya cümle yapısındaki farklılıkları, (3) dile veya kültüre özgü olmayan kelime, ifade veya cümle yapısındaki farklılıkları ve (4) madde formatındaki farklılıkları göstermişlerdir. Araştırmacıların bu çalışmaları Allalouf, Hambleton ve Sireci (1999)'nin bulgularıyla paralellik göstermiştir. Grisay (2003)'e göre uluslararası testlerdeki çeviri hataları maddelerin değişik fonksiyon göstermesindeki en önemli sebeptir ve kültürel farklılıklarla eğitim programlarındaki farklılıklarından kaynaklı DMF'lere göre daha sık karşılaşırlar.

Ercikan ve Koh (2005), TIMSS 1995 İngilizce ve Fransızca testlerinin eşdeğerliğini Kanada, İngiltere, Fransa ve ABD örneklemi üzerinde değerlendirmişlerdir. Araştırmacılar çalışmalarında ÇG-DFA ve MTK parametre karşılaştırma yöntemlerini kullanmışlardır. Analizler sonucunda matematik ve fen bilimleri testlerinin İngilizce ve Fransızca versiyonlarının ölçtüğü yapılar arasında önemli farklılıklar olduğu tespit edilmiştir. Karşılaştırmalarda matematik maddelerinin %14 ile %39, fen bilimleri maddelerinin ise %37 ile %79 arasında DMF gösterdiği belirlenmiştir. Yazarlar, matematik ve fen bilimleri performanslarının değerlendirilmesi amacıyla TIMSS sonuçları kullanılarak yapılan karşılaştırmalarda ciddi sınırlamalar olduğunu belirtmişlerdir.

PISA 2006 ön denemesi verisini kullanarak Walker (2007), Likert tipi tutum maddelerinin yanıtlanmasında farklı kültürel özellikteki ülkeler arasında ölçeğinin uç noktalarında yanıt verme eğiliminin olduğunu saptamıştır. Araştırmacı Likert tipi maddelerle ölçülen yapıların her bir ülke için ayrı olarak incelenmesi gerektiğini

vurgulamıştır. Paulhus (1991)'a göre, anket maddelerini yanıtlarken maddelerin ölçtüğü yapının haricindeki bir özelliğe bağlı olarak sistematik bir yanıtlama eğilimi gösterilmesi de bir yanlılık kaynağıdır. Hui ve Triandis (1989)'e göre iki farklı kültürel grup Likert tipindeki maddeleri yanıtlarken uç noktalarda farklı eğilim gösteriyorsa, hesaplanan korelasyon katsayıları etkilenecektir ve elde edilen faktörler gruplar arasında oldukça farklılık gösterecektir. Van Herk, Poortinga ve Verhallen (2004), olumlu cevap verme eğilimi ve ölçeğin uç noktalarında yanıt verme eğiliminin Akdeniz ülkelerindeki bireylerde Kuzeybatı Avrupa ülkelerindekilere göre daha fazla olduğunu tespit etmişlerdir.

PISA ve TIMSS uygulamalarında karşılaşılan çeviri ve uyarlamadan kaynaklı spesifik problemlere PISA çeviri ve uyarlama yönergelerinde (OECD, 2004, 2007b) örnekler verilmiş, test ve anket materyallerinin uyarlanmasında katılımcı ülkelere bu yönergelere uymaları beklenmiştir.

## **1.6 PISA VE PISA ÖĞRENCİ ANKETİ**

Öğrenciler gelecekte karşılaştıkları sorunların üstesinden gelmeye ne ölçüde hazırlar? Kendi fikirlerini etkili bir şekilde muhakeme, analiz ve ifade edebiliyorlar mı? Toplum ve ekonominin üretken bireyleri olarak hayatları boyunca gerekli olacak ilgi, merak ve becerilere ne derece sahiptirler? Ekonomik İşbirliği ve Kalkınma Örgütü (OECD), Uluslararası Öğrenci Değerlendirme Programı (PISA) 15 yaş öğrencilerine uyguladığı tarama (survey) araştırmaları ile bu sorulara yanıtlar bulmaya çalışmaktadır. PISA tarama araştırmaları OECD'ye üye ve üye olmayan bazı katılımcı ülkelerde üç yılda bir uygulanmaktadır.

PISA, okuma becerileri, matematik okuryazarlığı ve fen bilimleri okuryazarlığı konu alanlarındaki öğrenci yeterlikleri üzerine odaklanarak zorunlu eğitimin sonundaki öğrencilerin topluma tam bir katılım için gerekli olan bilgi ve kabiliyetleri ne derecede edindiklerini değerlendirmektedir. PISA uygulamalarında ayrıca öğrencilerin ev, aile ve okul ortamları, motivasyonları, kendileri hakkındaki görüşleri, öğrenme stilleri ve tutumlarıyla ilgili bilgiler de toplanmaktadır. Bunun için öğrenci, veli ve okul anketi gibi ölçme araçları kullanılmaktadır (MEB 2007; OECD, 1999, 2007a, 2009).

PISA üçer yıllık dönemler halinde uygulanmakta ve her bir dönemde bir konu alanına ağırlık verilmektedir. PISA 2000’de okuma becerileri, PISA 2003’te matematik okuryazarlığı ve PISA 2006’da ise fen bilimleri okuryazarlığı ağırlıklı olarak değerlendirilmiştir ve birinci dönem sona ermiştir. PISA 2009’dan itibaren ikinci dönem uygulamaları başlayacaktır. Ülkemiz ilk olarak 2003 yılında bu uygulamalara katılmıştır. Ülkemizin eğitimde hangi seviyede bulunduğu, giderilmesi gereken eksikliklerinin ve alınması gereken önlemlerin tespit edilmesi ve eğitim düzeyinin niteliğinin artırılması amacıyla OECD üyesi olarak ülkemiz PISA projelerine katılmaktadır (MEB, 2009; OECD, 2007a; Turner ve Adams, 2007).

PISA, OECD eğitim direktörlüğüne bağlı PISA yönetim kurulu tarafından yürütülmektedir. Kullanılan ölçme araçlarının geliştirilmesi, örneklem seçimi, verilerin analiz edilmesi, yorumlanması ve uluslararası raporların hazırlanması işlemleri bir konsorsiyum tarafından yapılmaktadır. Bu konsorsiyumun liderliğini ise Avustralya Eğitim Araştırmaları Konseyi (ACER) yapmaktadır. Ulusal düzeyde çeviri ve uyarlama işlemlerinin yapılması, projenin uygulanması, analizlerin yapılması ve ulusal raporların yazılmasını ise katılımcı ülkelerde belirlenen merkezler tarafından yapılmaktadır. Ölçme materyallerindeki kültürel ve dilsel eşdeğerliğin sağlanabilmesi için oldukça fazla emek ve kaynak harcanmaktadır. Bu amaç doğrultusunda çeviri yapılması, örneklem oluşturma ve veri toplama süreçlerinde uyulması zorunlu çok sıkı kalite güvence mekanizmaları uygulanmıştır. PISA uygulamalarının çerçevesi ve kavramsal temelleri katılımcı ülkelerdeki uzmanlar tarafından geliştirilmiş ve yapılan görüşmeler neticesinde fikir birliğine vararak onaylanmıştır (MEB, 2007, 2009; OECD, 2007a, 2009; Turner ve Adams, 2007).

Orijinal ölçme araçları İngilizce ve Fransızca olarak hazırlanmakta daha sonra ulusal dillere çeviri ve uyarlama için katılımcı ülkelere gönderilmektedir. Her bir ülke bu iki versiyonu kullanarak iki farklı bağımsız çeviri yapmakta (double translation) daha sonra ise bunları tek bir son versiyona dönüştürmektedir. Her bir dil versiyonu arasındaki denklığı sağlamak için ülkelerden, geliştirilmiş olan PISA çeviri ve uyarlama yönergelerine (OECD, 2007b) uymaları istenilmektedir. Ölçme araçlarının son ulusal versiyonları PISA tarafından eğitilmiş İngilizce, Fransızca ve hedef dilde uzman olan

profesyonel çevirmenler tarafından doğrulanmaktadır. Testler ve anketler hazırlandıktan sonra ön deneme uygulaması yapılmakta, verilerin analizi, psikometrik özelliklerinin incelenmesi, DMF analizlerinin ve gerekli düzeltmelerin yapılmasından sonra ana uygulamaya geçilmektedir. Ölçme araçlarının geliştirilmesi, ölçeklenmesi ve verilerin analiz edilmesinde MTK ve DFA metodolojilerinden yararlanılmaktadır (Grisay, 2002, 2003; OECD, 2002, 2004, 2007a, 2009).

PISA 2006 uygulaması OECD'ye üye 30 ülke ile üye olmayan 27 ülkeyi kapsamaktadır. Bu ülkelerdeki yaklaşık 20 milyon 15 yaş grubu öğrenciyi temsil edecek şekilde yaklaşık 400 000 öğrenci rasgele seçilmiştir. İki basamaklı tabakalı örnekleme yöntemi kullanılarak önce okullar daha sonra ise öğrenciler rasgele seçilmiştir. Yazılı sınav uygulaması yanında öğrencilere özgeçmişleri, öğrenme alışkanlıkları, motivasyonları ve fen bilimlerine yönelik tutumlarına ilişkin doldurması yaklaşık 30 dakika süren bir öğrenci anketi uygulanmıştır. Öğrenci anketindeki sorularla öğrenci performansıyla ilişkili sosyal, kültürel, ekonomik ve eğitimsel faktörlerin belirlenmesi amaçlanmıştır (MEB, 2007; OECD, 2005, 2007a; Turner ve Adams, 2007).

Öğrencilerin kazandığı bilimsel/teknolojik bilgi ve becerileri günlük yaşama nasıl uyguladıklarını değerlendirmeye ek olarak PISA 2006, öğrencilerin tutumları ve fen bilimleri ile ne kadar iç içe oldukları hakkında veri toplamaya çok fazla dikkat göstermiştir. PISA araştırmalarında tutumlar öğrencilerin fen bilimlerindeki yeterliklerinin anahtar bir parçası olarak görülmüştür (OECD, 2007a).

Tutumlara ek olarak PISA 2006 öğrenci anketinde, öğrenci ve ailesiyle ilgili demografik, eğitimsel ve kültürel bilgiler, öğrencilerin fen bilimlerine ilişkin görüşleri, fen bilimleri ile ilişkili kariyer planları, çevreyle ilgili düşünceleri, okul ve okul dışı zamanda öğrenmeye ayırdıkları zaman hakkında da bilgi toplanmıştır. Ankette çoğunluğu ikili ve dördümlü Likert tipinde kodlanmış 37 soru bulunmaktadır (OECD, 2007a).

PISA 2006'da öğrencilerin fen bilimlerine yönelik tutumları sadece öğrenci anketindeki sorularla değil aynı zamanda öğrenci performansının ölçüldüğü fen bilimleri testi içine entegre edilmiş bazı sorularla da ölçülmüştür. Tutumları ölçen sorular teori, kuram ve

yapılmış araştırma sonuçlarına dayanarak geliştirilmiştir. Oluşturulan her bir ölçek ve indisleri doğrulamak ve ülkeler arası karşılaştırılabilirliklerini sağlamak için DFA yönteminden faydalanılmıştır (OECD, 2002, 2007a). PISA 2006'da ölçülen, öğrencilerin fen bilimlerine yönelik tutumları Çizelge 1.2'de özetlenmiştir (OECD, 2007a):

#### Çizelge 1.2 PISA 2006'da Ölçülen Öğrenci Tutumları

##### Bilimsel Sorgulamaya Verilen Destek (support for scientific enquiry)

- Farklı bilimsel bakış açılarının ve argümanların göz önünde bulundurulmasının önemini vurgular.
- Gerçeklere dayalı bilginin ve mantıklı açıklamaların kullanımını destekler.
- Çıkarımda bulunurken mantıklı ve dikkatli hareket etmenin gerekliliğini vurgular.

##### Fen Bilimlerine Yönelik Öz İnanç (self belief as science learners)

- Bilimsel görevleri etkili bir şekilde yerine getirir.
- Bilimsel problemlerin çözümünde zorlukların üstesinden gelir.
- Güçlü bilimsel beceriler gösterir.

##### Fen Bilimlerine İlgi (interest in science)

- Fen bilimlerini ve fen bilimleriyle ilgili konu ve çalışmaları merak eder.
- Farklı kaynaklar ve yöntemler kullanarak daha fazla bilimsel bilgi ve beceri kazanmaya istekli olduğunu gösterir.
- Bilgileri araştırmaya istekli olduğunu, fen bilimlerine sürekli bir ilgisi olduğunu ve fen bilimleriyle ilişkili bir meslek seçmeyi düşündüğünü gösterir.

##### Çevre ve Kaynaklara Yönelik Sorumluluk (responsibility towards resources and environments)

- Sürdürülebilir bir çevrenin devamlılığı için kişisel bir sorumluluk hissi gösterir.
- Bireysel eylemlerin çevresel sonuçlarının farkında olduğunu gösterir.
- Doğal kaynakları korumak için harekete geçmeye gönüllü olduğunu gösterir.

Bilimsel sorgulamaya verilen destek fen bilimleri eğitiminin önemli bir amacı olarak kabul edilmiştir. Bilimsel sorgulamaları takdir etme ve destekleme, öğrencilerin bilimle ilişkili gerçek hayat durumlarıyla karşılaştıklarında bilimsel yollarla delil toplamaya, rasyonel biçimde sorgulama yapmaya, kritik ederek yanıtlamaya ve kendilerini ifade etmeye değer verdiklerini ima etmektedir. Öğrencilerin fen bilimlerindeki kendi yeteneklerine değer vermeleri fen bilimleri ile iç içe olmanın önemli bir parçası olduğundan, fen bilimlerine yönelik öz inanç değerlendirmeye alınmıştır. Literatürdeki

arařtırmalar fen bilimlerine ynelik erken ilginin ileriki bilimsel ğrenmeleri ve bilimsel/teknolojik alandaki kariyeri iyi bir řekilde yordadığını gstermiřtir. evre ve kaynaklara karřı sorumluluk ise gittike geliřen global bir endiředir (OECD, 2007a).

PISA arařtırmalarının en nemli amalarından birisi farklı anketler kullanarak ğrenci ve ğretmen zellikleriyle okuldaki ğrenme ortamları hakkında bilgi toplamaktır. ğrencilerin derslere karřı tutum, ilgi ve deęer yargıları nemli ğrenme rnleri olarak grlmekte, ğrenci performanslarındaki farklılıklar elde edilen bu bilgiler yardımıyla aıklanmaya alıřılmakta ve milletler arası karřılařtırmalar yapılmaktadır. Trkiye de dahil olmak zere dięer birok lkede eęitim programları elde edilen bu bilgiler iřığında yeniden yapılandırılmaktadır. Eęitim politikalarında nemli kararların alınmasına iřık tutacak bu tr uluslararası karřılařtırmanın anlamlı olabilmesi iin kullanılan testlerin ve anketlerin ltę yapıların eřdeęer olması gerekmektedir.

İlgili literatr incelendiğinde daha ok biliřsel testlerin eřdeęerlięinin gsterilmesine ynelik alıřmalar yapıldığı fakat anketlerin eřdeęerlięiyle ilgili arařtırmalara yeterince nem verilmedięi grlmektedir. Bu alıřmada PISA 2006 ğrenci anketinde yer alan bilimsel sorgulamaya verilen destek boyutunun Avustralya, Yeni Zelanda, Amerika Birleřik Devletleri ve Trkiye rneklemlerinde eřdeęer olarak llp llmedięi belirlenmeye alıřılacaktır.

Fen bilimlerine ynelik her bir tutum boyutu PISA 2006 uygulamasında oluřturulan indekslerle (alt boyut) llmeye alıřılmıřtır. Bilimsel sorgulamaya verilen destek ğrenci anketinde llen fen bilimlerinin genel deęeri (general value of science) ve fen bilimlerinin kiřisel deęeri (personal value of science) alt boyutlarından oluřmaktadır. PISA 2006 ğrenci anketinde, bilimsel sorgulamaya verilen destek tutumunu len maddelerin Avustralya, ABD ve Yeni Zelanda rneklemlerinde kullanılan İngilizce versiyonu Ek 1’de, Trkiye rnekleminde kullanılan Trke versiyonu ise Ek 2’de verilmiřtir.

## 1.7 PROBLEM CÜMLESİ

PISA 2006 öğrenci anketinin ölçtüğü yapılar farklı kültürler ve diller arasında ölçme eşdeğerliği göstermekte midir?

Alt Problemler

1. PISA 2006 Öğrenci Anketi kapsamında ölçülen “bilimsel sorgulamaya verilen destek” ölçme modeli Avustralya, Yeni Zelanda, Amerika Birleşik Devletleri ve Türkiye kültürlerinde aynı faktör yapısına sahip midir?
2. PISA 2006 Öğrenci Anketi kapsamında ölçülen “bilimsel sorgulamaya verilen destek” ölçme modelinin orijinal ve uyarlanan maddeleri kültürler ve diller arasında DMF göstermekte midir?
  - a. Maddeler Avustralya ve Yeni Zelanda arasında DMF göstermekte midir?
  - b. Maddeler Avustralya ve Amerika Birleşik Devletleri arasında DMF göstermekte midir?
  - c. Maddeler Avustralya ve Türkiye arasında DMF göstermekte midir?
3. Maddelerin kültürler ve diller arasında DMF göstermesinin olası nedenleri nelerdir?

## 1.8 ARAŞTIRMANIN AMACI

Bu çalışmanın amacı, PISA 2006 öğrenci anketinde yer alan yapıların kültürler arası eşdeğerliğinin ampirik olarak analiz edilmesidir. Diğer bir amaç ise test çevirisinden ve kültürel farklılıklardan kaynaklı DMF gösteren maddelerin tespiti ve DMF'nin olası sebeplerinin belirlenip incelenmesidir. Bu amaçla, PISA 2006 öğrenci anketindeki “bilimsel sorgulamaya verilen destek” tutumunu ölçen maddelerin oluşturduğu ölçme modelinin her bir kültürdeki faktör yapısı DFA ile incelenmiştir. Daha sonra, ÇG-DFA ile maddelerin farklı kültürler ve diller arasındaki ölçme değişmezliği yani DMF gösterip göstermediği belirlenmiştir. Son olarak ise, maddelerin DMF göstermesinin olası sebepleri uzman görüşlerine başvurularak değerlendirilmiştir.

## 1.9 ARAŞTIRMANIN ÖNEMİ

Araştırma PISA test maddelerinin kültürler arası eşdeğerliğinden ziyade uygulanan anketin eşdeğerliğinin incelendiği nadir araştırmalardan biri olması dolayısıyla literatüre önemli bir katkıda bulunacaktır. Bu çalışma ayrıca DMF analizlerinde serbest temel modellenmiş yaklaşımın kullanıldığı ülkemizdeki ilk çalışma niteliğindedir.

Eğer, anketin ölçtüğü psikolojik yapılar kültürler göre eşdeğer değilse ve maddeler kültürler göre yanlılık gösteriyorsa bu anketin ve anketten elde edilen verilerden yapılan karşılaştırmaların, yorumlamaların geçerliği ve anlamlılığı sorgulanabilir duruma gelecektir.

## 1.10 İLGİLİ ARAŞTIRMALAR

Grisay ve Monseur (2007) PISA 2000 ve 2001 verisi ile okuma becerileri parçalarının katılımcı ülkelere ait değişik dil grupları arasında madde seviyesinde eşdeğerliğini araştırmışlardır. Yazarlar madde formatının ve aynı pasajda yer alan maddelerin yerel bağımlılığının maddelerde DMF'ye sebep olabileceğini tespit etmişlerdir. Bu araştırmadaki diğer önemli bir bulgu ise kültürel ve coğrafi farklar minimize edilse bile (Kanada da İngilizce ve Fransızca konuşulan bölgeler) testlerin çevirisinin yapıldığı durumdaki DMF'li madde sayısının birbirinden oldukça uzak bile olsa aynı dili konuşan ülkeler (Amerika ve Yeni Zelanda) arasındaki DMF'li madde sayısından fazla olmasıdır. Araştırmacılara göre, çevirinin kalitesi ne olursa olsun, bir dilden başka bir dile bir testin çevrilmesi, testlerin eşdeğer olmaması sorunuyla her zaman karşı karşıyadır. Aynı dilin konuşulduğu ülkeler için tek bir ortak versiyonun geliştirilmesi ve madde formatıyla dilsel ve kültürel özellikler arasındaki etkileşimin daha detaylı olarak incelenmesi gerektiği önerilmiştir.

Grisay ve diğerleri (2007) PISA 2006 ön denemesi verisini kullanarak fen bilimleri sorularının ölçek bazında eşdeğerliğini araştırmışlardır. Madde zorluklarının farklarını kümeleme analizi yöntemiyle incelemişlerdir. Birkaç istisna dışında, aynı dili konuşan ülkeler, coğrafi olarak yakın olan ülkeler ve kültürel ve soyo-ekonomik benzerliği olan ülkeler aynı küme gruplarında toplanmıştır. Aynı analiz PISA 2006 asıl uygulama

verisine tekrar uygulanmış ve benzer sonuçlar elde edilmiştir. PISA 2006 fen bilimleri testindeki soruların tüm katılımcı ülkelerde aynı şekilde fonksiyon göstermediği sonucuna varılmıştır. Araştırmacılar daha sonra dilsel farklılıktan kaynaklı yanlılığın büyüklüğünü belirlemek için faktör analizi kullanarak madde zorluklarında açıklanan ve açıklanamayan varyans (yanlılık kaynağı) oranlarını karşılaştırmışlardır. Bu inceleme sonucunda özellikle Orta Doğu ve Asya ülkelerinde açıklanamayan varyans oranlarının diğer ülkelere göre oldukça yüksek olduğu bulunmuştur.

Schulz (2003) makalesinde PISA 2000 uygulamasında ölçülmeye çalışılan bazı yapıların ülkeler üzerinde geçerliğini ve eşdeğerliğini ÇG-DFA ve MTK yöntemleriyle incelemiştir. Zayıf ölçme değişmezliğini test ettiği çalışmada bazı ülkeler için maddelerin çevirisinde problemler olduğunu ortaya koymuştur. Ayrıca bazı maddelerin ölçmeye çalıştıkları yapıların yeterli birer temsilcisi olmadıklarını belirlemiştir.

Diğer bir çalışmada Schulz (2005) PISA 2006 ön deneme verisini kullanarak öğrenci anketinde yer alan maddelerin ölçme değişmezliğini, her bir grupta iki tane benzer dil ve kültürde ülke olacak şekilde, sekiz farklı grup üzerinde test etmiştir. Benzer dil ve kültürdeki ülkeler arasında zayıf değişmezlik sağlanırken, faktör yükleri tüm ülkeler arasında eşit olacak şekilde sınırlandırıldığında yeterli bir uyum elde edilememiştir.

Yıldırım (2008), PISA 2003 taramasında yer alan matematik okuryazarlığı sorularının Amerikan ve Türk öğrencilerden elde ettiği veri seti üzerinde DMF içerip içermediğini Mantel-Haenszel, olabilirlik oranı ve sınırlandırılmış faktör çözümlemesi yöntemleriyle araştırmıştır. Her üç yöntemin de ortak olarak bulunduğu maddeler dikkate alındığında, araştırmada incelenen maddelerin %32'sinde ölçülen özellik dışında performansı etkileyen unsurların varlığı tespit edilmiştir.

Uluslararası Okuma Becerilerinde Gelişim Projesi (PIRLS) 2001 uygulamasında yer alan okuma parçalarına ait testlerin psikometrik özellikleri Öğretmen (2006) tarafından incelemiştir. ÇG-DFA ölçülen yapıların kültürlere göre eşdeğer olup olmadığının tespitinde, MTK parametre karşılaştırma ve olabilirlik oranı testleri ise maddelerin kültürler arasında değişken bir fonksiyon gösterip göstermediğinin incelenmesinde

uygulanmıştır. Testlerin ölçtüğü yapıların referans ve odak kültürler arasında ölçme eşdeğerliğine sahip olmadığı ve toplam 98 sorudan 53 tanesinin kültürler arasında farklı fonksiyon gösterdiği araştırmacı tarafından tespit edilmiştir. Yüksek kavrama süreçlerini yoklayan maddelerde ve çok kategorili olarak puanlanan maddelerde daha fazla DMF gözlenmiştir.

Yıldırım ve Berberoğlu (2009) PISA 2003 matematik okuryazarlığı maddelerinde DMF gösteren maddeleri Sınırlandırılmış Faktör Çözümlemesi, Mantel-Haenszel ve MTK olabilirlik oranı yöntemleriyle belirlemeye çalışmıştır. Amerikan ve Türk öğrenciler araştırmanın örneklemini oluşturmuştur. Her üç yöntem tarafından 152 maddeden 10 tanesi DMF'li olarak işaretlenmiştir. Yetersiz çeviri, nicelik bildiren ve tam karşılığı olmayan kelimelerin kullanımı ve maddelerin bilişsel özellikleri uzmanlar tarafından DMF kaynakları olarak belirlenmiştir.

Yıldırım (2006) tez çalışmasında, Amerika ve Türkiye örnekleminde, TIMSS 1999 ve PISA 2003 Türkçe ve İngilizce versiyonlarında yer alan matematik testi maddelerini psikometrik özellikleri ve madde yanlılığı açısından incelemiştir. Matematik testi faktör yapısının her iki kültürde de karşılaştırılabilir olmadığı ortaya çıkmıştır. Madde düzeyindeki analizlerde ÇG-DFA, sınırlandırılmış faktör çözümlenmeleri, Mantel-Haenszel ve MTK olabilirlik oranı yöntemleri kullanılmış ve bunların arasında özellikle PISA çalışması için yüksek derecede uyum tespit edilmiştir. Sonuçta hem PISA hem de TIMSS çalışmalarında kullanılan bazı maddelerin kültürler arası karşılaştırmaya olanak sağlayacak şekilde denk olmadıkları ortaya konulmuştur. Bilgi düzeyinde ve rutin işlem becerisi, birden fazla durumu karşılaştırıp bir karara varılması ve bu kararların ifade edilmesini gerektiren gerçek yaşamla ilgili sorularda DMF tespit edilmiştir. Özellikle PISA soruları için eğitim programlarındaki farklılıkların ve çeviri problemlerinin DMF kaynağı olduğu iddia edilmiştir.

Le (2009), PISA 2006 ön uygulamasında yer alan fen bilimleri maddelerinin cinsiyet değişkeni bakımından DMF gösterip göstermediğini 60 farklı dil ve 50 farklı ülke verisi üzerinde incelemiştir. Analizlerinde MTK yöntemini kullanmıştır. Cinsiyete bağlı DMF'nin kullanılan madde formatı ve içerik alanıyla ilişkili olduğu tespit edilmiştir.

Değişik dil grupları üzerindeki analizlerde, maddelerin %24'ünün her iki cinsiyetten birine avantaj sağlayacak şekilde DMF içerdiği belirlenmiştir.

Diğer bir araştırmasında Le (2006) PISA 2006 ön uygulaması fen bilimleri maddelerinin farklı ülke, dil ve cinsiyet grupları üzerinde değişen bir fonksiyon gösterip göstermediğini incelemiştir. Cinsiyet, ülke ve dile bağlı DMF oranları sırasıyla yüzde 10, 25 ve 39-59 olarak tespit edilmiştir. Test çevirisiyle kültürel ve eğitim programlarındaki farklılıkların bu sonuçlara neden olabileceği belirtilmiştir.

Wu, Li ve Zumbo (2007) TIMSS 1999 verisiyle matematik testinin ölçme değişmezliğini incelemişlerdir. Öncelikle testte yer alan maddeleri içerik alanlarına göre gruplandırılıp (parceling) daha sonra her bir alt boyuttaki toplam puanlar hesaplanarak matematik testinin yedi farklı ülke verisi üzerinde eşdeğer olup olmadığı test edilmiştir. Ayrıca araştırmacılar benzer kültürel özelliklere sahip ülkeleri gruplandırarak kültürün ölçme değişmezliği testleri üzerindeki etkisini incelemişlerdir. Matematik testinin benzer kültürdeki ülkeler arasında ölçme değişmezliğini sağladığı farklı kültürdeki ülkeler arasında ise sağlamadığı ortaya konulmuştur.

PISA çeviri ve uyarlama yönergelerinde (OECD, 2004, 2007b) test ve anket materyallerinin çeviri ve uyarlanmasında sıklıkla karşılan problemlerden detaylıca bahsedilmiş ve PISA 2000, 2003 ve 2006 ön denemelerindeki çeviriden kaynaklı problemlere örnekler verilerek çeşitli önerilerde bulunulmuştur.

## 2. YÖNTEM

Bu bölümde araştırmanın türü, çalışma grubu, verilerin elde edilmesi, verilerin çözümlenmesi ve analizlerde kullanılan istatistik programlarına yer verilmiştir.

### 2.1 ARAŞTIRMANIN TÜRÜ

Araştırma, PISA 2006 öğrenci anketinin farklı kültürel ve dilsel gruplardaki psikometrik özelliklerinin ve ölçme eşdeğerliğinin incelenmesi yönü ile betimsel, uzman görüşlerine başvurularak maddelerin ölçme eşdeğerliği sağlamamasının nedenlerine karar verilmesi yönüyle de nitel bir özellik taşımaktadır.

### 2.2 ÇALIŞMA GRUBU

PISA 2006 uygulamasında ülkelere ait örneklem PISA konsorsiyumu tarafından iki basamaklı tabakalı örnekleme yöntemiyle belirlenmiştir. Bu çalışmaya Avustralya, Yeni Zelanda, ABD ve Türkiye örneklemi dahil edilmiştir. Ülkeler arasında maddelerin değişen fonksiyon göstermesinin kültürel farklılıklar ve/veya uyarlama problemlerinden mi kaynaklandığını belirlemek için bu ülkeler özellikle seçilmiştir.

Araştırma deseninin bu şekilde kurgulanması sayesinde DMF analizlerinde aynı kültür-aynı dil (AUS-NZL), farklı kültür-aynı dil (AUS-ABD) ve farklı kültür-farklı dil (AUS-TUR) karşılaştırmaları yapılabilmektedir. Örneklem büyüklüğünün fazla olması ve PISA ölçme araçlarının geliştirilmesinden sorumlu olan konsorsiyumun liderliğini yürüten şirketin Avustralya'da olması dolayısıyla Avustralya referans (kaynak) grup, diğer ülkeler ise odak (hedef) gruplar olarak belirlenmiştir.

Avustralya örneğinde 14170, Yeni Zelanda örneğinde 4823, ABD örneğinde 5611 ve Türkiye örneğinde 4942 öğrenci PISA 2006 kapsamında uygulanan öğrenci anketini yanıtlamışlardır. Öğrencilerin buldukları ülkelerin kültürel ve dilsel özelliklerini tam olarak yansıttıklarından emin olabilmek için testin uygulandığı ülkede doğan ve evlerinde testin uygulandığı dilin konuşulduğunu belirten öğrenciler

değerlendirmeye alınmıştır. Kayıp verilerle aykırı verilerin incelenmesi sonrasında oluşan, araştırmanın çalışma grubuna ait frekans ve yüzdeler Çizelge 2.1’de verilmiştir.

Çizelge 2.1 Çalışma Grubu Frekans ve Yüzdeleri

Gruplar	Frekans	Yüzde
Avustralya	11642	48,1
Yeni Zelanda	3500	14,5
ABD	4518	18,7
Türkiye	4535	18,7

### 2.3 VERİLERİN ELDE EDİLMESİ

Araştırma, Ekonomik İşbirliği ve Kalkınma Örgütü’nün (OECD) 2006 yılında düzenlemiş olduğu Uluslararası Öğrenci Değerlendirme Programı (PISA) kapsamında uygulanan öğrenci anketinden elde edilen veriler üzerinde yürütülmüştür. Çalışma için gerekli olan veri OECD PISA internet sitesinden ([www.pisa.oecd.org](http://www.pisa.oecd.org)) elde edilmiştir.

### 2.4 VERİLERİN ÇÖZÜMLENMESİ

Araştırmadaki veri analizleri, DFA sayıltılarının test edilmesi, ölçme modelinin uyumunun incelenmesi, referans değişkenin belirlenmesi, DMF gösteren maddelerin belirlenmesi ve DMF’nin olası sebeplerinin incelenmesi aşamalarından oluşmaktadır.

#### *DFA sayıltılarının test edilmesi*

Araştırma problemlerine ilişkin analizlere geçilmeden önce her bir ülkeye ait veri seti, aykırı değerler (outlier), kayıp veri (missing data), doğrusallık (linearity) ve tekli ve çoklu bağlantılılık (singularity, multicollinearity) sayıltıları açısından ayrı ayrı incelenmiştir. Parametre kestirimlerinde ağırlıklandırılmış en küçük kareler yöntemi (WLS) kullanıldığı (Jöreskog ve Sörbom, 1999, 2001, 2002) ve bu yöntemin de normallik sayıltısı olmadığı için veri setlerinde çok değişkenli normallik (multivariate) sayıltısı incelenmemiştir.

Her bir ülke verisindeki değişkenlere ilişkin kayıp veri oranları Ek 3'te belirtilmiştir. Kayıp veri oranları 0,004 ile 0,014 arasında değişmektedir. Her bir değişkendeki kayıp veri oranı %5'ten az olduğu ve sistematik bir yapı göstermediği için kayıp veriler analizden çıkarılmıştır (Tabacknick ve Fidell, 2007).

Avustralya, Yeni Zelanda, ABD ve Türkiye örneklemelerindeki tek değişkenli ve çok değişkenli aykırı değerleri tespit etmek için artıkların (residual) dağılımlarından faydalanılmıştır. Tek değişkenli aykırı değerler için standartlaştırılmış artıkların dağılımına, çok değişkenli aykırı değerler ve etkili (influential) gözlemlerin belirlenmesi için ise Mahalanobis uzaklıkları ve standartlaştırılmış DFFIT değerlerinin dağılımına bakılmıştır (Hair, Anderson, Tatham ve Black, 1998; Tabacknick ve Fidell, 2007). Aykırı değerler ve etkili gözlemler YEM'deki uyum indeksleri, parametre kestirimleri ve standart hatalar üzerinde etkili olabilirler. Bu yüzden bu tür gözlemlerin analizlerden önce tespit edilmesi ve rapor edilmesi gerekmektedir (Rensvold ve Cheung, 1999). Aykırı değerlerin analizine ilişkin grafikler Ek 4, 5, 6 ve 7'de gösterilmiştir. Tüm ülke verileri için standartlaştırılmış artıklar, (0,05) anlamlılık düzeyi için kritik t değerleri olan (-1,96; +1,96) aralığında yer almıştır. Dolayısıyla veri setlerinin tek değişkenli aykırı değerler içermediği tespit edilmiştir. Çok değişkenli aykırı değerler aynı zamanda etkili gözlem olup olmamaları açısından da incelenmiştir. Bu amaçla gözlemlere ilişkin Mahalanobis uzaklıkları ve standartlaştırılmış DFFIT değerlerinin dağılımları incelendiğinde, Avustralya, Yeni Zelanda ve Türkiye örneklemelerinde 2, ABD örnekleminde ise 3 adet gözlemin hem aykırı değer hem de sonuçlar üzerinde etkili gözlemler olduğu saptanmış ve veri setinden çıkarılmıştır.

Değişkenler arasındaki korelasyonların çok yüksek olmaması ve analizlerde kullanılan LISREL programının herhangi bir uyarı mesajı vermemesi dolayısıyla veri setlerinde tekli bağlantılılık olmadığı sonucuna varılmıştır (Tabacknick ve Fidell, 2007). Değişkenler arasındaki çoklu bağlantılılığın incelenmesi için CI (condition index), varyans şişme oranları (VIF) ve tolerans değerleri incelenmiştir. Çoklu bağlantılılığın olması için değişkenlere ait CI değerinin 30 ve üzeri, VIF değerinin 10'dan büyük veya tolerans değerlerinin 0,10 veya üzerinde olması gerekmektedir (Hair, Anderson, Tatham ve Black, 1998).

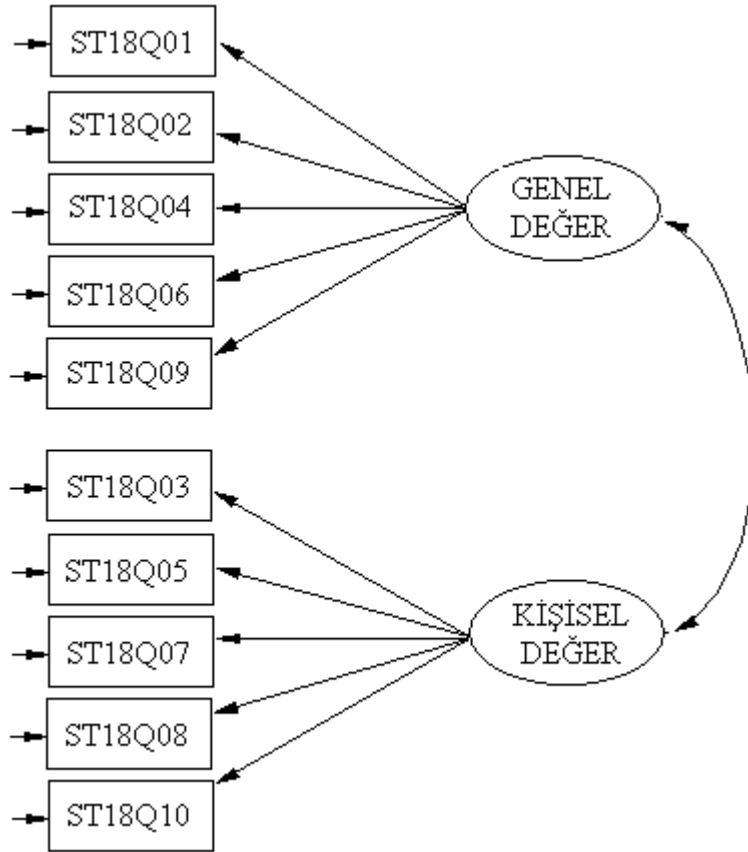
Bu incelemelere ilişkin çizelgeler Ek 8, 9, 10 ve 11’de verilmiştir. Her bir ülke verisine ait değerler incelendiğinde CI değerlerinin 30’dan, VIF değerlerinin 10’dan küçük olduğu ve tolerans değerlerinin de sıfırdan oldukça farklı olduğu ve sonuç olarak çoklu bağlantılılığın olmadığı görülmüştür. Doğrulayıcı faktör analizlerinde değişkenler arası doğrusal ilişkiler incelenmektedir. Bu araştırmadaki doğrusallık sayılıtısının her bir ülke verisi için test değerlendirilmesinde standartlaştırılmış artıklarla standartlaştırılmamış tahmini değerlerin dağılımları incelenmiş ve doğrusallık sayılıtısının da karşılandığı sonucuna varılmıştır. Yapılan analizler sonucunda DFA sayılıtlarının karşılandığı görülmüştür.

#### *Ölçme modelinin uyumunun incelenmesi*

Sireci ve Swaminathan (1996), farklı dillere uyarlanmış ölçme araçlarındaki maddelerin (orijinal madde ve uyarlanmış versiyonu) aynı olmadığını ve günümüzde kullanılan DMF analiz yöntemlerinin ise farklı gruplara uygulanmış aynı maddeleri değerlendirmek üzere dizayn edildiklerini belirtmiştir. Yazarlar uyarlaması yapılan maddelerin gruplar arasında farklı fonksiyon gösterip göstermediğinin belirlenmesinden önce ölçülen faktör yapısının her bir grup için ayrı ayrı incelenmesi gerektiğini vurgulamışlardır.

Analizlerin bu aşamasında PISA 2006 öğrenci anketi kapsamında ölçülmeye çalışılan, öğrencilerin fen bilimlerine yönelik tutumlarından birisi olan bilimsel sorgulamaya verilen destek ölçme modelinin birleştirilmiş veri seti ve ayrıca her bir ülke verisi üzerindeki uyumu değerlendirilmiştir.

Çalışmada incelenen bilimsel sorgulamaya verilen destek ölçme modeli Şekil 2.1’de verilmiştir.



Şekil 2.1 Bilimsel Sorgulamaya Verilen Destek Ölçme Modeli

Bilimsel sorgulamaya verilen destek ölçme modeli her biri beşer maddeli fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri alt boyutlarından oluşmaktadır. PISA 2006 öğrenci anketinde yer alan ST18Q01, ST18Q02, ST18Q04, ST18Q06 ve ST18Q09 kodlu maddeler öğrencilerin fen bilimlerine ilişkin genel değer yargılarını, ST18Q03, ST18Q05, ST18Q07, ST18Q08 ve ST18Q10 kodlu maddeler ise öğrencilerin fen bilimlerine ilişkin kişisel değer yargılarını ölçmektedir. Her iki boyutu da oluşturan tüm maddeler dörtlü Likert tipindedir. Maddeler Türkçe formunda tümüyle katılıyorum, katılıyorum, katılmıyorum ve hiç katılmıyorum, İngilizce formunda ise strongly agree, agree, disagree ve strongly disagree şeklinde yanıt kategorilerinden oluşmaktadır.

PISA 2006 öğrenci anketindeki maddeler sıralı (ordinal) veri biçiminde olduğundan belli bir ölçme birimleri ve sıfır noktaları yoktur. Jöreskog (2005) ve Lubke ve Muthen (2004) sıralı verilerin sürekli veri olmadığını ve sürekli veri kabul edilerek analiz

edilmemesi gerektiğini belirtmektedir. Jöreskog (2005)'a göre, her bir sıralı değişkenin ( $z$ ), temelinde yatan bir sürekli değişkenin ( $z^*$ ) olduğu varsayılır ve analizlerde  $z$  yerine  $z^*$  kullanılmalıdır. Böylece sürekli değişken ( $z^*$ ) sıralı değişkenin ( $z$ ) metriğini belirleyebilir. Sıralı değişkendeki kategorilere sürekli değişkende kategori sayısının bir eksiği kadar eşik değeri (threshold) karşılık gelmektedir. Jöreskog (2005) yapısal eşitlik modellerinde sıralı değişkenlerin analiz edilmesinde iki aşamalı bir yöntemi tavsiye etmektedir: Öncelikle PRELIS programıyla eşik değerleri kestirilmeli ve sonra bu eşik değerleri kullanılarak polikorik korelasyonlar ve ona karşılık gelen asimptotik kovaryans matrisi hesaplanmalıdır. Daha sonra ise bu matrisler ağırlıklandırılmış en küçük kareler (WLS) yöntemiyle model parametrelerinin kestirilmesinde LISREL programında kullanılmalıdır.

Jöreskog (2005)'un önerileri doğrultusunda bu çalışmada öncelikle her bir ülkeye ait veri setleri birleştirilmiş ve bu birleştirilmiş veri setinden PRELIS programı yardımıyla eşik değerler belirlenmiştir. Buna ilişkin SIMPLIS komut dosyası Ek 12'de verilmiştir. Daha sonra ise tüm ülkeleri ortak bir ölçek üzerinde karşılaştırabilmek için her bir ülke verisinin eşik değerleri elde edilen bu eşik değerlerine eşit olacak şekilde sabitlenmiş ve ülkelere ait polikorik korelasyon ve asimptotik kovaryans matrisleri hesaplanmıştır. Ülke verilerine ait eşik değerlerinin ortak eşit değerlerine sabitlenmesinde kullanılan SIMPLIS komut dosyalarından biri Ek 13'te verilmiştir.

PRELIS programının ürettiği matrisler ölçme modelinin uyumunun incelenmesinde, çoklu grup uygulamalarıyla referans değişkenin ve DMF'li maddelerin tespitinde LISREL programına okutulmuştur. LISREL programıyla yapılan tüm analizlerde ağırlıklandırılmış en küçük kareler (WLS) tahmin yöntemi kullanılmıştır. Her bir ülke verisine ait örneklem büyüklükleri 1000'den fazla olduğundan bu yöntemin dezavantajı olan büyük örneklem gerektirmesi analizlerde herhangi bir problem oluşturmamıştır (Flora ve Curran, 2004).

Doğrulayıcı faktör analizine ilişkin LISREL-SIMPLIS komut dosyasının bir örneği Ek 14'te verilmiştir. Bu değerlendirme çerçevesinde uygulanan her bir doğrulayıcı faktör

analizine ilişkin ülkelere ait faktör yükleri ve hata varyanslarının standartlaştırılmış değerleri ise Ek 15, 16, 17 ve 18’de verilmiştir.

Şekil 2.1’de verilen ölçme modelinin uyumunun değerlendirilmesinde öncelikle modeldeki değişkenlere ilişkin betimsel istatistikler incelenmiştir. Daha sonra modele ait genel uyum indeks değerleri, faktörlerin güvenilirlikleri, faktörler arası korelasyonlar ve her bir maddede ilişkili olduğu faktör tarafından açıklanan varyanslar ( $R^2$ ) incelenmiştir.

Uyum indeks değerleri, belirlenmiş yapısal eşitlik modelinin örneklem datasına ne kadar iyi uyum gösterdiğini belirtmektedir. Model data uyumunun değerlendirilmesinde Ki-kare, RMSEA, SRMR, CFI ve NNFI uyum indeksleri kullanılmıştır. PISA uygulamalarına ilişkin teknik raporlarda da RMSEA, RMR, CFI ve NNFI indeksleri rapor edilmektedir (OECD, 2009).

Faktörlerin güvenilirliklerinin incelenmesinde Cronbach’ın alpha değeri yerine Hair, Anderson, Tatham ve Black (1998)’in önerdiği yapı güvenirligi (construct reliability) ve açıklanan varyans oranı (variance extracted) değerleri kullanılmıştır. Yapı güvenirligi değerleri maddelerdeki hatalardan arınık olarak hesaplanan iç tutarlılığı, açıklanan varyans değerleri ise maddelerde, ilişkili oldukları faktörler tarafından açıklanan toplam varyansı ifade etmektedir. İyi bir uyum için yapı güvenirliginin 0,70 ve açıklanan varyans değerinin de 0,50 ve üzerinde olması önerilmektedir. Güvenirlige ilişkin bu değerlerin hesaplanmasında formül 4 ve 5 kullanılmıştır.

$$\text{Yapı Güvenirligi} = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum \varepsilon_j} \quad (4)$$

$$\text{Açıklanan Varyans} = \frac{\sum (\lambda_i^2)}{\sum (\lambda_i)^2 + \sum \varepsilon_j} \quad (5)$$

$\lambda_i$  = her bir maddeye ait standartlaştırılmış faktör yükü

$\varepsilon_j$  = her bir maddeye ait hata varyansı

### *Referans değişkenin belirlenmesi*

DFA'yı uygulayabilmek ve çoklu grup doğrulayıcı faktör analiziyle DMF gösteren maddeleri belirleyebilmek için her bir faktörün (örtük değişken) ölçeklenmesi gerekmektedir. Bu çalışmadaki faktörlerin ölçeklenmesinde referans değişken yöntemi kullanılmıştır. Referans değişkenin belirlenmesinde Stark, Chernyshenko ve Drasgow (2006) tarafından önerilen yöntem tercih edilmiştir. Referans değişkenlerin belirlenmesinde kullanılan sınırlandırılmış temel modele ait LISREL-SIMPLIS komut dosyasının bir örneği Ek 19'da verilmiştir. Hem referans değişkeni belirleme hem de DMF'li maddelerin belirlenmesinde Avustralya referans grup, Yeni Zelanda, ABD ve Türkiye ise odak gruplar olarak belirlenmiş ve analizlerde OKM kullanılmıştır.

Bu yönteme göre referans değişkenler belirlenirken yapılan her bir karşılaştırma için aşağıdaki adımlar izlenmiştir.

1. Tüm madde parametrelerinin (faktör yükleri ve regresyon sabitleri) referans ve odak gruplar arasında eşit olacak şekilde sınırlandırıldığı (sınırlandırılmış temel model) bir temel model oluşturulmuştur. Bu modelde referans grubun faktör ortalamaları sıfıra sabitlenmiş, odak grupların faktör ortalamalarının ise serbestçe kestirilmesine izin verilmiştir. OKM ile uygulanan çoklu grup doğrulayıcı faktör analizi bu model için bir ki-kare değeri,  $\chi^2(1)$ , üretmiştir.
2. Daha sonra herhangi bir maddenin (örneğin madde 1) faktör yükleri ve regresyon sabitleri gruplar üzerinde değişecek şekilde serbest bırakılmıştır. Bu model için de bir ki-kare değeri,  $\chi^2(2)$ , elde edilmiştir.
3.  $\chi^2(1) - \chi^2(2)$  farkı hesaplanmıştır. Bu fark 2 serbestlik derecesinde yaklaşık olarak ki-kare dağılımı göstermektedir.
4. Ki-kare farklılıklarının anlamlılığının değerlendirilmesinde Bonferroni düzeltmesi uygulanmıştır. Bonferroni düzeltmesi 10 karşılaştırma için uygulanırsa düzeltilmiş anlamlılık seviyesi ve bu anlamlılık düzeyi için kritik ki-kare değeri aşağıdaki gibi olmaktadır.

$$\alpha_{\text{düzeltilmiş}} = \frac{0,05}{10} = 0,005$$

$$p = 0,005 \text{ için } \chi^2(sd = 2) = 10,597$$

5.  $\chi^2(1)$ - $\chi^2(2)$  farkı hesaplanan bu ki-kare değerini geçtiği durumlarda madde DMF içeriyor şeklinde yorumlanmıştır.
6. Tüm maddeler için (1-5) arası adımlar tekrar edilmiştir.
7. DMF içermeyen (en yansız) ve ortak faktör üzerinde faktör yükü en yüksek olan madde referans değişkeni olarak belirlenmiştir.

Referans değişkenin belirlenmesine yönelik analiz sonuçları Ek 20, 21, 22'de verilmiştir. Avustralya-Yeni Zelanda karşılaştırması için ST18Q01 ve ST18Q08, Avustralya-ABD karşılaştırması için ST18Q02 ve ST18Q05 ve Avustralya-Türkiye karşılaştırması için ST18Q06 ve ST18Q07 kodlu maddeler referans değişkenler olarak belirlenmiştir. Avustralya-Türkiye karşılaştırmasında ST18Q06 kodlu madde DMF içermesine rağmen en düşük ki-kare değerine sahip olduğu için referans değişken olarak kabul edilmiştir.

#### *DMF gösteren maddelerin belirlenmesi*

Bu çalışmada incelenen ölçme modelinin kültürler arası eşdeğerliği madde bazındaki analizlerle test edilmiştir. Bu bağlamda Avustralya, Yeni Zelanda, ABD ve Türkiye örneklemeleri arasında DMF gösteren maddeler tespit edilmiştir. DMF'li maddelerin belirlenmesinde Stark, Chernyshenko ve Drasgow (2006) tarafından önerilen Bonferroni düzeltmeli serbest temel modellenli yaklaşım kullanılmıştır ve kullanılan serbest temel modele ait LISREL-SIMPLIS komut dosyasının bir örneği Ek 23'te verilmiştir.

Her bir karşılaştırma için aşağıdaki adımlar izlenmiştir.

1. Referans değişkenlerin faktör yüklerinin ve regresyon sabitlerinin sınırlandırılıp diğer tüm madde parametrelerinin serbest bırakıldığı bir temel model (serbest temel model) oluşturulmuştur. Bu modelde referans grubun faktör ortalamaları sıfıra sabitlenmiş, odak grupların faktör ortalamalarının ise serbestçe kestirilmesine izin verilmiştir. Ortalama ve kovaryans modeliyle uygulanan çoklu grup doğrulayıcı faktör analizi bu model için bir ki-kare değeri,  $\chi^2(1)$ , üretmiştir.

2. Referans deęişkenine ek olarak sadece bir maddenin (örneğin madde 1) faktör yükleri ve regresyon sabitleri gruplar üzerinde deęişmez olacak şekilde sınırlandırılmıştır. Bu model için de bir ki-kare deęeri,  $\chi^2(2)$ , elde edilmiştir.
3.  $\chi^2(2) - \chi^2(1)$  farkı hesaplanmıştır. Bu fark 2 serbestlik derecesinde yaklaşık olarak ki-kare dağılımı göstermektedir.
4. Ki-kare farklılıklarının anlamlılıęının deęerlendirilmesinde Bonferroni düzeltmesi uygulanmıştır. Ölçme modelinde 10 madde olmasına karşın bunlardan iki tanesi referans deęişken olarak atanmıştır. DMF analizlerinde dolayısıyla yapılan karşılaştırma sayısı 8'dir. Bonferroni düzeltmesi 8 karşılaştırma için uygulanırsa düzeltilmiş anlamlılık seviyesi ve bu anlamlılık düzeyi için kritik ki-kare deęeri aşıęıdaki gibi olmaktadır.

$$\alpha_{\text{düzeltilmiş}} = \frac{0,05}{8} = 0,006$$

$$p = 0,006 \text{ için } \chi^2(sd = 2) = 10,232$$

5.  $\chi^2(2) - \chi^2(1)$  farkı hesaplanan bu ki-kare deęerini geçtięi durumlarda madde DMF içeriyor şeklinde yorumlanmıştır.
6. Referans deęişkenler hariç tüm maddeler için (1-5) arası adımlar tekrar edilmiştir.

#### *DMF nin olası sebeplerinin incelenmesi*

Analiz sonucunda maddelerin gruplar arasında deęişen fonksiyon göstermesinin kaynaęının saptanması amacıyla üniversitelerin yabancı diller eğitimi ya da ölçme deęerlendirme bölümlerinde görev yapan anadili İngilizce veya Türkçe olan on akademisyenin görüşüne başvurulmuştur. Uzman görüşlerini almak için madde inceleme formu (Ek 24) kullanılmıştır. Allalouf, Hambleton ve Sireci (1999)'nin geliştirdięi madde inceleme formu bu çalışma için modifiye edilmiştir. Maddeleri incelemeden önce uzmanlara DMF hakkında özet bilgi verilmiş ve DMF'li maddelere örnekler gösterilmiştir. Tüm uzmanlar birbirlerinden baęımsız olarak maddeleri incelemiş ve madde inceleme formundaki soruları DMF gösteren her bir madde için yanıtlamışlardır. Uzmanlardan ayrıca yanıt kategorilerinin orijinal ve uyarlanan versiyonunu da incelemeleri istenmiştir.

Bu alıřmada, betimsel istatistiklerin hesaplanması ve DFA sayılılarının test edilmesinde SPSS 17, lme modeline iliřkin parametrelerin kestirilmesi, referans deęiřkenin belirlenmesi ve DMF gsteren maddelerin tespit edilmesinde PRELIS 2.71 , LISREL 8.71 programları kullanılmıřtır.

### 3. BULGULAR VE YORUMLAR

Bu bölüm alt problemlere ilişkin bulguları ve yorumları içermektedir. Araştırma sonucunda elde edilen bulgular çizelge ve açıklamalarla ifade edilmiş ve yorumlanmıştır.

PISA 2006 öğrenci anketinde yer alan, öğrencilerin fen bilimlerine yönelik tutumlarından birisi olan “bilimsel sorgulamaya verilen destek” tutumunun Avustralya, Yeni Zelanda, Amerika Birleşik Devletleri ve Türkiye kültürleri arasında eşdeğerliği ÇG-DFA ile incelenmiştir.

Bu incelemeler sırasında öncelikle, alt gruplara ilişkin madde istatistiklerine ardından ölçme modelinin alt gruplardaki uyumu ile ÇG-DFA sonuçlarına ve elde edilen sonuçlara göre DMF'nin olası sebeplerine yer verilmiştir.

#### 3.1 ALT PROBLEM 1'E İLİŞKİN BULGULAR VE YORUMLAR

*Alt Problem 1: PISA 2006 Öğrenci Anketi kapsamında ölçülen “Bilimsel Sorgulamaya Verilen Destek” ölçme modeli Avustralya, Yeni Zelanda, Amerika Birleşik Devletleri ve Türkiye kültürlerinde aynı faktör yapısına sahip midir?*

Bu alt problem kapsamında incelenen fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri faktörlerine ait betimsel istatistikler Çizelge 3.1 ve Çizelge 3.2'de verilmiştir. Bilimsel sorgulamaya verilen destek tutumunu oluşturan her iki faktörde de beşer madde yer almaktadır. Yorumlama kolaylığı açısından maddelerin kodlaması tersine çevrilmiştir, yüksek değerler daha olumlu tutum belirtmektedir.

Çizelge 3.1 Fen Bilimlerinin Genel Değeri Faktörünü Oluşturan Maddelere Ait Betimsel İstatistikler

	ST18Q01	ST18Q02	ST18Q04	ST18Q06	ST18Q09
Avustralya					
Ortalama	3,194	3,283	3,037	3,162	2,778
Standart sapma	0,645	0,612	0,633	0,647	0,739
Yeni Zelanda					
Ortalama	3,152	3,241	3,049	3,087	2,725
Standart sapma	0,633	0,612	0,624	0,668	0,730
ABD					
Ortalama	3,324	3,370	3,155	3,244	2,946
Standart sapma	0,652	0,619	0,669	0,656	0,746
Türkiye					
Ortalama	3,502	3,415	3,204	3,461	3,290
Standart sapma	0,616	0,620	0,734	0,636	0,685

*Not.* Maddeler 4 dereceli Likert formatındadır (1= hiç katılmıyorum, 2= katılmıyorum, 3= katılıyorum, 4= tümüyle katılıyorum).

Fen bilimlerine verilen genel değeri ölçen maddelerin aritmetik ortalamalarına bakıldığında, öğrencilerin genelde katılıyorum veya tümüyle katılıyorum şeklinde görüş bildirdikleri görülmektedir. Türk öğrenciler diğer öğrencilere göre fen bilimlerine genel olarak daha fazla değer verdiklerini bildirirken Avustralyalı ve Yeni Zelandalı öğrencilerin ortalamaları diğer ülke öğrencilerinininkinden daha düşüktür. Tüm ülkeler için ST18Q04 ve ST18Q09 kodlu maddelerin aritmetik ortalamalarının diğer maddelere göre biraz daha düşük olduğu görülmektedir.

Çizelge 3.2 Fen Bilimlerinin Kişisel Değeri Faktörünü Oluşturan Maddelere Ait Betimsel İstatistikler

	ST18Q03	ST18Q05	ST18Q07	ST18Q08	ST18Q10
Avustralya					
Ortalama	2,676	2,737	2,614	2,872	2,677
Standart sapma	0,720	0,823	0,852	0,767	0,877
Yeni Zelanda					
Ortalama	2,637	2,725	2,579	2,895	2,712
Standart sapma	0,702	0,824	0,852	0,765	0,881
ABD					
Ortalama	2,865	2,898	2,807	3,023	2,902
Standart sapma	0,732	0,807	0,817	0,748	0,809
Türkiye					
Ortalama	3,028	3,084	2,807	3,080	2,674
Standart sapma	0,727	0,772	0,867	0,750	0,888

*Not.* Maddeler 4 dereceli Likert formatındadır (1= hiç katılmıyorum, 2= katılmıyorum, 3= katılıyorum, 4= tümüyle katılıyorum).

Fen bilimlerine verilen kişisel değeri ölçen maddelerin aritmetik ortalamalarına bakıldığında, öğrencilerin genelde katılmıyorum veya katılıyorum şeklinde görüş bildirdikleri görülmektedir. Türk öğrenciler diğer öğrencilere göre fen bilimlerine kişisel olarak daha fazla değer verirken Avustralyalı ve Yeni Zelandalı öğrencilerin ortalamaları diğer ülke öğrencilerinininkinden daha düşüktür.

Çizelge 3.1 ve Çizelge 3.2 incelendiğinde öğrencilerin fen bilimlerine verdikleri genel değer ortalamalarının fen bilimlerine verilen kişisel değer ortalamalarından daha yüksek olduğu görülmektedir. Her iki alt boyuta ilişkin ortalamalar ve standart sapmalar incelendiğinde, öğrencilerin yanıtlarında benzer bir örüntü olduğu görülmektedir.

Kültürler arası çoklu grup uygulamalarından önce, “bilimsel sorgulamaya verilen destek” tutumu faktör yapısının her bir kültür içinde de geçerli olup olmadığını incelemek için model uyumu birleştirilmiş veri ve her bir ülke verisi için ayrı ayrı değerlendirilmiştir.

DFA sonucunda elde edilen faktör yükleri (Ek 15-18) 0,05 anlamlılık düzeyinde değerlendirildiğinde hepsinin anlamlı olduğu görülmüştür. Faktör yükleri Avustralya örnekleme için 0,68-0,93, Yeni Zelanda örnekleme için 0,62-0,91, ABD örnekleme için 0,72-0,89 ve Türkiye örnekleme için 0,65-0,83 arasında değerler almıştır. Ölçme modeline ait genel uyum indeks değerleri ise Çizelge 3.3'te özetlenmiştir.

Çizelge 3.3 Bilimsel Sorgulamaya Verilen Destek Ölçme Modeli Uyum İstatistikleri

	$\chi^2$	sd	RMSEA	CFI	NNFI	SRMR
Birleşik veri	2665,099*	34	0,057	0,950	0,934	0,079
Avustralya	1512,769*	34	0,061	0,960	0,947	0,090
Yeni Zelanda	475,228*	34	0,061	0,949	0,932	0,092
ABD	475,351*	34	0,054	0,946	0,929	0,078
Türkiye	469,880*	34	0,053	0,943	0,925	0,073

*Not.* sd = serbestlik derecesi, RMSEA= root mean square error of approximation , CFI= comparative fit index , NNFI=non-normed fit index , SRMR = standardized root mean square residual.

\* $p < 0,01$

Fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri alt boyutlarından oluşan iki faktörlü ölçme modelinin uyumu Çizelge 3.3'te verilen ki-kare ve indeks değerleri kullanılarak incelenmiştir. Ki-kare değerlerinin örneklem büyüklüğünden etkilendiği bilindiğinden model uyumuna ilişkin kararlar diğer indeks değerlerine bakılarak verilmiştir. RMSEA ve SRMR değerleri sırasıyla 0,08 ve 0,10'un altında ve CFI ile NNFI değerleri de 0,90'in üstünde olduğundan iki faktörlü ölçme modeline ait indeks değerlerinin hepsinin kabul edilebilir uyum kriterlerini sağladığı görülmüştür. Doğrulayıcı faktör analizi sonuçları ölçme modelinin her bir ülke verisinde aynı sayıda faktöre ve maddeye sahip olduğunu ve faktörlerle maddeler arası ilişkilerinde benzer olduğunu ortaya koymuştur.

Ölçme modelinin her bir ülkedeki uyumunun değerlendirilmesinde modeldeki faktörlerin güvenilirlikleri ve faktörler arası ilişkiler de incelenmiştir. Bilimsel sorgulamaya verilen destek ölçme modeline ait fen bilimlerinin genel değeri ve fen

bilimlerinin kişisel değeri faktörlerinin her bir kültürdeki güvenilirlikleri ve aralarındaki korelasyonlar Çizelge 3.4'te verilmiştir.

Çizelge 3.4 Fen Bilimlerinin Genel Değeri ve Fen Bilimlerinin Kişisel Değeri Faktörlerinin Güvenirlikleri

	Yapı güvenirligi		Açıklanan varyans		Faktörler arası korelasyon
	GD	KD	GD	KD	
Avustralya	0,897	0,926	0,637	0,717	0,856
Yeni Zelanda	0,872	0,909	0,579	0,670	0,854
ABD	0,902	0,905	0,650	0,656	0,869
Türkiye	0,881	0,880	0,598	0,597	0,827

*Not.* GD = fen bilimlerinin genel değeri, KD= fen bilimlerinin kişisel değeri.

Çizelge 3.4'teki yapı güvenirligi değerleri iç tutarlılığı, açıklanan varyans değerleri ise maddelerde ilişkili oldukları faktörler tarafından açıklanan toplam varyansı ifade etmektedir.

Yapı güvenirligi değerleri fen bilimlerinin genel değeri faktörü için 0,872 ile 0,902 arasında değerler alırken fen bilimlerinin kişisel değeri faktörü içinse 0,880 ile 0,926 arasında değişmektedir. Açıklanan varyans oranları fen bilimlerinin genel değeri faktörü için 0,579-0,650, fen bilimlerinin kişisel değeri faktörü için 0,597-0,717 ranjındadır. Avustralya ve Yeni Zelanda örneklemlerinde fen bilimlerinin kişisel değeri faktörünün güvenirliginin fen bilimlerinin genel değeri faktörünün güvenirliginden daha yüksek olduğu tespit edilmiştir. ABD ve Türkiye için ise her iki faktörün güvenirlikleri yaklaşık olarak aynıdır.

Fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri faktörlerinin yapı güvenirliklerinin, tüm ülkeler için, önerilen seviye olan 0,70 den yüksek olduğu görülmektedir. Diğer bir güvenirlilik göstergesi olan faktörlere ait açıklanan varyans oranlarının da tavsiye edilen (Hair, Anderson, Tahtam ve Black, 1998) 0,50 sınırının üzerinde olduğu tespit edilmiştir.

Faktörler arası korelasyonlar incelendiğinde, fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri faktörlerinin tüm ülkelerde oldukça ilişkili olduğu gözlenmiştir. Faktörler arası korelasyonlar Avustralya ve Yeni Zelanda örneklemelerinde birbirine oldukça yakın değerler alırken ABD ve Türkiye örneklemelerinde farklılık göstermektedir. Türkiye örneğinde faktörler arası korelasyon diğer ülkelere göre biraz daha düşüktür.

Her bir maddede ilişkili olduğu faktör tarafından açıklanan varyans miktarı diğer bir ifadeyle madde güvenilirliği de bir başka model uyum göstergesidir (Schulz, 2000). Bu değerler LISREL çıktısındaki her bir maddeye ait ( $R^2$ ) değerleridir. Fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri faktörlerine ait maddelerin güvenilirlikleri sırasıyla Çizelge 3.5 ve Çizelge 3.6'da verilmiştir.

Çizelge 3.5 Fen Bilimlerinin Genel Değeri Faktörünü Oluşturan Maddelerin Güvenirlikleri

	ST18Q01	ST18Q02	ST18Q04	ST18Q06	ST18Q09
Avustralya	0,595	0,723	0,549	0,782	0,536
Yeni Zelanda	0,557	0,693	0,498	0,700	0,448
ABD	0,600	0,729	0,581	0,801	0,540
Türkiye	0,642	0,684	0,422	0,647	0,593

Çizelge 3.5'e göre fen bilimlerinin genel değeri faktörünün her bir maddedeki varyansın genelde %50'den fazlasını açıkladığı görülmektedir. ST18Q06 kodlu maddenin bu faktörün en iyi temsilcisi olduğu sonucuna varılabilir. ST18Q04 ve ST18Q09 kodlu maddelerde açıklanan varyans oranları %50'nin altına inmektedir.

Çizelge 3.6 Fen Bilimlerinin Kişisel Değeri Faktörünü Oluşturan Maddelerin Güvenirlikleri

	ST18Q03	ST18Q05	ST18Q07	ST18Q08	ST18Q10
Avustralya	0,458	0,807	0,864	0,820	0,737
Yeni Zelanda	0,389	0,780	0,822	0,680	0,680
ABD	0,513	0,695	0,742	0,695	0,636
Türkiye	0,487	0,565	0,674	0,660	0,507

Çizelge 3.6'ya göre fen bilimlerinin kişisel değeri faktörünün, her bir maddedeki varyansın genelde %50'den fazlasını açıkladığı görülmektedir. ST18Q07 kodlu maddenin bu faktörün en iyi temsilcisi olduğu sonucuna varılabilir. ST18Q03 kodlu madde de açıklanan varyans oranları %50'nin altına inmektedir. Diğer maddelerle karşılaştırıldığında bu maddenin güvenilirlik değerlerinin oldukça düşük olması maddenin ifade tarzında bir problem olabileceğini veya bu faktörün iyi bir temsilcisi olmadığını düşündürmektedir.

Betimsel istatistikler, global model uyum indeksleri, faktör ve madde güvenilirlikleri incelendiğinde ölçme modelinin her bir kültürde iki faktörlü olduğu ve maddelerin faktörlerle aynı örüntüye sahip olduğu sonucuna varılmıştır. Bununla birlikte, iki faktörlü ölçme modelinin tüm gruplarda oldukça iyi uyum vermesi, model parametrelerinin gruplar arasında eşdeğer olacağı anlamına gelmemelidir.

### 3.2 ALT PROBLEM 2A'YA İLİŞKİN BULGULAR VE YORUMLAR

*Alt Problem 2a: PISA 2006 Öğrenci Anketi kapsamında ölçülen “Bilimsel Sorgulamaya Verilen Destek” ölçme modelinin orijinal ve uyarlanan maddeleri Avustralya ve Yeni Zelanda arasında DMF göstermekte midir?*

Çizelge 3.7 Bilimsel Sorgulamaya Verilen Destek Ölçme Modeli Maddeleri için Avustralya ve Yeni Zelanda Örneklemelerinde Yapılan DMF Analizi Sonuçları

	$\chi^2$	$\Delta\chi^2$	$\Delta sd$	p	DMF
Serbest temel model	1987,998	-	-	-	-
GD					
ST18Q02	1988,229	0,231	2	0,891	Yok
ST18Q04	1991,048	3,050	2	0,218	Yok
ST18Q06	1989,148	1,150	2	0,563	Yok
ST18Q09	1991,181	3,183	2	0.204	Yok
KD					
ST18Q03	1993,574	5,576	2	0,062	Yok
ST18Q05	1989,523	1,525	2	0,466	Yok
ST18Q07	1989,085	1,087	2	0,581	Yok
ST18Q10	1988,490	0,492	2	0,782	Yok

*Not.* sd = serbestlik derecesi, DMF= değişen madde fonksiyonu, GD = fen bilimlerinin genel değeri, KD= fen bilimlerinin kişisel değeri.

Avustralya ve Yeni Zelanda örneklemeleri üzerinde yapılan bilimsel sorgulamaya verilen destek ölçme modeli maddelerine ilişkin DMF analizi sonuçları Çizelge 3.7’de verilmiştir. ST18Q01 ve ST18Q08 kodlu maddeler referans değişkeni olarak kullanıldığından bu çizelgede yer almamaktadır. Fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri maddelerine ait ki-kare farklılıkları Bonferroni düzeltmesi uygulanmış anlamlılık düzeyinde değerlendirildiğinde hiçbir maddenin DMF içermediği sonucuna varılmıştır. ST18Q03, ST18Q04 ve ST18Q09 kodlu maddelerin ki-kare farklılık değerlerinin diğer maddelerinkinden biraz daha yüksek olduğu ortaya çıkmıştır. Avustralya ve Yeni Zelanda örneklemeleri arasında hiçbir madde DMF göstermemiştir.

Avustralya ve Yeni Zelanda’da aynı dil kullanıldığı ve benzer kültürel özelliklere sahip olduğu için maddelerin bu ülkelerdeki öğrenciler tarafından aynı şekilde anlaşıldığı ve DMF göstermediği düşünülmektedir.

### 3.3 ALT PROBLEM 2B’YE İLİŞKİN BULGULAR VE YORUMLAR

*Alt Problem 2b: PISA 2006 Öğrenci Anketi kapsamında ölçülen “Bilimsel Sorgulamaya Verilen Destek” ölçme modelinin orijinal ve uyarlanan maddeleri Avustralya ve ABD arasında DMF göstermekte midir?*

Çizelge 3.8 Bilimsel Sorgulamaya Verilen Destek Ölçme Modeli Maddeleri için Avustralya ve ABD Örneklerinde Yapılan DMF Analizi Sonuçları

	$\chi^2$	$\Delta\chi^2$	$\Delta sd$	p	DMF
Serbest temel model	1988,120	-	-	-	-
GD					
ST18Q01	1990,034	1,914	2	0,384	Yok
ST18Q04	1990,287	2,167	2	0,339	Yok
ST18Q06	1988,643	0,523	2	0,770	Yok
ST18Q09	1994,381	6,261	2	0,044	Yok
KD					
ST18Q03	2016,864	28,744	2	0,000	Var
ST18Q07	1988,147	0,027	2	0,987	Yok
ST18Q08	1999,802	11,682	2	0,003	Var
ST18Q10	1989,499	1,379	2	0,502	Yok

*Not.* sd = serbestlik derecesi, DMF= değişen madde fonksiyonu, GD = fen bilimlerinin genel değeri, KD= fen bilimlerinin kişisel değeri.

Avustralya ve ABD örnekleri üzerinde yapılan bilimsel sorgulamaya verilen destek ölçme modeli maddelerine ilişkin DMF analizi sonuçları Çizelge 3.8’de verilmiştir. ST18Q02 ve ST18Q05 kodlu maddeler referans değişkeni olarak kullanıldığından bu çizelgede yer almamaktadır. Fen bilimlerinin genel değeri ve fen bilimlerinin kişisel

değeri maddelerine ait ki-kare farklılıkları Bonferroni düzeltmesi uygulanmış anlamlılık düzeyinde değerlendirildiğinde ST18Q03 ve ST18Q08 kodlu maddelerin DMF içerdiği sonucuna varılmıştır. Özellikle ST18Q03 kodlu maddeye ilişkin ki-kare farklılık değeri diğer maddelerinkine göre oldukça yüksektir. Avustralya ve ABD örneklemeleri arasında iki madde DMF göstermiştir. Bu ülkelerde aynı dil kullanılmasına karşın kültürel açıdan farklılıklar olduğundan dolayı bu iki maddenin Avustralya ve ABD'deki öğrenciler tarafından aynı şekilde algılanmadığı ve dolayısıyla DMF gösterdiği düşünülmektedir.

### 3.4 ALT PROBLEM 2C'YE İLİŞKİN BULGULAR VE YORUMLAR

*Alt Problem 2c: PISA 2006 Öğrenci Anketi kapsamında ölçülen "Bilimsel Sorgulamaya Verilen Destek" ölçme modelinin orijinal ve uyarlanan maddeleri Avustralya ve Türkiye arasında DMF göstermekte midir?*

Çizelge 3.9 Bilimsel Sorgulamaya Verilen Destek Ölçme Modeli Maddeleri için Avustralya ve Türkiye Örneklemelerinde Yapılan DMF Analizi Sonuçları

	$\chi^2$	$\Delta\chi^2$	$\Delta sd$	p	DMF
Serbest temel model	1982,649	-	-	-	-
GD					
ST18Q01	2022,892	40,243	2	0,000	Var
ST18Q02	2035,419	52,770	2	0,000	Var
ST18Q04	2014,059	31,410	2	0,000	Var
ST18Q09	2060,827	78,178	2	0,000	Var
KD					
ST18Q03	2074,006	91,357	2	0,000	Var
ST18Q05	1999,324	16,675	2	0,000	Var
ST18Q08	2007,537	24,888	2	0,000	Var
ST18Q10	2014,681	32,032	2	0,000	Var

*Not.* sd = serbestlik derecesi, DMF= değişen madde fonksiyonu, GD = fen bilimlerinin genel değeri, KD= fen bilimlerinin kişisel değeri.

Avustralya ve Türkiye örneklemi üzerinde yapılan bilimsel sorgulamaya verilen destek ölçme modeli maddelerine ilişkin DMF analizi sonuçları Çizelge 3.9’da verilmiştir. ST18Q06 ve ST18Q07 kodlu maddeler referans değişkeni olarak kullanıldığından bu çizelgede yer almamaktadır. Fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri maddelerine ait ki-kare farklılıkları Bonferroni düzeltmesi uygulanmış anlamlılık düzeyinde değerlendirildiğinde referans değişkenleri hariç tüm maddelerin DMF içerdiği sonucuna varılmıştır. Özellikle ST18Q03 kodlu maddeye ilişkin ki-kare farklılık değeri diğer maddelerinkine göre daha yüksektir. Avustralya ve Türkiye örneklemi arasında ST18Q07 kodlu madde hariç diğer tüm maddeler DMF göstermiştir. Bu ülkelerde hem kullanılan dil hem de kültürel açıdan büyük farklılıklar olduğundan dolayı bu maddelerin Avustralya ve Türkiye’deki öğrenciler için aynı anlamı vermediği ve dolayısıyla DMF gösterdiği düşünülmektedir.

### 3.5 ALT PROBLEM 3’E İLİŞKİN BULGULAR VE YORUMLAR

*Alt Problem 3: Maddelerin kültürler ve diller arasında farklı fonksiyon göstermesinin olası nedenleri nelerdir?*

Avustralya-ABD ve Avustralya-Türkiye arasında sırasıyla 2 ve 9 maddenin DMF gösterdiği tespit edilmiştir. Avustralya-Yeni Zelanda karşılaştırmasında ise DMF içeren madde bulunmamaktadır.

#### *Avustralya-Türkiye karşılaştırması*

Bilimsel sorgulamaya verilen destek tutumunu ölçen maddelerin yer aldığı soru kökü İngilizce formunda “How much do you agree with the statements below? (Please tick one box in each row)” ve yanıt kategorileri “1= strongly agree, 2= agree, 3= disagree, 4= strongly disagree” şeklindedir. Türkçe formunda ise soru kökü “Aşağıdaki ifadelerle ne ölçüde katılıyorsunuz? (Her sırada sadece bir kutuyu işaretleyiniz)” ve yanıt kategorileri “1= tümüyle katılıyorum, 2= katılıyorum, 3= katılmıyorum, 4= hiç katılmıyorum” şeklindedir.

Uzmanların yedisi soru ifadesinin ve yanıt kategorilerinin uyarlanmasında herhangi anlam kaybı olmadığı ve çevirisi kalitesinin de çok iyi olduğu konusunda ortak görüş

bildirmişlerdir. Uzmanlardan iki tanesi ise çevirinin çok zayıf olduğu bunun da anlamda değişikliğe yol açtığını belirtmiştir. Uzmanlara göre “tümüyle” kelimesinin İngilizce karşılığının “strongly” değil, “totally” olması gerekmektedir. Bir uzman ise “strongly agree” ifadesi “tümüyle katılıyorum” şeklinde çevrildiğinden dolayı “strongly disagree” ifadesinin de “tümüyle katılmıyorum” biçiminde çevrilmesi gerektiğini belirtmiştir.

ST18Q01 kodlu madde İngilizce formunda “Advances in science and technology usually improve people’s living conditions.”, Türkçe formunda ise “Fen bilimleri ve teknolojideki ilerlemeler genellikle insanların yaşam koşullarını iyileştirir.” şeklindedir. Bu maddenin uyarlanmasıyla ilgili olarak uzmanların yedisi çeviri kalitesinin iyi olduğu ve anlam kaybı olmadığını belirtirken, üç uzman ise çeviride problemler olduğunu ifade etmiştir. Uzmanlardan iki tanesi bilimde “advance” kelimesinin “ilerleme” yerine “gelişme” olarak kullanıldığını belirtirken, diğer uzman ise “fen bilimleri” teriminin “science ” terimine denk olmadığını belirtmiştir. Maddenin niçin DMF gösterdiğine dair ortak bir kanıya ulaşılamamıştır.

ST18Q02 kodlu madde İngilizce formunda “Science is important for helping us to understand the natural world.”, Türkçe formunda ise “Fen bilimleri, doğal dünyayı anlamamıza yardımcı olması açısından önemlidir.” şeklindedir. Bu maddedeki “natural world” ifadesinin Türkçe’ye “doğal dünya” olarak uyarlanmasının yanlış olduğu konusunda uzmanlar ortak görüş bildirmişlerdir. Uzmanlara göre bu ifadenin “doğa” veya “tabiat” olarak çevrilmesi gerekmektedir. Çevirideki hatadan dolayı maddenin DMF göstermiş olabileceği bildirilmiştir.

ST18Q03 kodlu madde İngilizce formunda “Some concepts in science help me see how I relate to other people.”, Türkçe formunda ise “Fen bilimlerindeki bazı kavramlar, diğer insanlarla nasıl bir bağlantı içinde olduğumu görmeme yardımcı olur.” şeklindedir. Bu maddeyle ilgili olarak tüm uzmanlar çevirinin çok zayıf olduğunu ve anlam değişikliğinin de fazla olduğunu ifade etmişlerdir. Ayrıca uzmanların yarısı bu maddenin hem Türkçe hem de İngilizce ifadesinin çok anlaşılır olmadığını belirtmiştir. Uzmanların hepsi “relate to” ifadesinin “bağlantı içinde olmak” anlamına gelmediğini belirtmiştir. Bununla birlikte, Türkçe çevirisindeki “bağlantı” dan ne kastedildiğinin de

belli olmadığı, bunun sosyal mi yoksa fiziki yoksa biyolojik bir bağlantı mı olduğunun anlaşılacağı dile getirilmiştir. Bazı uzmanlara göre “relate to” ifadesi “alaka” bazılarına göre ise “etkileşim” anlamında kullanılmıştır. Bununla birlikte, bu maddedeki “some concepts (bazı kavramlar)” ifadesinin belirsiz bir anlam içermesi, uzmanlar tarafından bu maddenin DMF göstermesinin kaynağı olabileceği şeklinde yorumlanmıştır.

ST18Q04 kodlu madde İngilizce formunda “Advances in science and technology usually help improve the economy.”, Türkçe formunda ise “Fen bilimleri ve teknolojiye ilerlemeler genellikle ekonominin iyileştirilmesine yardımcı olur.” şeklindedir. Uzmanlardan ikisi bu maddedeki “iyileştirilmesine” kelimesinin “iyileşmesine” şeklinde çevrilmesi gerektiğini ifade etmişlerdir. Bir başka uzman ise “fen bilimlerinde” ilerleme olmayacağını “bilimde” ilerleme olacağını belirtmiştir. Bu maddenin niçin değişen fonksiyon gösterdiğine ilişkin ortak bir kanıya varılamamıştır.

ST18Q05 kodlu madde İngilizce formunda “I will use science in many ways when I am an adult.”, Türkçe formunda ise “Bir yetişkin olduğumda fen bilimlerinden birçok yönde yararlanacağım.” şeklindedir. Bu maddenin çevirisinin çok zayıf olduğu ve bu maddedeki uyarlamadan kaynaklı anlam değişikliğinin fazla olduğu konusunda ortak görüş bildirilmiştir. Uzmanlar cümlenin İngilizce’sinin basit ve net olduğunu Türkçe ifadesinin anlaşılmasının daha zor olduğunu dile getirmişlerdir. Uzmanlara göre bu maddedeki “use” kelimesi “kullanmak” anlamındayken Türkçe’ye “yararlanmak” şeklinde çevrilmiştir. Uzmanların bu maddeyle ilişkili olarak değindikleri diğer bir nokta ise “adult” ve “yetişkin” terimlerinin her iki dilde aynı zorlukta olmadığıdır. “Adult” kelimesi İngilizce’de günlük dilde daha sık kullanıldığından, 15 yaş grubu öğrencileri için, bu kelimenin Avustralyalı ve Türk ve öğrenciler tarafından farklı şekilde algılanabileceği bazı uzmanlar tarafından vurgulanmıştır. Çevirinin yetersiz olmasının ve kelimelerin her iki dilde farklı zorlukta olmasının bu maddenin farklı fonksiyon göstermesinin nedenleri olabileceği belirtilmiştir.

ST18Q06 kodlu madde İngilizce formunda “Science is valuable to society.”, Türkçe formunda ise “Fen bilimleri toplum için yararlıdır.” şeklindedir. Avustralya-Türkiye

karşılaştırmasında bu madde referans değişken olarak kullanılmasına rağmen, referans değişkeni belirleme analizlerinde bu maddenin de DMF’li olduğu tespit edilmiştir. Uzmanların çoğu “valuable” kelimesinin bu maddedeki Türkçe karşılığının “yararlı” değil “değerli” olduğu şeklinde ortak görüş bildirmişlerdir. Birkaç uzman ise çeviride herhangi bir problem olmadığını belirtmiştir.

ST18Q08 kodlu madde İngilizce formunda “I find that science helps me to understand the things around me.”, Türkçe formunda ise “Fen bilimlerinin, çevremdeki şeyleri anlamada bana yardımcı olduğu görüşümdedir.” şeklindedir. Uzmanlara göre bu maddedeki problemlerden biri “I find that” ifadesiyle ilgilidir. “I find that” ifadesi bu maddede bazı uzmanlara göre “olduğunu düşünüyorum”, bazılarına göre “farkındayım” veya “idrakindeyim” diğer bir kısmına göre ise “görüyorum” şeklinde Türkçe’ye çevrilmelidir. Bazı uzmanlar bu maddenin İngilizce ifadesindeki “I find that” ifadesinin gereksiz olduğunu ve bu ifadenin cümleden çıkarıldığında herhangi bir anlam kaybı olmayacağını dile getirmiştir. Bu uzmanlara göre, bu ifade cümleyi Avustralyalı ve Türk öğrenciler için farklı ölçüde karmaşık hale getirmektedir. Bu maddeyle ilgili uzmanların ortak görüş bildirdiği diğer bir problem ise maddenin orijinal halindeki “things around me” ifadesidir. Uzmanlara göre “çevremdeki şeyler” ifadesi farklı kültürdeki kişiler için farklı anlamlar taşıyabilir. Çevirideki sıkıntılar ve kelimelerin kültürlerde farklı anlamlar içermesinden dolayı bu maddedeki anlam değişikliğinin fazla olduğu uzmanlar tarafından belirtilmiştir.

ST18Q09 kodlu madde İngilizce formunda “Advances in science and technology usually bring social benefits.”, Türkçe formunda ise “Fen bilimleri ve teknolojiye ilerlemeler genellikle toplumsal yararlar sağlar.” şeklindedir. Bu maddedeki “social benefits” kavramının kültürel açıdan farklı anlamlar içerebileceği uzmanların çoğu tarafından belirtilmiştir. Başka bir uzmana göre ise bu maddenin çevirisi “Fen bilimleri ve teknolojiye ilerlemelerin genellikle toplumsal yararları olur.” şeklinde olmalıdır.

ST18Q10 kodlu madde İngilizce formunda “When I leave school there will be many opportunities for me to use science.”, Türkçe formunda ise “Okulu bitirdikten sonra, fen bilimlerinden yararlanmam için pek çok fırsat olacaktır.” şeklindedir. Uzmanlar bu

maddede yer alan “use” kelimesinin “yararlanma” anlamındaki çevirisinin yanlış olduğu konusunda ortak görüş bildirmişlerdir. Doğru çevirinin “kullanma” anlamında olması gerekmektedir. Ayrıca iki uzmana göre, “when I leave school” ifadesi “okulu bitirdikten sonra” değil “okulu bitirince” şeklinde Türkçe’ye çevrilmelidir.

Uzmanlar ST18Q01 ve ST18Q04 kodlu maddelerin Avustralya ve Türkiye örneklemeleri arasında niçin DMF gösterdiği konusunda uzmanlar ortak bir kanıya ulaşamamıştır. Diğer maddeler için çeviri ve uyarlama problemleri uzmanlar tarafından maddelerin Avustralya ile Türkiye arasında farklı fonksiyon göstermesinin asıl kaynağı olarak gösterilmiştir. Avustralya ve Türkiye arasında maddelerin farklı fonksiyon göstermesinin diğer nedenleri ise, orijinal maddenin anlamındaki belirsizlik, maddelerde belirsizlik bildiren ifadelerin veya sözcüklerin kullanılması ve bazı sözcüklerin kültürlerde değişen sıklıkta kullanılmasından dolayı maddelerin farklı kültürlerdeki öğrenciler tarafından farklı şekilde algılanabileceği şeklinde yorumlanmıştır. Benzer sonuçlar Allalouf, Hambleton ve Sireci (1999), Elosua ve Jauregúi (2007), Ercikan (1998), Ercikan ve diğerleri (2004), Schulz (2003), Van de Vijver ve Tanzer (2004) tarafından da rapor edilmiştir.

#### *Avustralya-ABD karşılaştırması*

Uzmanlara göre, ST18Q03 kodlu maddedeki “some concepts” ve ST18Q08 kodlu maddedeki “things around me” ifadelerinin belirsiz anlamlar içermesi bu maddelerin Avustralya ve Amerikan kültürlerinde farklı fonksiyon göstermesinin nedeni olabilir. Amerikalı bir uzmana göre örneğin “things around me” ifadesi Amerika’da somut şeyler için kullanılmaktadır. Türk kültüründe ise “şey” kelimesi hem somut hem de soyut anlamlar içermektedir. Benzer şekilde bu ifade Avustralya kültüründe de soyut anlamlar içerecek şekilde kullanılıyor olabilir. Aynı uzmana göre ST18Q03 maddesindeki “some concepts” ifadesi yerine daha spesifik bir deyişin kullanılması uygun olabilir.

Maddelerin Avustralya-Türkiye ve Avustralya-ABD arasında DMF göstermesinin olası nedenlerine yönelik uzman görüşleri Çizelge 3.10’da özetlenmiştir.

Çizelge 3.10 Maddelerin DMF Göstermesinin Olası Nedenlerine İlişkin Uzman Görüşleri

		Yanıt Kategorileri	ST18Q01	ST18Q02	ST18Q03	ST18Q04	ST18Q05	ST18Q06	ST18Q07	ST18Q08	ST18Q09	ST18Q10
AUS-TUR	Çeviri/Uyarlama Problemleri	3	3	9	10	3	8	7	-	8	4	8
	Kültürel Farklılıklar	0	0	0	6	0	3	0	-	7	6	0
AUS-ABD	Kültürel Farklılıklar	-	-	-	7	-	-	-	-	7	-	-

Çizelgeden de görülebileceği gibi, uzmanlara göre maddelerin DMF göstermesinin nedenleri çeviri ve uyarlama problemleriyle kültürel farklılıklardır. Avustralya-Türkiye karşılaştırması için DMF gösteren maddelerin tamamında çeviri problemleri olduğu görülmektedir. Özellikle ST18Q03 kodlu madde için uzmanların 10'u da maddenin uyarlanmasında problem olduğunu belirtmiştir. Ayrıca bu karşılaştırma için dört maddede kültürel farklılıklardan kaynaklı DMF olduğu uzmanlar tarafından belirtilmiştir. Avustralya-ABD karşılaştırması içinse her iki soru için de yedi uzman kültürel farklılıkların DMF'ye neden olabileceğini belirtmiştir. Elde edilen bu sonuçlar literatürle de paralellik göstermektedir. Ölçme araçlarının kültürler arası eşdeğerliğiyle ilgili çalışmalar incelendiğinde, maddelerin kültürler ve diller arası DMF göstermesinin olası nedenleri arasında çeviri ve uyarlama problemleri (Allalouf, Hambleton ve Sireci, 1999; Ercikan, 1998, 2002; Ercikan ve diğerleri, 2004; Grisay ve diğerleri, 2007; Grisay ve Monseur, 2007; Hambleton, 1994; Sireci ve Allalouf, 2003; Yıldırım, 2006; Yıldırım ve Berberoğlu, 2009; Van de Vijver ve Tanzer, 2004) ve kültürel farklılıklar (Allalouf, Hambleton ve Sireci, 1999; Elousa ve Jauregui, 2007; Gierl ve Khaliq, 2001; Grisay 2003; Grisay ve Monseur, 2007; Van de Vijver ve Tanzer, 2004; Walker, 2007) en önemli sebepler arasında yer almaktadır.

## 4. SONUÇ VE ÖNERİLER

Bu bölümde, her bir alt probleme ilişkin sonuçlar özetlenmiştir. Ayrıca analizlerden elde edilen bulgular ışığında, yeni çalışmalara yönelik önerilere de yer verilmiştir.

### 4.1 SONUÇLAR

#### 4.1.1 Alt Problem 1'e İlişkin Sonuçlar

Öğrencilerin fen bilimlerine yönelik tutumlarından birisi olan bilimsel sorgulamaya verilen destek boyutu PISA uygulamalarında, fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri faktörleriyle ölçülmüştür. İki faktörlü ölçme modelinin betimsel istatistikleri incelendiğinde tüm öğrencilerin fen bilimlerine genel ve kişisel olarak değer verdikleri belirlenmiştir. Tüm ülke öğrencilerinin yanıtlarında benzer bir örüntü olduğu gözlenmiştir.

Ölçme modelinin genel uyumunun incelenmesinde, RMSEA, CFI, NNFI ve SRMR indeks değerlerinin birleştirilmiş veri ve tüm ülke verileri için kabul edilebilir uyum kriterlerini sağladığı tespit edilmiştir. Dolayısıyla iki faktörlü ölçme modelinin faktör yapısının her bir örnekleme de geçerli olduğu kabul edilmiştir. Fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değerine ilişkin güvenilirlik değerleri ve açıklanan varyans oranları da oldukça yüksek çıkmıştır. Avustralya, Yeni Zelanda, ABD ve Türkiye örneklemlerinde her iki faktör arasındaki korelasyon değerleri incelendiğinde, faktörler arası ilişkinin de benzer olduğu görülmüştür. Her bir maddede ilişkili oldukları faktörler tarafından açıklanan varyans miktarları değerlendirildiğinde, her bir maddedeki varyansın genelde %50'den fazlasının açıklandığı ortaya konulmuştur. Uluslararası çalışmalarda, her bir ülke verisinde aynı faktör yapısının ölçüldüğünün varsayılması doğru değildir. Analizler sonucunda ölçme modelinin her bir kültürde iki faktörlü olduğu ve maddelerin faktörlerle benzer örüntüye sahip olduğu sonucuna varılmıştır.

#### 4.1.2 Alt Problem 2'ye İlişkin Sonuçlar

Bilimsel sorgulamaya verilen destek ölçme modeline ait maddeler için Avustralya ve Yeni Zelanda örneklemeleri arasında yapılan analizler sonucunda hiçbir maddenin DMF içermediği sonucuna varılmıştır. Avustralya ve ABD örneklemeleri karşılaştırmasında fen bilimlerinin kişisel değeri faktörüne ait ST18Q03 ve ST18Q08 kodlu iki maddenin bu kültürler arasında eşdeğer olmadığı belirlenmiştir. Fen bilimlerinin genel değerini ölçen maddelerde DMF tespit edilmemiştir. Avustralya ve Türkiye karşılaştırmasında ise ST18Q07 kodlu madde hariç diğer tüm maddelerin bu kültürler arasında farklı fonksiyon gösterdiği görülmüştür. Maddelere ait ki-kare farklılık değerlerinin diğer bir ifadeyle DMF miktarlarının, diğer karşılaştırmalara oranla oldukça yüksek olduğu tespit edilmiştir. Bu çalışmadaki DMF analizlerinin ortaya koyduğu en önemli sonuç, dil ve kültür farklılıklarının maddelerin DMF göstermesinde oldukça etkili olabileceğidir. DMF analizleri bu farklılıklara göre incelendiğinde Çizelge 4.1'de verilen sonuçlara ulaşılmıştır.

Çizelge 4.1 Kültür ve Dil Farklılıklarına Göre DMF Gösteren Maddelerin Dağılımı

	Madde sayısı	DMF'li madde sayısı	% DMF
Aynı dil – Benzer kültür (AUS-NZL)	10	0	0
Aynı dil – Farklı kültür (AUS-ABD)	10	2	20
Farklı dil – Farklı kültür (AUS-TUR)	10	9	90

Çizelge 4.1'den de görüldüğü gibi kültürel ve dilsel farklılıklar arttıkça değişen fonksiyon gösteren maddelerin sayısı da artmaktadır. Ayrıca Türkiye ile Avustralya arasındaki kültürel farklılığın ABD ve Avustralya arasındaki kültürel farklılıktan daha fazla olduğu da yorumlamalarda göz önünde bulundurulmalıdır. Elde edilen bulgular benzer çalışmalar tarafından da desteklenmektedir (Ercikan ve Koh, 2005; Grisay ve diğerleri, 2007; Grisay ve Monseur 2007). Uluslararası testlerdeki çeviri hataları maddelerin DMF göstermesinin en önemli nedenidir ve kültürel farklılıklarla eğitim programlarındaki farklılıklardan kaynaklı DMF'ye göre daha sık karşılaşılır (Grisay, 2003).

Özetle, DFA sonuçları ölçme modelinin faktör yapısının tüm kültürlerde benzer yapıda olduğunu desteklemiştir. Diğer yandan ortalama ve kovaryans modeliyle yapılan ÇG-DFA sonuçları ise Avustralya-ABD ve Avustralya-Türkiye karşılaştırmalarında bazı maddelerin eşdeğer olmadığı yani DMF gösterdiğini ortaya koymuştur. Bilimsel sorgulamaya verilen destek tutumuna ilişkin, ülkeler veya kültürler arası karşılaştırmaların sadece Avustralya ve Yeni Zelanda arasında yapılabileceği sonucuna varılmıştır. Avustralya-Türkiye ve Avustralya-ABD karşılaştırmalarının tüm ölçek bazında yapılması ise anlamlı olmayacaktır. ITC Test Uyarlama Yönergelerinde (ITC, 2001) de, farklı gruplar üzerindeki karşılaştırmaların ancak ölçme değişmezliğinin sağlandığı ölçekler için yapılabileceği belirtilmektedir (Yönerge 21).

#### 4.1.3 Alt Problem 3'e İlişkin Sonuçlar

Maddelerin kültürler ve diller arasında farklı fonksiyon göstermesinin olası nedenlerinin neler olduğu konusunda uzman görüşlerine başvurulmuştur. Van de Vijver ve Hambleton (1996) testlerin çevirisinin ve uyarlamasının yapılması sürecinde DMF nin olası sebepleri hakkında bilgi sahibi olan psikologlardan ve dilbilimcilerden yararlanılmasını önermişlerdir.

Avustralya-Türkiye karşılaştırmasında, maddelerin DMF göstermesinin asıl nedeninin çeviri ve uyarlama yanlışlıkları olduğu belirlenmiştir. Ayrıca kültürel farklılıklardan dolayı, bazı maddelerin bu ülke öğrencileri tarafından farklı şekilde algılanmış olabileceği sonucuna varılmıştır. Belirsiz anlam içeren ifadelerin kullanılmasından dolayı iki maddenin Avustralyalı ve ABD'li öğrenciler için aynı anlamı ifade etmediği ve DMF gösterdiği tespit edilmiştir. Kültürel ve dilsel farklılıklar arttıkça ölçme değişmezliği sağlamayan maddelerin sayısının ve içerdikleri DMF miktarının arttığı gözlenmiştir. Grisay (2003)'e göre, kelime ve cümlelerin ortalama uzunlukları diller arasında farklılık gösteren bir özelliktir ve ölçme araçlarının uyarlanması aşamasında çevirmenler tarafından tamamen kontrol edilemez. Bu bakımdan, yazılı materyal içeren ölçme araçlarının tamamen eşdeğer versiyonlarının geliştirilmesi imkansız bir görevdir.

## 4.2 SINIRLILIKLAR

Bu çalışma, Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) tarafından düzenlenen Uluslararası Öğrenci Değerlendirme Programı (PISA) 2006 tarama araştırmasına ait verilerle sınırlıdır. Araştırmada kullanılan öğrenci tutumlarına ait veriler tamamen öğrencilerin kendi beyanlarına dayanmaktadır. Araştırma kapsamında görüşlerine başvurulmuş uzmanların tamamı maddelerin DMF göstermesinin kaynağını değerlendirmede istenilen yeterlikte olmayabilir.

Araştırmadaki diğer bir sınırlılık ise ülkelere ait örneklem büyüklüklerinin farklı olmasıdır. Analizlerde kullanılan ki-kare değerleri ve örneklem büyüklüğüne duyarlı diğer parametre kestirimleri eşit olmayan bu örneklem büyüklüklerinden farklı oranlarda etkilenmiş olabilir.

Ölçme değişmezliği analizlerinde, seçilen referans değişkenlerin DMF içermemesi gerekmektedir. Bu çalışmada Avustralya-Türkiye karşılaştırması için seçilen referans değişkenlerden biri DMF içermiştir. Bu durum, analizlerde birinci tip hata yapma olasılığını artırmış olabilir. Ayrıca çalışmadaki DMF analizlerinde tek bir yöntemin kullanılmış olması da diğer bir sınırlılıktır.

## 4.3 ÖNERİLER

### 4.3.1 PISA uygulayıcılarına öneriler

PISA kapsamında uygulanan anketlerden elde edilen veriler yardımıyla öğrenci başarısıyla ilişkili faktörler belirlenmeye çalışılmakta ve eğitim politikaları belirlenmektedir. Anketlerin madde ve ölçek bazında kültürler ve diller arası eşdeğerliğiyle ilgili çalışmalara araştırmacılar tarafından yeterince önem verilmediği görülmektedir. PISA ve diğer uluslararası çalışmalar kapsamında uygulanan, öğrenci, veli, okul ve yönetici anketleri üzerinde DMF analizleri düzenli olarak yapılmalı ve bu tür uygulamalardaki olası DMF sebepleri tespit edilmelidir.

Maddelerin kültürel farklılıklardan çok uyarlamadan kaynaklı problemlerden dolayı DMF gösterdiği sonucu dikkate alındığında, özellikle ülkemizde PISA uygulamalarından sorumlu MEB Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı'nın uygulamalarda görev alan çevirmenlerin seçimi ve eğitimi konusuna daha fazla önem vermesi gerektiği görülmektedir.

#### 4.3.2 Bundan sonra yapılacak araştırmalar için öneriler

Bu araştırma ve diğer birçok uluslararası karşılaştırma çalışmalarında, kültürler ve diller arasında genelde ölçme değişmezliğinin sağlanamadığı görülmektedir. Farklı dillere çevrilmiş, farklı kültürlerde ve eğitim sistemlerinde uygulanmış anket veya test maddelerinin tam olarak aynı şekilde yanıtlanmasının beklenmesi de pek mantıklı değildir. Asıl sorun ne seviyedeki ölçme değişmezliğinin problem yarattığı ve karşılaştırma çalışmalarında yanlılığa sebep olduğudur. Farklı seviyelerdeki ölçme değişmezliğinin etkileri yapılacak çalışmalarla incelenebilir.

Bu çalışmada referans değişkenin belirlenmesinde Stark, Chernyshenko ve Drasgow (2006) tarafından önerilen yöntem kullanılmıştır. Referans değişkenin DMF içermesi durumunda hem test hem de madde bazındaki ölçme değişmezliği analizleri gerçek sonuçları yansıtmayacaktır. Dolayısıyla referans değişkenin gruplar arasında gerçekten ölçme değişmezliğini sağladığından emin olunmalıdır. Bundan sonra yapılacak çalışmalarda referans değişkenin belirlenmesi konusunda literatürde önerilen farklı yöntemler kullanılıp bunların DMF analizleri üzerindeki etkileri belirlenebilir.

DMF gösteren maddelerin tespit edilmesinde bu çalışmada OKM'ye dayalı ÇG-DFA yöntemi kullanılmıştır. Hambleton ve Patsula (1999), kültürler arası karşılaştırmalarda, testin farklı dil versiyonlarının eşdeğerliğinden emin olabilmek için bir veya birden fazla yöntemle DMF analizlerinin yapılmasını önermektedirler. Araştırmacılar, Madde tepki kuramına dayalı DMF belirleme yöntemlerini, Mantel-Haenszel ya da lojistik regresyon gibi diğer yöntemleri kullanarak değişen fonksiyon gösteren maddeleri birden fazla yöntemle belirleyip sonuçları karşılaştırabilirler.

## KAYNAKÇA

- Allalouf, A., Hambleton, R., & Sireci, S. (1999). Identifying the Causes of DIF in Translated Verbal Items. *Journal of Educational Measurement*, 36, 185-198.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Angoff, W. H., & Cook, L. L. (1988). *Equating the Scores of the Prueba de Aptitud Academica and the Scholastic Aptitude Test* (Research Report No. 88-2). New York: College Entrance Examination Board.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.
- Brislin, R. W., Lonner, W., & Thorndike, R. M. (1973). *Cross Cultural Research Methods*. New York: Wiley.
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.
- Byrne, B. M. (2008). Testing for Multigroup Equivalence of a Measuring Instrument: A Walk Through the Process. *Psicothema*, 20 (4), 872-882
- Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance. *Psychological Bulletin*, 105, 456-466.
- Byrne, B. M., & Stewart, S. M. (2006). The MACS Approach to Testing for Multigroup Invariance of a Second-Order Structure: A Walk Through the Prcess. *Structural Equation Modeling*, 13(2), 287-321.
- Chan, D. (2000). Detection of Differential Item Functioning on the Kirton Adaptation-Innovation Inventory Using Multiple-Group Mean and Covariance Structure Analyses. *Multivariate Behavioral Research*, 35, 169-199.
- Cheung, G. W., & Rensvold, R. B. (1999). Testing Factorial Invariance Across Groups: A Reconceptualization and Proposed New Method. *Journal of Management*, 25, 1-7.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-fit Indexes for Testing MI. *Structural Equation Modeling*, 9, 233-255.
- De Klerk, G. (2008). Cross Cultural Testing. In M. Bron, C.D. Foxcroft & R. Butter (Eds.), *Online Readings in Testing and Assessment*, International Test Commission, <http://www.intestcom.org/Publications/ORTA.php>, Erişim Tarihi: 12.02.2009.

- Drasgow, F., & Kanfer, R. (1985). Equivalence of Psychological Measurement in Heterogeneous Populations. *Journal of Applied Psychology*, 70(4), 662-680.
- Ellis, B. B., & Raju, N. S. (2003). Test and Item Bias: What they Are, What they Aren't, and How to Detect Them. In: *Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators* (pp. 88- 98).
- Elousa, P., & López-Jaúregui, A. (2007). Potencial Sources of Differential Item Functioning. *International Journal of Testing*, 7 (1), 39-52
- Ercikan, K. (1998). Translation Effects in International Assessments. *International Journal of Educational Research*, 29, 543-553.
- Ercikan, K. (2002). Disentangling the Sources of Differential Item Functioning in Multilingual Assessments. *International Journal of Testing*, 2 (3&4), 199-215.
- Ercikan, K., Gierl, M. J., McCreith, T., Puhan, G., & Koh, K. (2004). Comparability of Bilingual Versions of Assessments: Sources of Incomparability of English and French Versions of Canada's National Achievement Tests. *Applied Measurement in Education*, 17(3), 301-321.
- Ercikan, K., & Koh, K. (2005). Examining the Construct Comparability of the English and French Versions of TIMSS. *International Journal of Testing*, 5(1), 23-35.
- Flora, D. B., & Curran, P. J. (2004). An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data. *Psychological Methods*, 9(4), 466-491.
- Gierl, M. J. (2000). Construct Equivalence on Translated Achievement Tests. *Canadian Journal of Education*, 25(4), 280-296.
- Gierl, M., J., & Khaliq, S., N. (2001). Identifying Sources of Differential Item and Bundle Functioning on Translated Achievement Tests: A Confirmatory Analysis. *Journal of Educational Measurement*, 38(2), 164-187.
- Grisay, A. (2002). Translation and Cultural Appropriateness of the Test and Survey Material. In R. J. Adams & M. Wu (Eds.). *PISA 2000. Technical Report* (pp. 57-70). Paris: OECD Publications.
- Grisay, A. (2003). Translation Procedures in OECD/PISA 2000 International Assessment. *Language Testing*, 20(2), 225-240
- Grisay, A., de Jong, J. H. A. L., Gebhardt, E., Berenzer, A., & Halleux-Monseur, B. (2007). Translation Equivalence Across PISA Countries. *Journal of Applied Measurement*, 8 (3), 249-266.
- Grisay, A. & Monseur, C. (2007). Measuring the Equivalence of Item Difficulty in the Various Versions of an International Test. *Studies in Educational Evaluation*, 33, 69-86.

- Hair, J. F., Anderson, R. E., Tatham, R. L. & Black, W. (1998). *Multivariate Data Analysis* (5th ed.). New York: Prentice Hall.
- Hambleton, R.K. (1994). Guidelines for Adapting Educational and Psychological Tests: A Progress Report. *European Journal of Psychological Assessment, 10*, 229-240.
- Hambleton, R.K. (2002). Adapting Achievement Tests into Multiple Languages for International Assessments. In A. Porter y A. Gamoran (Ed.), *Methodological Advances in Large-scale Cross-national Education Surveys* (pp. 58-79) Washington: National Academy of Sciences.
- Hambleton, R.K. (2005). Issues, Designs, and Technical Guidelines for Adapting Tests into Multiple Languages and Cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (pp. 3-38). Hillsdale, NJ: Lawrence Erlbaum.
- Hambleton, R.K., & de Jong, J. H. A. L. (2003). Advances in Translating and Adapting Educational and Psychological Tests. *Language Testing, 20*, 127-134.
- Hambleton, R. K., & Patsula, L. (1998). Adapting Tests for Use in Multiple Languages and Cultures. *Social Indicators Research, 45*, 153-171.
- Hambleton, R. K., & Patsula, L. (1999). Increasing the Validity of Adapted Tests: Myths to be Avoided and Guidelines for Improving Test Adaptation Practices. *Journal of Applied Testing Technology, 1*(1), 1-30.
- Harkness, J. (1998). *Response Scales in Cross-National Survey Research*. Paper Presented at the Meeting of the American Psychological Association, Toronto, Canada.
- Harrington, D. (2009). *Confirmatory Factor Analysis*. (Pocket Guide Series), New York: Oxford University Press.
- Hernandez, A., & Gonzalez-Roma, V. (2003). Evaluating the Multiple-Group Mean and Covariance Structure Analysis Model for The Detection of Differential Item Functioning in Polytomous Ordered Items. *Psicothema, 15*(2), 322-327.
- Hui, C. H., & Triandis, H. C. (1985). Measurement in Cross-Cultural Psychology: A Review and Comparison of strategies. *Journal of Cross-Cultural Psychology, 16*(2), 131-152.
- Hui, C. H., & Triandis, H. C. (1989). Effects of Culture and Response Format on Extreme Response Style. *Journal of Cross-Cultural Psychology, 20*(3), 296-309.
- International Test Commission (2001). *International Test Commission Guidelines for Test Adaptation*. London: Author.
- Johnson, T. P. (1998). Approaches to Equivalence in Cross-Cultural and Cross-National Survey Research. *ZUMA Nachrichten Spezial, Cross Cultural Survey Equivalence, 3*, 1-40.

- Johnson, E. C., & Meade, A. W. (2007). *The Role of Referent Indicators in Tests of Measurement Invariance*. Paper presented at the 22nd Annual Meeting of the Society for Industrial and Organizational Psychology, New York.
- Jöreskog, K. G. (1971). Simultaneous Factor Analysis in Several Populations. *Psychometrika*, *36*, 409-426.
- Jöreskog, K. G. (2005). *Structural Equation Modeling with Ordinal Variables Using LISREL*. [www.ssicentral.com/lisrel/techdocs/ordinal.pdf](http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf), Erişim Tarihi: 15.04.2008.
- Jöreskog, K. G., & Sörbom, D., (1999). *Lisrel 8.30: Structural Equation Modeling with the Simplis Command Language*, Chicago, IL: Scientific Software International, Inc.
- Jöreskog, K. G., & Sörbom, D. (2001). *Lisrel 8: User's Reference Guide*. Chicago: Scientific Software International.
- Jöreskog, K. G., & Sörbom, D. (2002). *PRELIS 2: User's Reference Guide*. Chicago: Scientific Software International.
- Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. (2<sup>nd</sup> ed.) New York: the Guilford Press.
- Le, L. T. (2006). *Analysis of Differential Item Functioning*. Paper Presented at the Annual Meeting of American Educational Research Association in San Francisco, 7-11 April 2006.
- Le, L. T. (2009). Investigating Gender Differential Item Functioning Across Countries and Test Languages for PISA Science Items. *International Journal of Testing*, *9*(2), 122-133.
- Little, T. D. (1997). Mean and Covariance Structures (MACS)Analyses of Cross-cultural Data: Practical and Theoretical Issues. *Multivariate Behavioral Research*, *32*, 53-76.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Lubke, G. H., & Muthen, B. O. (2004). Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons. *Structural Equation Modeling*, *11*(4), 514-534.
- Meade, A. W., & Lautenschlager, G. J. (2004). A Comparison of Item Response Theory and Confirmatory Factor Analytic Methodologies for Establishing Measurement Equivalence/Invariance. *Organizational Research Methods*, *7*(4), 361-388.
- MEB (2007). *PISA 2006 Uluslararası Öğrenci Değerlendirme Programı Ulusal Ön Rapor*. MEB, Ankara

- MEB Web Sayfası: <http://earged.meb.gov.tr/pisa/dil/tr/pisanedir.html>, Erişim Tarihi: 12.10.2009.
- Mellenbergh, G. J. (1982). Contingency Table Models for Assessing Item Bias. *Journal of Educational Statistics*, 7, 105-118.
- Mellenbergh, G. J. (1989). Item Bias and Item Response Theory. *International Journal of Educational Research*, 13, 127-143.
- Meredith, W. (1964). Notes on Factorial Invariance. *Psychometrika*, 29, 177-185.
- Meredith, W. (1993). Measurement Invariance, Factor Analysis and Factorial Invariance. *Psychometrika*, 58, 525-543.
- Meredith, W. & Millsap, R. E. (1992). On the Misuse of Manifest Variables in the Detection of Measurement Invariance. *Psychometrika*, 57(2), 289-311.
- Meredith, W., & Teresi, J. A. (2006). An Essay on Measurement and Factorial Invariance. *Medical Care*, 44, 69-77.
- Millsap, R. E. (2001). When Trivial Constraints Are Not Trivial: The Choice of Uniqueness Constraints in Confirmatory Factor Analysis. *Structural Equation Modeling*, 8, 1-17.
- OECD (1999). *Measuring Student Knowledge and Skills: A new Framework for Assessment*. Paris: OECD Publications
- OECD (2002). *PISA 2000 Technical Report*. Paris: OECD Publications.
- OECD (2004). *PISA 2006 Translation and Adaptation Guidelines*. Doc: NPM (0409)13, National Project Managers' Meeting, Bratislava, Slovak Republic.
- OECD (2005). *School Sampling Preparation Manuel: PISA 2006 Main Study*. Paris: OECD Publications.
- OECD (2007a). *PISA 2006: Science Competencies for Tomorrow's World: Volume 1 and Volume 2*. Paris: OECD Publications.
- OECD (2007b). *PISA 2009 Translation and Adaptation Guidelines*. Doc: NPM (0709)1, National Project Managers' Meeting, Dubrovnik, Croatia.
- OECD (2009). *PISA 2006 Technical Report*. Paris: OECD Publications.
- Oort, F. J. (1998). Simulation Study of Item Bias Detection With Restricted Factor Analysis. *Structural Equation Modeling*, 5, 107-125.
- Öğretmen, T. (2006). *Uluslararası Okuma Becerilerinde Gelişim Projesi (PIRLS) 2001 Testinin psikometrik Özelliklerinin İncelenmesi: Türkiye- Amerika Birleşik Devletleri Örneği*. Yayınlanmamış Doktora Tezi. Eğitim Bilimleri Bölümü, Hacettepe Üniversitesi, Ankara.

- Paulhus, D. L. (1991). Measurement and Control of Response Bias. In J.P. Robinson, P.R. Shaver & L.S. Wrightsman (eds.), *Measures of Personality and Social Psychological Attitudes* (Vol. 1, pp. 17-59). San Diego: Academic Press.
- PISA Web Site: [www.pisa.oecd.org](http://www.pisa.oecd.org), Erişim Tarihi: 21.01.2009.
- Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement Equivalence: A Comparison of Methods Based on Confirmatory Factor Analysis and Item Response Theory. *Journal of Applied Psychology*, 87(3), 527-529.
- Raykow, T., & Marcoulides, G. A. (2006). *A First Course in Structural Equation Modeling* (2 nd. ed). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory Factor Analysis and Item Response Theory: Two Approaches for Exploring Measurement Invariance. *Psychological Bulletin*, 114, 552-566
- Rensvold, R. B., & Cheung, G. W. (1999). Identification of Influential Cases in Structural Equation Models Using the Jackknife Method. *Organizational Research Methods*, 2(3), 293-308
- Schmitt, N., & Kuljanin, G. (2008). Measurement Invariance: Review of Practice and Implications. *Human Resource Management Review*, 18, 210-222.
- Schulz, W. (2003). *Validating Questionnaire Constructs in International Studies. Two Examples from PISA 2000*. Paper Presented at the Annual Meetings of the American Educational Research Association (AERA) in Chicago, 21-25 April.
- Schulz, W. (2005). *Testing Parameter Invariance for Questionnaire Indices Using Confirmatory Factor Analysis and Item Response Theory*. Paper Presented at the Annual Meetings of the American Educational Research Association (AERA) in San Francisco, 7-11 April.
- Schulz, W. (2008). *Questionnaire Construct Validation in the International Civic and Citizenship Education Study*. Paper presented to the 3rd IEA International Research Conference in Taipei, September.
- Schumacker, R. E., & Lomax, R. G. (2004). *A Beginner's Guide to Structural Equation Modeling, Second Edition*. Mahwah, NJ: Erlbaum and Associates.
- Sireci, S. G. (2005). Using Bilinguals to Evaluate the Comparability of Different Language Versions of a Test. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (pp. 93-115). Hillsdale, NJ: Lawrence Erlbaum.
- Sireci, S. G., & Allalouf, A. (2003). Appraising Item Equivalence Across Multiple Languages and Cultures. *Language Testing* 20(2), 148-166.

- Sireci, S. G., & Bastari, B. (1998). *Evaluating Construct Equivalence Across Adapted Tests*. Paper Presented at the Annual Meeting of the American Psychological Association in San Francisco, CA, 14-18 August.
- Sireci, S. G., & Berberoğlu, G. (2000). Using Bilingual Respondents to Evaluate Translated – Adapted Items. *Applied Measurement in Education*, 13(3), 229-248.
- Sireci, S. G., Patsula, L., & Hambleton, R. K. (2005). Statistical Methods for Identifying Flaws in the Test Adaptation Process. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (pp. 93-115). Hillsdale, NJ: Lawrence Erlbaum.
- Sireci, S. G., & Swaminathan, H. (1996). *Evaluating Translation Equivalence: So What's the Big Dif?* Paper Presented at the Annual Meeting of the Northeastern Educational Research Association, Ellenville, NY, October.
- Sörbom, D. (1974). A General Method for Studying Differences in Factor Means and Factor Structures between Groups. *British Journal of Mathematical and Statistical Psychology*, 27, 229-239.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting Differential Item Functioning with Confirmatory Factor Analysis and Item Response Theory: Toward a Unified Strategy. *Journal of Applied Psychology*, 91(6), 1292-1306.
- Şimşek, Ö. F. (2007). *Yapısal Eşitlik Modellemesine Giriş: Temel İlkeler ve LISREL Uygulamaları*. Kızılay: Ankara. Ekinoks Eğitim Danışmanlık Hiz. ve Bas. Yay. Dağ. San. ve Tic. Ltd. Şti.
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Boston MA: Allyn & Bacon.
- Turner, R., & Adams, R. J. (2007). The Programme for International Student Assessment: an Overview. *Journal of Applied Measurement*, 8(3), 237-248.
- Walker, M (2007). Ameliorating Culturally Based Extreme Response Tendencies to Attitude Items. *Journal of Applied Measurement*, 8(3), 267-278.
- Van de Vijver, F. J. R (1998). Towards a Theory of Bias and Equivalence. *ZUMA Nachrichten Spezial, Cross Cultural Survey Equivalence*, 3, 41-65.
- Van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating Tests: Some Practical Guidelines. *European Psychologists*, 1, 89-99.
- Van de Vijver, F. J. R., & Poortinga, Y. H. (2005). Conceptual and Methodological Issues in Adapting Tests. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (pp. 39-63). Hillsdale, NJ: Lawrence Erlbaum.

- Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and Equivalence in Cross-Cultural Assessment. *European Review of Applied Psychology, 54*, 119-135.
- Van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response Styles in Rating Scales: Evidence of Method Bias in Data From Six EU Countries. *Journal of Cross-Cultural Psychology, 35*(3), 346-360
- Vandenberg, R. J. (2002). Toward a Further Understanding of and Improvement in Measurement Invariance Methods and Procedures. *Organizational Research Methods, 5*(2), 139-158.
- Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the MI Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods, 3*, 4-69.
- Widaman, K. F., & Reise, S. P. (1997). Exploring the MI of Psychological Instruments: Applications in the Substance Use Domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The Science of Prevention: Methodological Advances From Alcohol and Substance Abuse Research* (pp. 281-324). Washington, DC: American Psychological Association.
- Wu, Amery D., Li, Zhen & Zumbo, Bruno D. (2007). Decoding the Meaning of Factorial Invariance and Updating the Practice of Multi-group Confirmatory Factor Analysis: A Demonstration with TIMSS Data. *Practical Assessment Research & Evaluation, 12*(3). <http://pareonline.net/pdf/v12n3.pdf>, Erişim Tarihi: 07.07.2009.
- Yayan, B., & Berberoğlu, G. (2004). A Re-Analysis of The TIMSS 1999 Mathematics Assessment data of the Turkish Students. *Studies in Educational Evaluation, 30*, 87-104.
- Yıldırım, S. (2008). Comparison of Restricted Factor Analysis with Likelihood-Ratio and Mantel-Haenszel Methods in DIF Analyses. *H.U. Journal of Education, 34*, 297-307.
- Yıldırım, H. H (2006). The Differential Item Functioning (DIF) Analysis of Mathematics Items in the International Assessments Programs. Yayınlanmamış Doktora Tezi. Orta Doğu Teknik Üniversitesi, Ankara.
- Yıldırım, H. H. & Berberoğlu, G. (2009). Judgemental and Statistical Analyses of the PISA-2003 Mathematics Literacy Items. *International Journal of Testing, 9*(2), 108-121.
- Yoon, M., & Millsap, R. E. (2007). Detecting Violations of Factorial Invariance Using Data-based Specification Searches: A Monte Carlo Study. *Structural Equation Modeling, 14*, 435-463.

- Zumbo, B. D., (2003). Does Item-Level DIF Manifest Itself in Scale-Level Analyses? Implications for Translating Language Tests. *Language Testing*, 20 (2), 136-147.
- Zumbo, B. D. (2005). Manifestation Of Differences In Item-Level Characteristics in Scale-Level Measurement Invariance Tests Of Multi-Group Confirmatory Factor Analyses. *Journal of Modern Applied Statistical Methods*, 4 (1), 275-282.
- Zumbo, B. D., Sireci, S. G., & Hambleton, R. K. (2003). *Re-Visiting Exploratory Methods for Construct Comparability: Is There Something to be Gained From the Ways of Old?* Paper Presented in the Symposium Construct Comparability Research: Methodological Issues and Results, National council on Measurement in Education, April, Chicago, IL.
- Zumbo, B. D. (2007). Validity: Foundational Issues and Statistical Methodology. In C. R. Rao and S. Sinharay (Eds.), *Handbook of Statistics, Vol.26: Psychometrics* (pp. 45-79). Elsevier Science B. V.: The Netherlands.

**Ek 1 Avustralya, Yeni Zelanda ve ABD Öğrenci Anketlerinde Bilimsel  
Sorgulamaya Verilen Destek Tutumunu Ölçmek için Kullanılan Maddeler**

<b>Q18</b>	<b>How much do you agree with the statements below?</b> (Please tick one box in each row)
ST18Q01	a) Advances in science and technology usually improve people's living conditions
ST18Q02	b) Science is important for helping us to understand the natural world
ST18Q03	c) Some concepts in science help me see how I relate to other people
ST18Q04	d) Advances in science and technology usually help improve the economy
ST18Q05	e) I will use science in many ways when I am an adult
ST18Q06	f) Science is valuable to society
ST18Q07	g) Science is very relevant to me
ST18Q08	h) I find that science helps me to understand the things around me
ST18Q09	i) Advances in science and technology usually bring social benefits
ST18Q10	j) When I leave school there will be many opportunities for me to use science

*Not.* Maddeler 4 dereceli Likert formatındadır (1= strongly agree, 2= agree, 3= disagree, 4 = strongly disagree). Bu soru Avustralya öğrenci anketinde 19. soru olarak yer almıştır.

**Ek 2 Türkiye Öğrenci Anketinde Bilimsel Sorgulamaya Verilen Destek Tutumunu Ölçmek için Kullanılan Maddeler**

<b>S18</b>	<b>Aşağıdaki ifadelere ne ölçüde katılıyorsunuz?</b> (Her sırada sadece bir kutuyu işaretleyiniz)
ST18Q01	a) Fen bilimleri ve teknolojideki ilerlemeler genellikle insanların yaşam koşullarını iyileştirir
ST18Q02	b) Fen bilimleri, doğal dünyayı anlamamıza yardımcı olması açısından önemlidir
ST18Q03	c) Fen bilimlerindeki bazı kavramlar, diğer insanlarla nasıl bir bağlantı içinde olduğumu görmeme yardımcı olur
ST18Q04	d) Fen bilimleri ve teknolojideki ilerlemeler genellikle ekonominin iyileştirilmesine yardımcı olur
ST18Q05	e) Bir yetişkin olduğumda fen bilimlerinden birçok yönde yararlanacağım
ST18Q06	f) Fen bilimleri toplum için yararlıdır
ST18Q07	g) Fen bilimleri beni çok ilgilendiriyor
ST18Q08	h) Fen bilimlerinin, çevremdeki şeyleri anlamada bana yardımcı olduğu görüşündeyim
ST18Q09	i) Fen bilimleri ve teknolojideki ilerlemeler genellikle toplumsal yararlar sağlar
ST18Q10	j) Okulu bitirdikten sonra, fen bilimlerinden yararlanmam için pek çok fırsat olacaktır

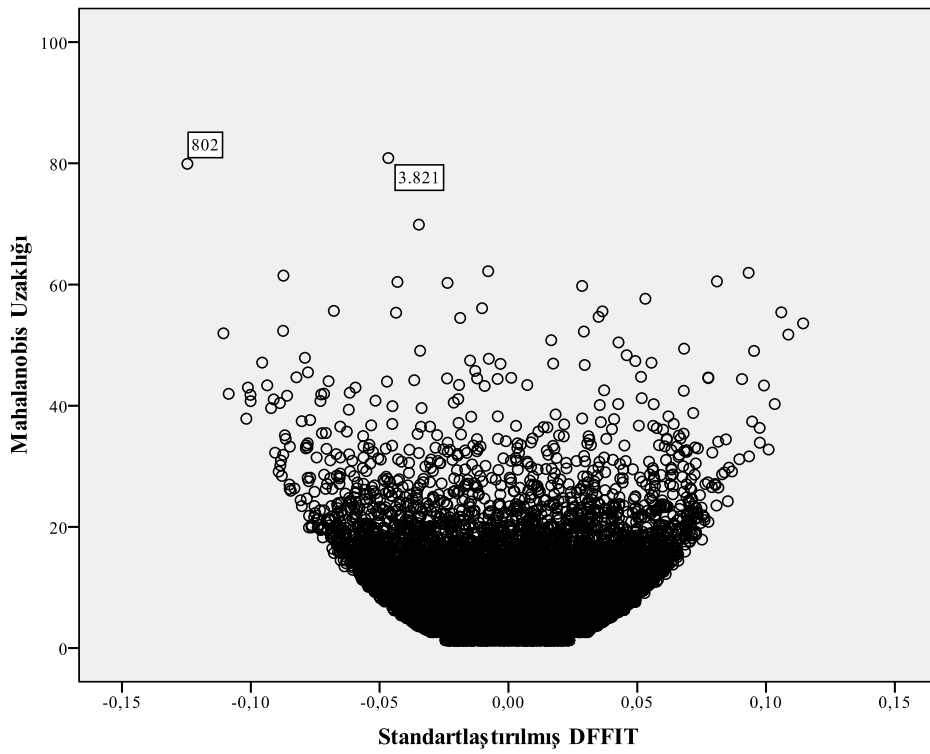
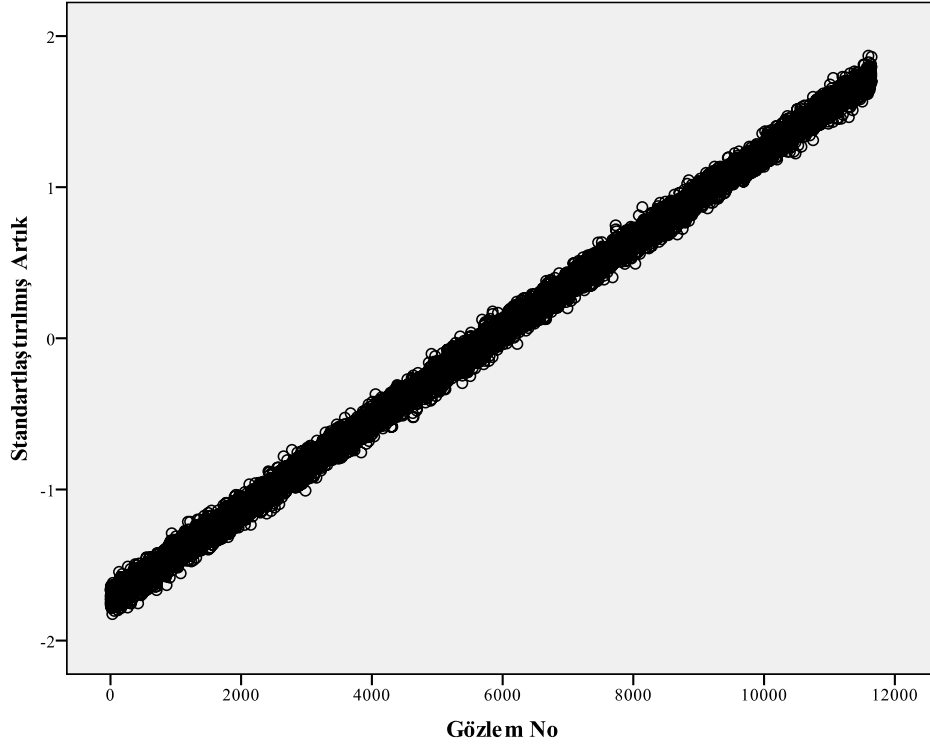
*Not.* Maddeler 4 dereceli Likert formatındadır (1= tümüyle katılıyorum, 2= katılıyorum, 3= katılmıyorum, 4 = hiç katılmıyorum).

### Ek 3 Kayıp Veri Oranları

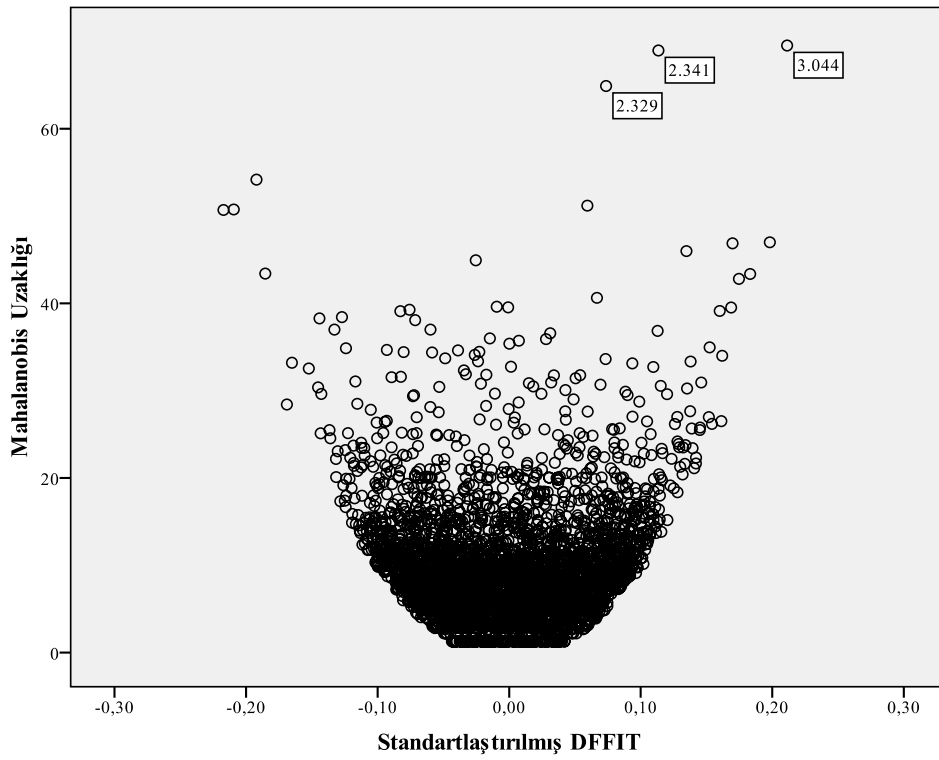
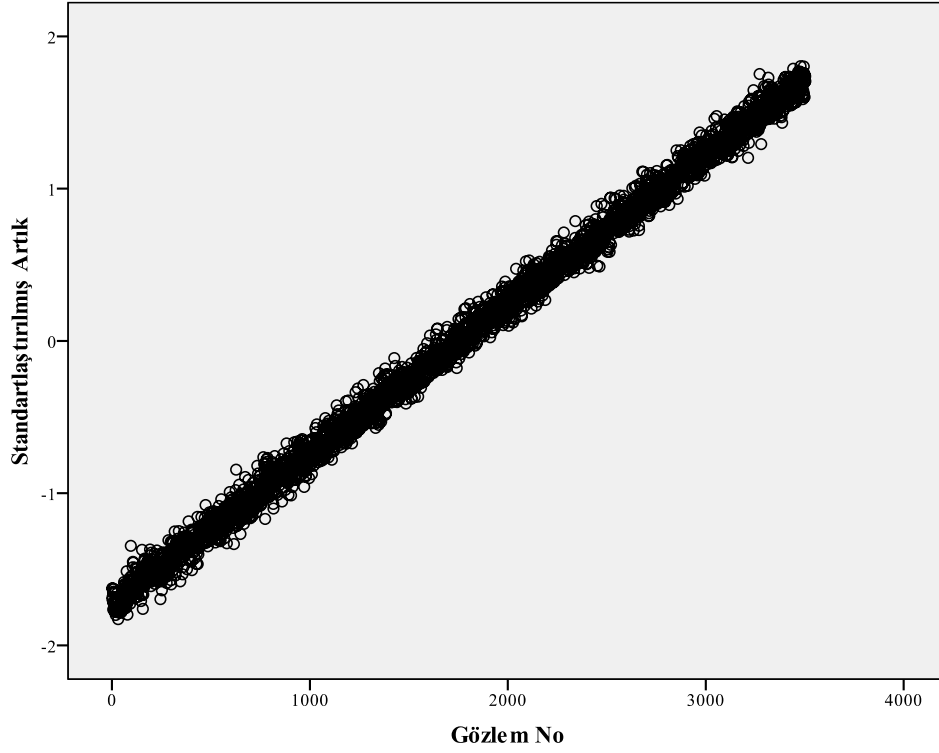
	Avustralya		Yeni Zelanda		ABD		Türkiye	
	Geçerli veri	Kayıp veri	Geçerli veri	Kayıp veri	Geçerli veri	Kayıp veri	Geçerli veri	Kayıp veri
ST18Q01	11891	55	3609	19	4626	45	4660	19
ST18Q02	11886	60	3608	20	4624	47	4655	24
ST18Q03	11879	67	3609	19	4622	49	4641	38
ST18Q04	11850	96	3599	29	4612	59	4647	32
ST18Q05	11855	91	3596	32	4615	56	4648	31
ST18Q06	11877	69	3603	25	4612	59	4650	29
ST18Q07	11862	84	3594	34	4605	66	4616	63
ST18Q08	11867	79	3604	24	4611	60	4641	38
ST18Q09	11862	84	3590	38	4611	60	4642	37
ST18Q10	11866	80	3600	28	4617	54	4648	31

*Not.* Her bir madde için kayıp veri oranı 0,004 – 0,014 arasında değişmektedir.

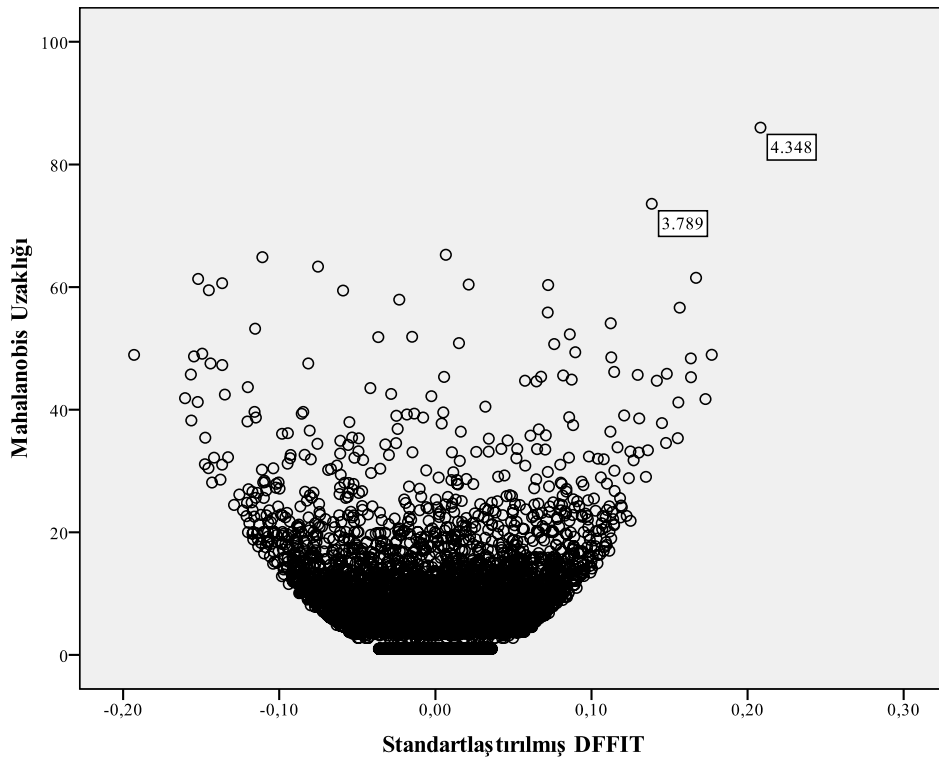
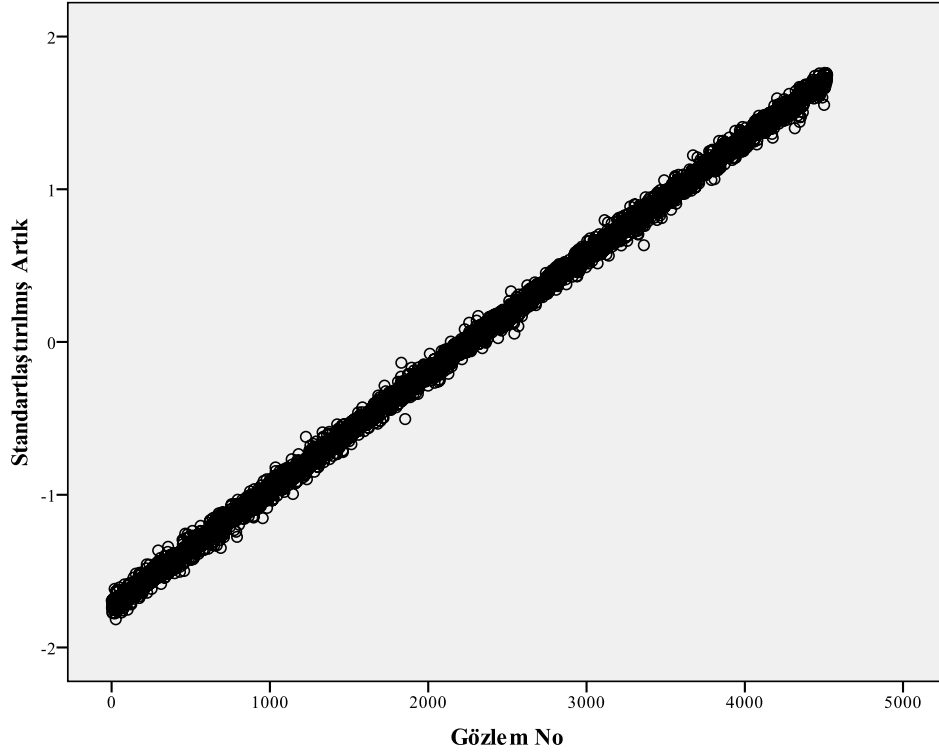
#### Ek 4 Avustralya Örnekleme için Aykırı Değer Analizi Grafikleri



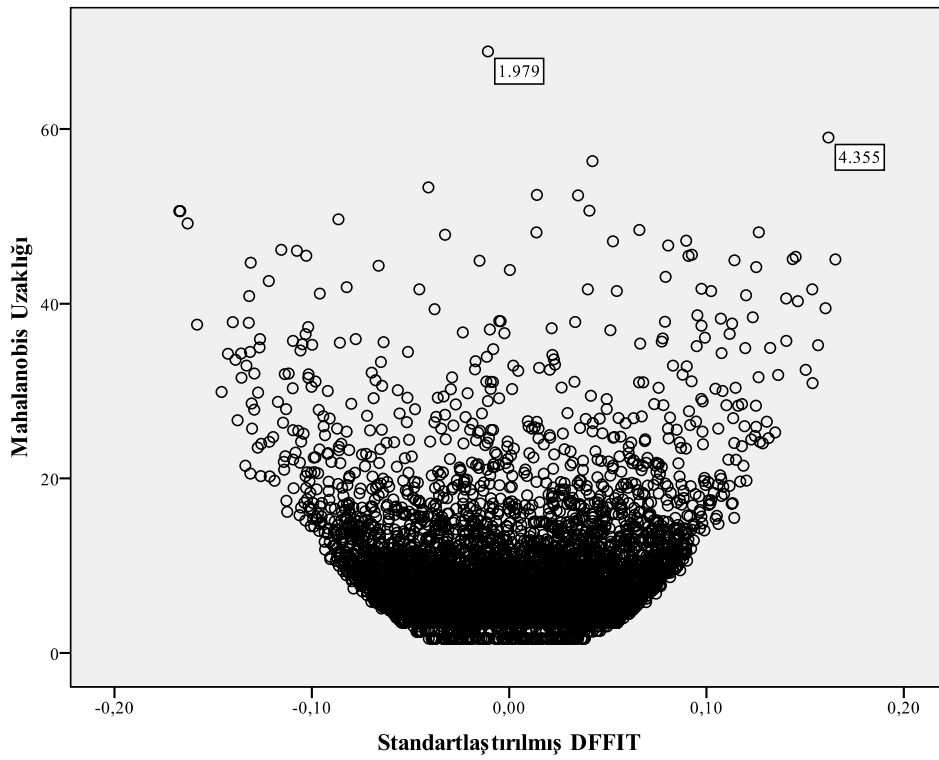
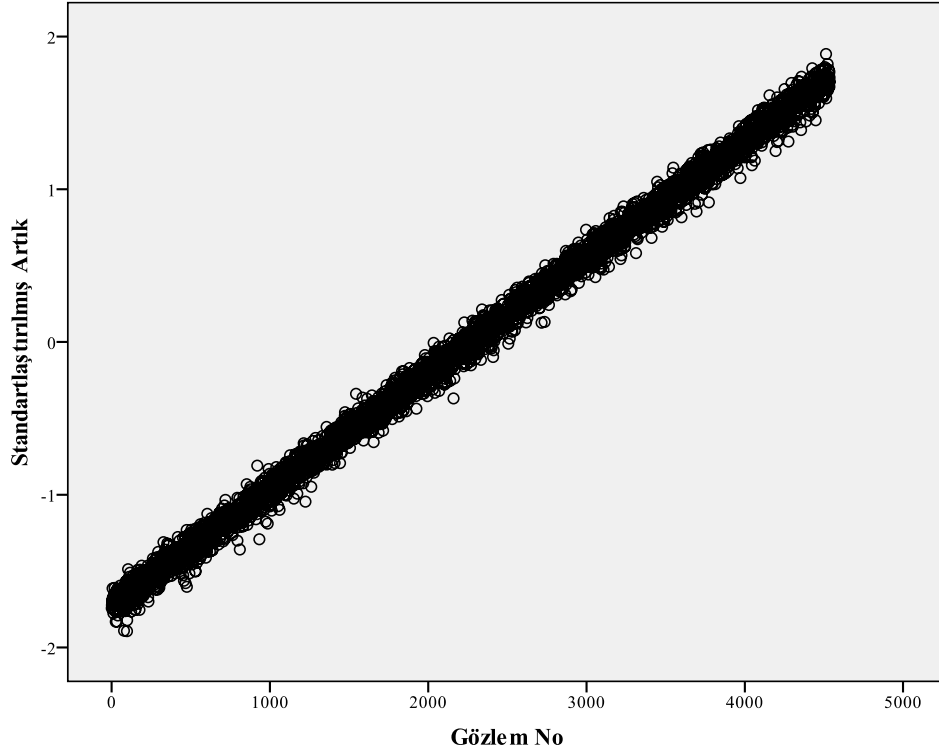
### Ek 5 Yeni Zelanda Örneklemi için Aykırı Değer Analizi Grafikleri



### Ek 6 ABD Örnekleme için Aykırı Değer Analizi Grafikleri



### Ek 7 Türkiye Örneklemi için Aykırı Değer Analizi Grafikleri



**Ek 8 Avustralya Örneklemi için Çoklu Bağlantılılık Analizi Sonuçları**

Dimension	Condition Index	ST18Q01	ST18Q02	ST18Q03	ST18Q04	ST18Q05	ST18Q06	ST18Q07	ST18Q08	ST18Q09	ST18Q10
1	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2	8,78	0,09	0,03	0,01	0,05	0,05	0,00	0,04	0,02	0,01	0,08
3	11,52	0,11	0,15	0,21	0,01	0,00	0,09	0,00	0,00	0,08	0,00
4	12,85	0,10	0,20	0,24	0,05	0,01	0,00	0,00	0,11	0,21	0,05
5	13,41	0,12	0,00	0,03	0,02	0,06	0,15	0,00	0,15	0,39	0,06
6	13,94	0,32	0,00	0,01	0,36	0,00	0,26	0,00	0,07	0,09	0,00
7	14,86	0,02	0,11	0,02	0,36	0,01	0,10	0,00	0,48	0,07	0,01
8	15,45	0,13	0,32	0,01	0,14	0,00	0,18	0,01	0,04	0,12	0,05
9	15,70	0,09	0,14	0,35	0,00	0,04	0,09	0,03	0,06	0,00	0,12
10	17,50	0,02	0,04	0,14	0,00	0,34	0,10	0,11	0,00	0,03	0,63
11	18,94	0,00	0,01	0,00	0,00	0,48	0,02	0,80	0,08	0,00	0,00
Tolerance		0,60	0,54	0,69	0,64	0,37	0,50	0,34	0,49	0,64	0,41
VIF		1,66	1,85	1,45	1,57	2,70	1,99	2,97	2,07	1,57	2,42

**Ek 9 Yeni Zelanda Örneklemi için Çoklu Bağlantılılık Analizi Sonuçları**

Dimension	Condition Index	ST18Q01	ST18Q02	ST18Q03	ST18Q04	ST18Q05	ST18Q06	ST18Q07	ST18Q08	ST18Q09	ST18Q10
1	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2	8,52	0,07	0,02	0,01	0,05	0,05	0,00	0,04	0,02	0,02	0,09
3	11,89	0,08	0,25	0,15	0,00	0,00	0,09	0,00	0,02	0,13	0,01
4	12,32	0,12	0,10	0,26	0,11	0,02	0,00	0,00	0,18	0,03	0,07
5	13,12	0,11	0,02	0,08	0,01	0,02	0,29	0,00	0,04	0,41	0,05
6	14,04	0,08	0,00	0,05	0,02	0,02	0,42	0,00	0,32	0,17	0,03
7	14,38	0,04	0,15	0,01	0,56	0,00	0,00	0,00	0,30	0,14	0,02
8	15,09	0,47	0,41	0,03	0,18	0,00	0,05	0,03	0,05	0,02	0,13
9	16,07	0,02	0,05	0,00	0,02	0,27	0,15	0,17	0,02	0,06	0,54
10	16,64	0,01	0,00	0,40	0,04	0,00	0,01	0,02	0,02	0,03	0,05
11	18,20	0,00	0,00	0,00	0,01	0,61	0,00	0,73	0,03	0,00	0,01
Tolerance		0,63	0,58	0,75	0,66	0,39	0,55	0,38	0,53	0,69	0,46
VIF		1,59	1,71	1,34	1,51	2,54	1,82	2,61	1,91	1,44	2,16

**Ek 10 ABD Örnekleme için Çoklu Bağlantılılık Analizi Sonuçları**

Dimension	Condition Index	ST18Q01	ST18Q02	ST18Q03	ST18Q04	ST18Q05	ST18Q06	ST18Q07	ST18Q08	ST18Q09	ST18Q10
1	1,000	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2	8,656	0,13	0,05	0,00	0,06	0,07	0,00	0,06	0,02	0,00	0,08
3	11,066	0,12	0,10	0,19	0,04	0,01	0,11	0,00	0,00	0,11	0,01
4	12,041	0,00	0,09	0,31	0,11	0,00	0,01	0,01	0,07	0,31	0,05
5	12,700	0,15	0,02	0,00	0,00	0,09	0,06	0,00	0,28	0,18	0,10
6	13,093	0,25	0,00	0,00	0,59	0,04	0,04	0,00	0,00	0,24	0,02
7	13,726	0,25	0,50	0,00	0,02	0,02	0,01	0,09	0,06	0,01	0,27
8	13,996	0,08	0,00	0,00	0,13	0,02	0,31	0,03	0,29	0,03	0,18
9	14,154	0,01	0,01	0,30	0,00	0,08	0,01	0,00	0,16	0,06	0,00
10	15,461	0,00	0,21	0,19	0,00	0,20	0,42	0,07	0,00	0,05	0,26
11	16,257	0,00	0,01	0,01	0,04	0,46	0,02	0,73	0,11	0,01	0,03
Tolerance		0,60	0,53	0,65	0,62	0,45	0,48	0,43	0,51	0,62	0,51
VIF		1,67	1,88	1,53	1,62	2,23	2,09	2,34	1,98	1,62	1,98

**Ek 11 Türkiye Örneklemi için Çoklu Bağlantılılık Analizi Sonuçları**

Dimension	Condition Index	ST18Q01	ST18Q02	ST18Q03	ST18Q04	ST18Q05	ST18Q06	ST18Q07	ST18Q08	ST18Q09	ST18Q10
1	1,000	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2	8,089	0,10	0,04	0,00	0,06	0,08	0,02	0,08	0,02	0,03	0,10
3	10,213	0,08	0,18	0,13	0,12	0,01	0,17	0,00	0,00	0,18	0,00
4	10,300	0,06	0,02	0,05	0,55	0,00	0,23	0,01	0,01	0,01	0,01
5	11,026	0,38	0,00	0,29	0,02	0,04	0,07	0,01	0,10	0,01	0,07
6	11,555	0,00	0,00	0,00	0,13	0,14	0,18	0,01	0,01	0,46	0,14
7	12,173	0,00	0,02	0,08	0,09	0,01	0,13	0,05	0,47	0,00	0,06
8	12,757	0,02	0,02	0,13	0,02	0,40	0,16	0,00	0,05	0,23	0,18
9	13,414	0,29	0,64	0,05	0,01	0,02	0,00	0,00	0,14	0,05	0,08
10	13,903	0,03	0,03	0,18	0,01	0,29	0,02	0,40	0,14	0,01	0,02
11	14,055	0,05	0,04	0,08	0,01	0,02	0,03	0,44	0,05	0,02	0,34
Tolerance		0,62	0,58	0,67	0,72	0,55	0,62	0,51	0,56	0,62	0,62
VIF		1,61	1,74	1,49	1,40	1,82	1,62	1,98	1,80	1,60	1,62

**Ek 12 Birleştirilmiş Veriden Ortak Eşik Değerlerin Tespit Edilmesinde Kullanılan  
PRELIS Komut Dosyası**

```
!PRELIS SYNTAX: Equal Threlsholds  
SY='POOLED.PSF'  
ET ST18Q01 ST18Q02 ST18Q03 ST18Q04 ST18Q05 ST18Q06 ST18Q07 ST18Q08  
ST18Q09 ST18Q10  
OU MA=KM SM=POOLED.COR ME=POOLED.ME AC=POOLED.ACM XM  
TH=POOLED.THR
```

**Ek 13 Eşik Değerlerin Sabitletmesinde Kullanılan PRELIS Komut Dosyası**

```
!PRELIS SYNTAX: Fix Thresholds
SY='TUR.PSF'
FT=POOLED.THR ST18Q01
FT ST18Q02
FT ST18Q03
FT ST18Q04
FT ST18Q05
FT ST18Q06
FT ST18Q07
FT ST18Q08
FT ST18Q09
FT ST18Q10
OU MA=KM SM=TUR.COR ME=TUR.ME AC=TUR.ACM XM
```

**Ek 14 Doğrulayıcı Faktör Analizi LISREL-SIMPLIS Komut Dosyası**

TÜRKİYE DFA

Observed Variables

ST18Q01 ST18Q02 ST18Q03 ST18Q04 ST18Q05

ST18Q06 ST18Q07 ST18Q08 ST18Q09 ST18Q10

Correlation matrix from file TUR.COR

Asymptotic covariance matrix from file TUR.ACM

Sample Size = 4535

Latent Variables GD KD

Relationships

ST18Q01 = 1\*GD

ST18Q02 = GD

ST18Q04 = GD

ST18Q06 = GD

ST18Q09 = GD

ST18Q03 = 1\*KD

ST18Q05 = KD

ST18Q07 = KD

ST18Q08 = KD

ST18Q10 = KD

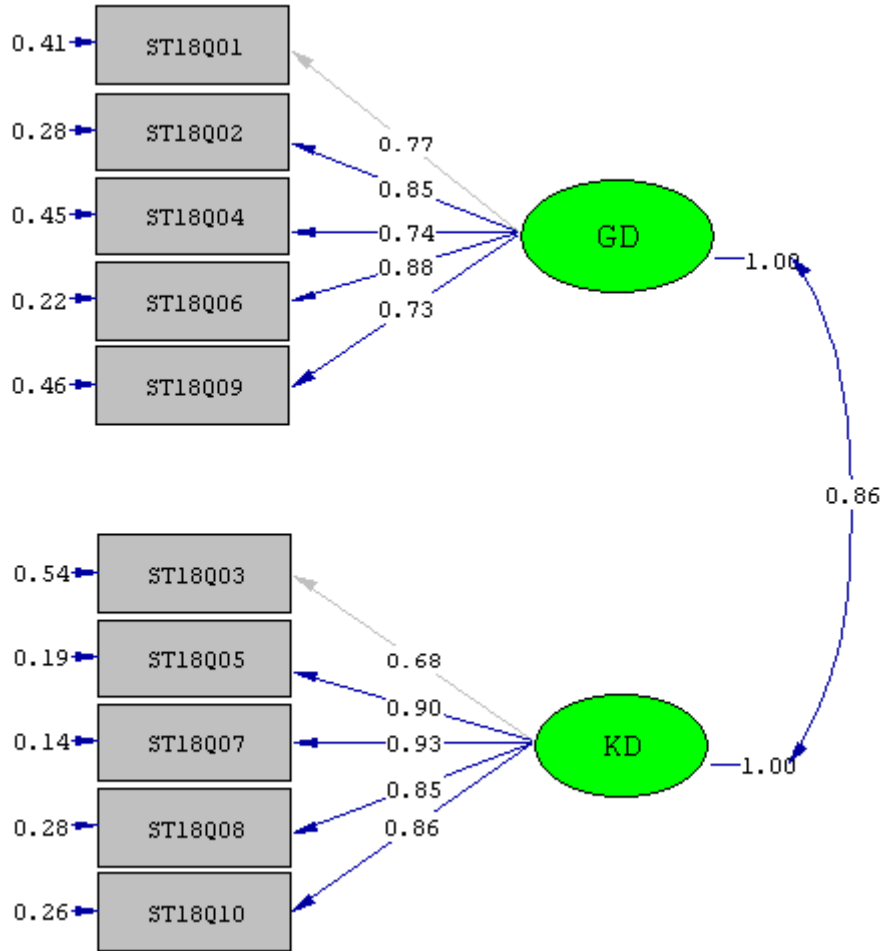
Method of Estimation Weighted Least Squares

Options ND=3 MI SC

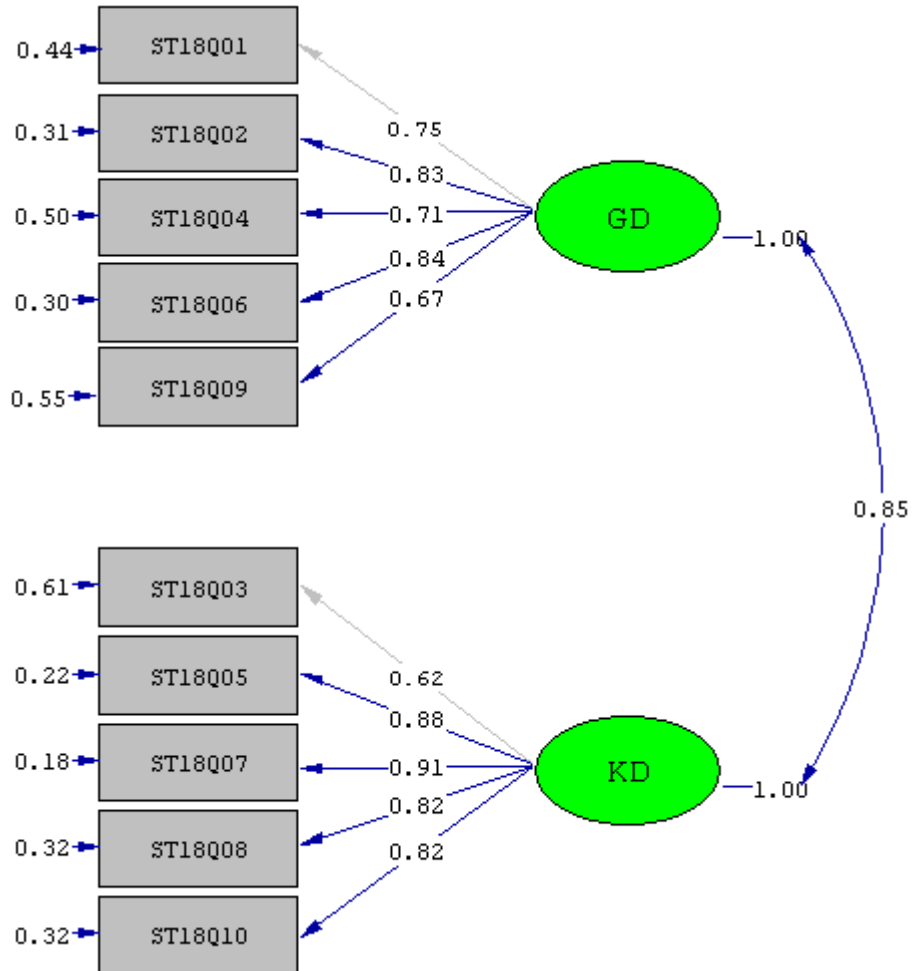
Path Diagram

End of Problem

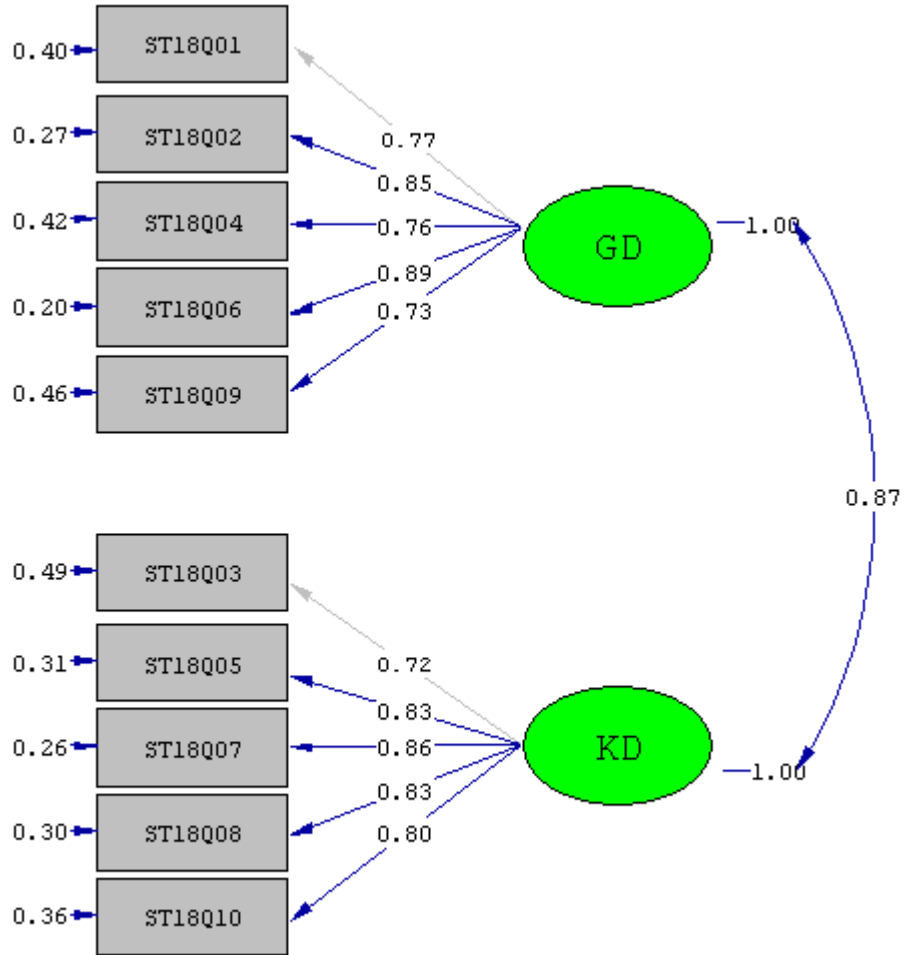
**Ek 15 Avustralya Örneklemi için Standartlaştırılmış Faktör Yükleri ve Hata Varyansları**



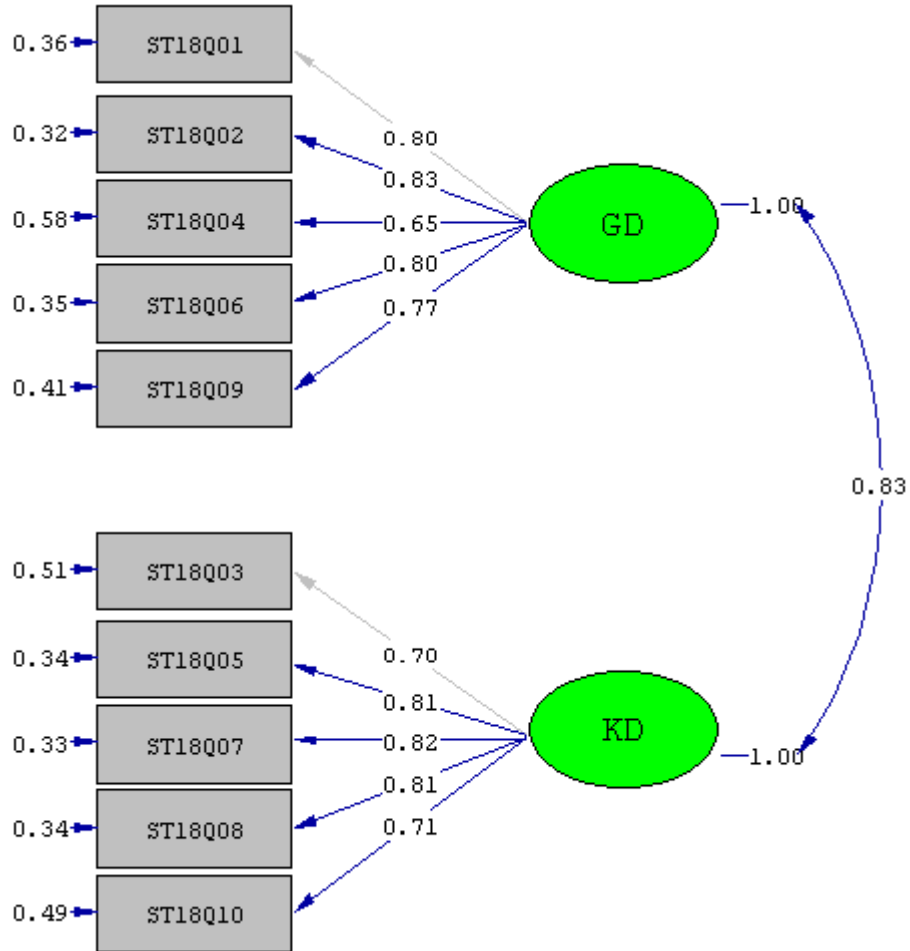
**Ek 16 Yeni Zelanda Örnekleme için Standartlaştırılmış Faktör Yükleri ve Hata Varyansları**



**Ek 17 ABD Örnekleme için Standartlaştırılmış Faktör Yükleri ve Hata Varyansları**



**Ek 18 Türkiye Örnelemi için Standartlaştırılmış Faktör Yükleri ve Hata Varyansları**



**Ek 19 Referans Değişkeni Belirlemede Kullanılan Sınırlandırılmış Temel Model  
LISREL-SIMPLIS Komut dosyası**

Group Reference AVUSTRALYA

Observed Variables

ST18Q01 ST18Q02 ST18Q03 ST18Q04 ST18Q05

ST18Q06 ST18Q07 ST18Q08 ST18Q09 ST18Q10

Correlation matrix from file AUS.COR

Asymptotic covariance matrix from file AUS.ACM

Means from file AUS.ME

Sample Size = 11642

Latent Variables GD KD

Relationships

ST18Q01 = CONST GD

ST18Q02 = CONST GD

ST18Q04 = CONST GD

ST18Q06 = CONST GD

ST18Q09 = CONST GD

ST18Q03 = CONST KD

ST18Q05 = CONST KD

ST18Q07 = CONST KD

ST18Q08 = CONST KD

ST18Q10 = CONST KD

Group Focal TÜRKİYE

Observed Variables

ST18Q01 ST18Q02 ST18Q03 ST18Q04 ST18Q05

ST18Q06 ST18Q07 ST18Q08 ST18Q09 ST18Q10

Correlation matrix from file TUR.COR

Asymptotic covariance matrix from file TUR.ACM

Means from file TUR.ME

Sample Size = 4535

Latent Variables GD KD

Relationships

GD = CONST

KD = CONST

Set the Error Variance of ST18Q01 Free

Set the Error Variance of ST18Q02 Free

Set the Error Variance of ST18Q03 Free

(Ek-19'un devamı)

Set the Error Variance of ST18Q04 Free  
Set the Error Variance of ST18Q05 Free

Set the Error Variance of ST18Q06 Free  
Set the Error Variance of ST18Q07 Free  
Set the Error Variance of ST18Q08 Free  
Set the Error Variance of ST18Q09 Free  
Set the Error Variance of ST18Q10 Free

Set the Variance of GD Free  
Set the Variance of KD Free  
Set the Covariance of GD - KD free

Method of Estimation Weighted Least Squares  
Options MI ND=3 SC  
End of Problem

**Ek 20 Avustralya ve Yeni Zelanda Karşılaştırması için Referans Değişkeni Belirleme Sonuçları**

	$\chi^2$	$\Delta\chi^2$	$\Delta sd$	DMF	$\lambda$
Temel Model (Sınırlandırılmış)	2016,278	-	-	-	-
<b>GD</b>					
<b>ST18Q01</b>	<b>2016,075</b>	<b>0,203</b>	<b>2</b>	<b>Yok</b>	<b>0,765</b>
ST18Q02	2012,812	3,466	2	Yok	0,846
ST18Q04	2008,506	7,772	2	Yok	0,732
ST18Q06	2011,163	5,115	2	Yok	0,874
ST18Q09	2013,993	2,285	2	Yok	0,719
<b>KD</b>					
ST18Q03	2011,309	4,969	2	Yok	0,665
ST18Q05	2014,255	2,023	2	Yok	0,895
ST18Q07	2013,652	2,626	2	Yok	0,924
<b>ST18Q08</b>	<b>2016,076</b>	<b>0,202</b>	<b>2</b>	<b>Yok</b>	<b>0,843</b>
ST18Q10	2008,112	8,166	2	Yok	0,851

*Not.* sd = serbestlik derecesi, DMF= değişen madde fonksiyonu, GD = fen bilimlerinin genel değeri, KD= fen bilimlerinin kişisel değeri. Referans değişkeni olarak belirlenen maddeler koyu renkte gösterilmiştir.

\* $p < 0,005$

**Ek 21 Avustralya ve ABD Karşılaştırması için Referans Değişkeni Belirleme Sonuçları**

	$\chi^2$	$\Delta\chi^2$	$\Delta sd$	DMF	$\lambda$
Temel Model (Sınırlandırılmış)	2062,042	-	-	-	-
GD					
ST18Q01	2060,253	1,789	2	Yok	0,771
<b>ST18Q02</b>	<b>2060,092</b>	<b>1,950</b>	<b>2</b>	<b>Yok</b>	<b>0,851</b>
ST18Q04	2060,410	1,632	2	Yok	0,748
ST18Q06	2034,971	27,071*	2	Var	0,888
ST18Q09	2047,096	14,946*	2	Var	0,735
KD					
ST18Q03	2033,704	28,338*	2	Var	0,692
<b>ST18Q05</b>	<b>2059,807</b>	<b>2,235</b>	<b>2</b>	<b>Yok</b>	<b>0,882</b>
ST18Q07	2049,095	12,947*	2	Var	0,913
ST18Q08	2051,514	10,528	2	Yok	0,844
ST18Q10	2058,804	3,238	2	Yok	0,842

*Not.* sd = serbestlik derecesi, DMF= değişen madde fonksiyonu, GD = fen bilimlerinin genel değeri, KD= fen bilimlerinin kişisel değeri. Referans değişkeni olarak belirlenen maddeler koyu renkte gösterilmiştir.

\* $p < 0,005$

**Ek 22 Avustralya ve Türkiye Karşılaştırması için Referans Değişkeni Belirleme Sonuçları**

	$\chi^2$	$\Delta\chi^2$	$\Delta sd$	DMF	$\lambda$
Temel Model (Sınırlandırılmış)	2622,489	-	-	-	-
GD					
ST18Q01	2599,306	23,183*	2	Var	0,775
ST18Q02	2482,412	140,077*	2	Var	0,821
ST18Q04	2564,916	57,573*	2	Var	0,711
<b>ST18Q06</b>	<b>2606,938</b>	<b>15,551*</b>	<b>2</b>	<b>Var</b>	<b>0,864</b>
ST18Q09	2370,708	251,781*	2	Var	0,748
KD					
ST18Q03	2524,947	97,542*	2	Var	0,687
ST18Q05	2601,585	20,904*	2	Var	0,868
<b>ST18Q07</b>	<b>2620,481</b>	<b>2,008</b>	<b>2</b>	<b>Yok</b>	<b>0,903</b>
ST18Q08	2605,546	16,943*	2	Var	0,839
ST18Q10	2465,824	156,665*	2	Var	0,814

*Not.* sd = serbestlik derecesi, DMF= değişen madde fonksiyonu, GD = fen bilimlerinin genel değeri, KD= fen bilimlerinin kişisel değeri. Referans değişken olarak belirlenen maddeler koyu renkte gösterilmiştir.

\* $p < 0,005$

**Ek 23 DMF Gösteren Maddeleri Belirlemede Kullanılan Serbest Temel Model  
LISREL-SIMPLIS Komut dosyası**

Group Reference AVUSTRALYA

Observed Variables

ST18Q01 ST18Q02 ST18Q03 ST18Q04 ST18Q05

ST18Q06 ST18Q07 ST18Q08 ST18Q09 ST18Q10

Correlation matrix from file AUS.COR

Asymptotic covariance matrix from file AUS.ACM

Means from file AUS.ME

Sample Size = 11642

Latent Variables GD KD

Relationships

ST18Q01 = CONST GD

ST18Q02 = CONST GD

ST18Q04 = CONST GD

ST18Q06 = CONST 1\*GD

ST18Q09 = CONST GD

ST18Q03 = CONST KD

ST18Q05 = CONST KD

ST18Q07 = CONST 1\*KD

ST18Q08 = CONST KD

ST18Q10 = CONST KD

Group Focal TÜRKİYE

Observed Variables

ST18Q01 ST18Q02 ST18Q03 ST18Q04 ST18Q05

ST18Q06 ST18Q07 ST18Q08 ST18Q09 ST18Q10

Correlation matrix from file TUR.COR

Asymptotic covariance matrix from file TUR.ACM

Means from file TUR.ME

Sample Size = 4535

Latent Variables GD KD

Relationships

ST18Q01 = CONST GD

ST18Q02 = CONST GD

ST18Q04 = CONST GD

ST18Q09 = CONST GD

ST18Q03 = CONST KD

ST18Q05 = CONST KD

(Ek-23'ün devamı)

ST18Q08 = CONST KD

ST18Q10 = CONST KD

GD = CONST

KD = CONST

Set the Error Variance of ST18Q01 Free

Set the Error Variance of ST18Q02 Free

Set the Error Variance of ST18Q03 Free

Set the Error Variance of ST18Q04 Free

Set the Error Variance of ST18Q05 Free

Set the Error Variance of ST18Q06 Free

Set the Error Variance of ST18Q07 Free

Set the Error Variance of ST18Q08 Free

Set the Error Variance of ST18Q09 Free

Set the Error Variance of ST18Q10 Free

Set the Variance of GD Free

Set the Variance of KD Free

Set the Covariance of GD - KD free

Method of Estimation Weighted Least Squares

Options MI ND=3 SC

End of Problem

**Ek 24 Madde İnceleme Formu**

Reviewer's name:

You will be asked to evaluate the comparability in meaning between the English form and the Turkish form. The items below were identified as functioning differently across groups. This means that Australian and Turkish examinees with the same personality/trait had a different likelihood of answering the item positively or negatively. Make sure that you understand the concept of differential item functioning (DIF) before completing these questionnaires. For each item, please answer the following questions:

1. Do you think this item is comparable in meaning between these two forms?
  - (1) No change in meaning
  - (2) Minor change in meaning
  - (3) Major change in meaning
  
2. Quality of translation
  - (1) Very good
  - (2) Acceptable
  - (3) Very poor
  
3. What is the reason for the DIF?
  - (1) Translation is not correct
  - (2) Words do not have the same level of difficulty
  - (3) There are differences in cultural relevance
  - (4) Others
  
4. Additional Comments

## ÖZGEÇMİŞ

### Kişisel Bilgiler

Adı Soyadı : Mustafa Asil  
Doğum Yeri ve Tarihi : Konya/Beyşehir - 23.07.1977

### Eğitim Durumu

Yüksek Lisans : The Ohio State University-2005  
School of Education  
Quantitative Research, Evaluation, and  
Measurement in Education (QREME)  
Lisans : Marmara Üniversitesi-1999  
Atatürk Eğitim Fakültesi  
Matematik Öğretmenliği  
Yabancı Dil : İngilizce (KPDS 98)

### İş Deneyimi

1999-2001 : Kartal Yakacık Lisesi - Matematik Öğretmeni  
2005-... : Öğrenci Seçme ve Yerleştirme Merkezi  
Test Araştırma Birimi- Araştırmacı

### İletişim

Adres : ÖSYM TAB Bilkent / Ankara  
Elektronik Posta : [asilmustafa@yahoo.com](mailto:asilmustafa@yahoo.com)  
Tarih : 25.01.2010

