

T.C.
ABANT İZZET BAYSAL ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
ANABİLİMDALI

PISA 2006 ÖĞRENCİ ANKETİ MADDE YANLILIĞININ
SIRALI LOJİSTİK REGRESYON VE POLY-SIBTEST
TEKNİKLERİ İLE TEST EDİLMESİ

GÖZDE SIRGANCI

ARALIK – 2012

T.C.
ABANT İZZET BAYSAL ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANABİLİMDALI

PISA 2006 ÖĞRENCİ ANKETİ MADDE YANLILIĞININ SIRALI
LOJİSTİK REGRESYON VE POLY-SIBTEST TEKNİKLERİ İLE
TEST EDİLMESİ

Yüksek Lisans Tezi

Hazırlayan
Gözde SİRGANCI

Danışman
Doç. Dr. Mehtap ÇAKAN

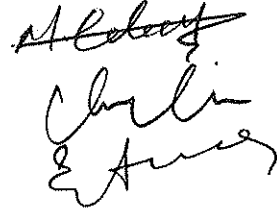
ARALIK - 2012

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ'NE,

Gözde Sırgancı'ya ait PISA 2006 Öğrenci Anketi Madde Yanlılığının Sıralı Lojistik Regresyon ve Poly-Sıbtest Teknikleri İle Test Edilmesi adlı çalışma, jürimiz tarafından Eğitimde Ölçme ve Değerlendirme Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir..../..../2012

Akademik Unvan ve Adı Soyadı

Üye (Tez Danışmanı) : Doç. Dr. Mehtap Çakan
Üye : Doç. Dr. Zekeriya Nartgün
Üye : Doç. Dr. Erkan Tekinarslan



Eğitim Bilimleri Enstitüsünün Onayı

Prof. Dr. Soner DURMUŞ
Enstitü Müdürü

BİLDİRİM

Yüksek lisans tezi olarak sunduğum, PISA 2006 Öğrenci Anketi Madde Yanlılığının Sıralı Lojistik Regresyon ve Poly-Sıbtest Teknikleri ile Test Edilmesi başlıklı çalışmanın yazılmasında, bilimsel ve etik kurallara uyulduğunu, başkalarının eserlerinden yararlanılması durumunda aitta bulunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, tezin tamamının ya da bir kısmının bu üniversite veya başka bir üniversitede bir tez çalışması olarak sunulmadığını beyan ederim. .../12/2012.

GÖZDE SIRGANCI

ABSTRACT

SIRGANCI, Gözde. *Testing the Item Bias of PISA 2006 Student Questionnaire by Using Sequential Logistic Regression and Poly-SIBTEST Methods*, Master Thesis, Bolu, 2012.

This study investigated whether student questionnaire of PISA 2006 show differential item functioning with respect to different cultural and linguistic characteristics (English-Turkish) with the samples of Australia, New Zealand, United States of America and Turkey on a comparative basis.

The sample of this study consisted of 4 countries selected among 57 countries participating in the PISA implementation and students selected among approximately 400000 students, namely 14170 students from Australia, 4823 students from New Zealand, 5605 students from United States of America, and 4942 students from Turkey. In this study, due to cultural and linguistic differences addressed as the possible causes of differential item functioning, the data of the 4 countries having similar and different cultural-linguistic characteristics was studied. Australia has been taken for reference and the countries were compared with the same language –similar culture (Australia- New Zealand), the same language –different culture (Australia –America) and different language-different culture (Australia- Turkey) .

In the first phase of the study, the factor structure of the student survey was examined by using a confirmatory factor analysis (CFA). In the second phase of the study, it was investigated whether questionnaire items show differential item functioning between different kültürse ant languraysa bay usan sequential logistic regression and poly-SIBTEST.

The results of the confirmatory factor analysis showed that the measurement model had the same factor structure in all cultures. The findings of sequential logistic regression and poly-SIBTEST revealed that there were items showing differential item functioning between the countries. While these methods showed consistency in determining differential item functioning, they differed in the levels of differential item functioning.

As a result of the analyses, it was observed that when linguistic and cultural differences increased among countries, the number of items showing differential item functioning increased, as well. In conclusion, the items may show differential item functioning in terms of cultural and linguistic differences. Thus, this fact should be taken in consideration when using these tests on different populations from different cultures and languages.

Keywords: PISA, differential item functioning, confirmatory factor analysis, sequential logistic regression, poly-SIBTEST

ÖZET

SIRGANCI, Gözde. *PISA 2006 Öğrenci Anketi Madde Yanlılığının Sıralı Lojistik Regresyon ve Poly- SIBTEST Yöntemleri İle Test Edilmesi*, Yüksek Lisans Tezi, Bolu, 2012.

Bu çalışmada, PISA 2006 kapsamında uygulanan öğrenci anketinin kültürel ve dilsel farklılıklara (İngilizce- Türkçe) göre değişen madde fonksiyonu gösterip göstermediği, Avustralya, Yeni Zelanda, Amerika Birleşik Devletleri ve Türkiye örneklemeleri üzerinde karşılaştırmalı olarak incelenmiştir.

Araştırmanın örneklemini, PISA uygulamasına katılan 57 ülke ve yaklaşık 400000 öğrenci içinden seçilen 14170 öğrenci ile Avustralya, 4823 öğrenci ile Yeni Zelanda, 5605 öğrenci ile ABD ve 4942 öğrenci ile Türkiye oluşturmaktadır. Çalışmada değişen madde fonksiyonunun olası nedenleri olarak kültürel ve dilsel farklılıklar ele alındığından dolayı benzer ve farklı kültürel - dilsel özelliklere sahip bu dört ülke verisi üzerinde çalışma yapılmıştır. Çalışmada Avustralya referans ülke alınmış ve diğer ülkeler ile aynı dil- benzer kültür (Avustralya- Yeni Zelanda, aynı dil-farklı kültür (Avustralya- Amerika) ve farklı dil-farklı kültür (Avustralya- Türkiye) karşılaştırması yapılmıştır.

Araştırmanın birinci aşamasında, öğrenci anketinin faktör yapısı doğrulayıcı faktör analizi yöntemiyle incelenmiştir. İkinci aşamada ise sıralı lojistik regresyon ve poly-SIBTEST analizleri ile anket maddelerinin farklı kültürler ve diller arasında değişen madde fonksiyonu gösterip göstermediği araştırılmıştır.

Doğrulayıcı Faktör Analizi sonuçları ölçme modelinin tüm kültürlerde aynı faktör yapısına sahip olduğunu göstermiştir. Sıralı Lojistik Regresyon ve Poly-SIBTEST bulguları ülkeler arasında farklı fonksiyonlaşan maddelerin olduğunu ortaya koymuştur.

Bu yöntemler değişen madde fonksiyonu belirlemede uyum gösterirken değişen madde fonksiyonu düzeylerinde farklılık göstermişlerdir.

Analizler sonucunda ülkeler arasında dilsel ve kültürel farklılıklar arttıkça değişen madde fonksiyonu gösteren maddelerin sayısının arttığı gözlenmiştir. Maddelerin kültürel ve dilsel farklılıklardan ötürü değişen madde fonksiyonu gösterebileceği sonucuna varılmıştır. Bu testlerin farklı kültür ve dillerdeki öğrencilere uygulanmasında madde yanlılığı olabileceği gerçeği gözönünde bulundurulmalıdır.

Anahtar kelimeler: PISA, değişen madde fonksiyonu, doğrulayıcı faktör analizi, ordinal lojistik regresyon, poly-SIBTEST

TEŞEKKÜR

Tez çalışma sürecimde çalışmalarına destek olan danışmanım Doç. Dr. Mehtap Çakan'a, çalışmalarımın sonuçlandırılmasında görüşlerine başvurduğum jüri üyeleri Doç Dr. Zekeriya Nartgün ve Do. Dr. Erkan Tekinarslan'a, bu uzun süreçte her zaman en büyük destekçim ve güç kaynağım olan annem Lerzan Sırgancı, babam Kubilay Sırgancı ve ablam Seda Sırgancı olmak üzere aileme teşekkür ederim.

Ayrıca, tezimde kullandığım bazı programları elde etme ve çalıştırma aşamalarında desteğini aldığım Kübra Atalay ve Mustafa Asil'e, çeviri konusunda yardımlarını esirgemeyen arkadaşlarım Berkan Çelik'e ve Nihal Yaman'a, teknik ve manevi desteğini aldığım arkadaşım Raşit Köşlü'ye, bölüm arkadaşım Alperen Yandı'ya ve bana destek olan bütün arkadaşlarıma teşekkür ederim.

İÇİNDEKİLER

BİLDİRİM.....	IV
ABSTRACT.....	V
ÖZET.....	VII
TEŞEKKÜR.....	IX
İÇİNDEKİLER DİZİNİ.....	X
TABLolar DİZİNİ.....	XIII
ŞEKİLLER DİZİNİ.....	XV
KISALTMALAR.....	XVI
BÖLÜM I.....	1
1. Giriş.....	1
1.1. Problem Durumu.....	1
1.1.1. Test ve madde yanlılığı.....	4
Yanlılık nedir? Ne değildir?	6
1.1.2. Değişen madde fonksiyonu (DMF)	9
Tek biçimli DMF (TB-DMF)	11
Tek biçimli olmayan DMF (TBO-DMF).....	12
Dışsal yöntemler	13
İçsel yöntemler.....	15
Değişen madde fonksiyonunun olası sebepleri.....	15
1.1.3. Pisa ve Pisa öğrenci anketi.....	16

1.1.4. Değişen madde fonksiyonunu belirlemede kullanılan istatistiksel yöntemler	20
Çok kategorili verilerde değişen madde fonksiyonu belirleme yöntemleri	22
Sıralı lojistik regresyon (OLR)	23
Poly- SIBTEST	26
1.2. Problem İfadesi	27
1.3. Araştırmanın Önemi	28
1.4. Araştırmanın Amacı	29
1.5. Araştırma Soruları	29
1.6. Sayıtlılar	30
1.7. Sınırlılıklar	30
1.8. Tanımlar	30
BÖLÜM II	32
2. Literatür Taraması	32
BÖLÜM III	44
3. Yöntem	44
3.1. Araştırmanın Türü	44
3.2. Evren ve Örneklem	44
3.3. Veri Toplama Araçları ve Verilerin Elde Edilmesi	45
3.4. Verilerin Çözümlemesi	47
BÖLÜM IV	50
4. Bulgular ve Yorum	50
4.1. DFA sayıtlılarının test edilmesi	50
4.2. Araştırma Sorusu 1' e İlişkin Bulgular ve Yorumlar	51
4.3. Araştırma Sorusu 2'ye İlişkin Bulgular	58
4.3.1. Araştırma sorusu 2a'ya ilişkin bulgular	58
4.3.2. Araştırma Sorusu 2b'ye ilişkin bulgular	62

4.3.3. Araştırma Sorusu 2c'ye ilişkin bulgular	66
BÖLÜM V	70
5.Sonuç ve Öneriler.....	70
5.1.Sonuçlar	70
5.1.1. Araştırma sorusu 1'e ilişkin sonuçlar	70
5.1.2. Araştırma sorusu 2'ye ilişkin sonuçlar	71
Araştırma sorusu 2a'ya ilişkin sonuçlar.....	71
Araştırma sorusu 2b'ye ilişkin sonuçlar	72
Araştırma sorusu 2c'ye ilişkin sonuçlar.....	72
5.2. Öneriler	74
5.2.1. PISA uygulayıcılarına yönelik öneriler	74
5.2.2. Araştırmacılara yönelik öneriler	75
KAYNAKLAR	77
EKLER	87

TABLOLAR DİZİNİ

Tablo 1-1. Pısa 2006 Fen Bilimlerine Yönelik Tutum.....	19
Tablo 1-2. Çok Kategorili Verilerde Dmf Belirleme Yöntemleri.....	22
Tablo 1-3. ΔR^2 Değerini Yorumlama Ölçütleri	25
Tablo 1-4. B Değerlerini Yorumlama Ölçütleri	27
Tablo 3-1. Çalışma Grubuna Ait Frekans Ve Yüzdelere	45
Tablo 4-1. Fen Bilimlerinin Genel Değeri Faktörünü Oluşturan Maddelere Ait Betimsel İstatistikler	52
Tablo 4-2. Fen Bilimlerinin Kişisel Değeri Faktörünü Oluşturan Maddelere Ait Betimsel İstatistikler	53
Tablo 4-3. Bilimsel Sorgulamaya Verilen Destek Ölçme Modeli Uyum Değerleri	54
Tablo 4-4. Fen Bilimlerinin Genel Değeri ve Fen Bilimlerinin Kişisel Değeri Faktörlerinin Güvenirlikleri	55
Tablo 4-5. Fen Bilimlerinin Genel Değeri Faktörünü Oluşturan Maddelerin Güvenirlikleri.....	56
Tablo 4-6. Fen Bilimlerinin Genel Değeri Faktörünü Oluşturan Maddelerin Güvenirlikleri.....	57
Tablo 4-7. Avustralya Ve Yeni Zelanda Örneklemelerinde Orf Yöntemi ile Yapılan Dmf Sonuçları	59

Tablo 4-8. Avustralya ve Yeni Zelanda Örneklemelerinde Poly-Sıbtest Yöntemi İle Yapılan Dmf Sonuçları	60
Tablo 4-9. Avustralya ve Yeni Zelanda Örneklemeleri Arasında Orl ve Poly-Sıbtest Analizlerinin Karşılaştırılması.....	61
Tablo 4-10. Avustralya ve ABD Örneklemelerinde Orl Yöntemi ile Yapılan Dmf Sonuçları.....	63
Tablo 4-11. Avustralya ve ABD Örneklemelerinde Poly-Sıbtest Yöntemi ile Yapılan Dmf Sonuçları.....	64
Tablo 4-12. Avustralya ve ABD Örneklemeleri Arasında Orl ve Poly-Sıbtest Analizlerinin Karşılaştırılması.....	65
Tablo 4-13. Avustralya ve Türkiye Örneklemelerinde Orl Yöntemi ile Yapılan Dmf Sonuçları	67
Tablo 4-14. Avustralya ve Türkiye Örneklemelerinde Poly-Sıbtest Yöntemi ile Yapılan Dmf Sonuçları.....	68
Tablo 4-15. Avustralya ve ABD Örneklemeleri Arasında Orl ve Poly-Sıbtest Analizlerinin Karşılaştırılması.....	69
Tablo 5-1. Kültür ve Dil Farklılıklarına Göre Orl ve Poly-Sıbtest Sonuçları.....	74

ŞEKİLLER DİZİNİ

Şekil 1-1. Dmf Bulunmayan Madde.....	10
Şekil 1-2. Erkek Ve Kız Öğrencilerin Farklı Yetenek Düzeyindeki Maddeyi Doğru Yanıtlama Olasılıkları.....	11
Şekil 1-3. Tek Biçimli Dmf Bulunan Madde Grafiği.....	12
Şekil 1-4. Erkek Ve Kız Öğrencilerin Farklı Yetenek Düzeyindeki Maddeyi Doğru Yanıtlama Olasılıkları.....	12
Şekil 1-5. Tek Biçimli Olmayan Dmf Bulunan Madde Grafiği	13
Şekil 1-6. Yanlılık Gösteren Madde	14
Şekil 1-7. Yansız Madde	14

KISALTMALAR

- PISA: Uluslararası Öğrenci Değerlendirme Programı
TIMSS: Üçüncü Uluslararası Fen Bilgisi ve Matematik Çalışması
PIRLS: Uluslararası Okuma Becerilerinde Gelişim Projesi
OECD: Ekonomik İşbirliği ve Kalkınma Örgütü
DMF: Değişen madde fonksiyonu
DFA: Doğrulayıcı Faktör Analizi
LR: Lojistik regresyon
SLR: Sıralı lojistik regresyon
M-H: Mantel Haenszel
GMH: Genelleştirilmiş Mantel Haenszel
LLM: Log-lineer model
LDFA: Lojistik Diskriminant Fonksiyon Analizi
MTK: Madde Tepki Kuramı
KTK: Klasik Test Teorisi
AUS: Avustralya
ABD: Amerika Birleşik Devletleri
NZL: Yeni Zelanda
TUR: Türkiye
TBO: Tek Biçimli Olmayan
TB: Tek Biçimli

BÖLÜM I

1. Giriş

Bu bölümde araştırmanın; problem durumu, problem ifadesi, araştırma soruları, sınırlılıkları, sayıltıları, tanımlar ve kısıtlamaları, amacı ve önemine yer verilmiştir.

1.1. Problem Durumu

Günümüzün küreselleşen dünyasında bilginin niteliği sürekli olarak değişmekte, bilgiye duyulan ihtiyaç giderek artmaktadır. İnternet ve diğer iletişim teknolojilerinin öngörülemez biçimde ilerlemesi, ulaşım olanaklarının artması ve ucuzlaması, 21. yüzyılda dünya vatandaşı olmak kavramını ortaya çıkarmıştır (EARGED, 2011). Bir ülkenin eğitim sistemi, vatandaşlarının eğitim düzeyi ve niteliği ulusal gelişmişliğin bir göstergesidir. Eğitim, sadece bir ülkenin sosyal ve ekonomik yönlerden gelişmesinin temeli olmayıp toplumun ve insanların hayat standardının yükselmesinin de temel belirleyicisidir (EARGED, 2011). Bunun bilincinde olan ülkeler eğitime daha çok kaynak ayırmaya çalışmakta, insan kaynaklarına daha çok yatırım yapmaktadır. İnsan kaynakları açısından yetişmiş elemanı olan, bilgi ve teknoloji üreten ülkeler ayakta kalmakta, diğerleri ise bu ülkelerin takipçisi ve pazarı konumuna düşmektedir (EARGED, 2011). Kamu bütçesinden en çok kaynak ayrılan alanlardan biri olan eğitim sisteminin performansının değerlendirilmesi ve buna bağlı düzeltici önlemlerin alınması gerekliliği ülkeleri hangi tür eğitim programlarının, daha yüksek öğrenci başarısını sağladığı konusunda veriler sağlama arayışına sokmuş ve TIMSS-Uluslararası Matematik ve Fen Eğilimleri Çalışması (Trends in International Mathematics and Science Study), PISA (Programme for International Student Assessment) gibi uluslar arası sınavlar yapılmaya başlanmıştır. Bu sınavlarda toplanan

veriler, ülkelerin öğrencilerinin belli alanlardaki başarı durumlarını, öğretim programlarını, öğrenci, öğretmen özelliklerini değerlendirmelerine ve eğitim sistemleri hakkında bilgi sahibi olmalarına olanak tanır. Öte yandan ülkeler, kendi eğitim sistemlerini sorgular, eksiklikleri belirler, gerekli düzeltme ve yenilikleri yapar ve diğer ülkelerin eğitim sistemleri ile kendi eğitim sistemlerini karşılaştırabilirler (EARGED, 2011).

Sonuçlarına bakılarak önemli kararların alındığı bu sınavlarda kullanılan testlerin veya anketlerin farklı dillerdeki versiyonlarının eşdeğerliğinin sağlandığından emin olunmalıdır. Kullanılan her bir maddenin bir kültürdeki birey için verdiği anlam ile diğer kültürlerdeki bireyler için verdiği anlam aynı olmalıdır (Hui ve Triandis, 1985).

Farklı kültür ve dilden bireyleri karşılaştırma çalışmalarında, karşılaştırmanın anlamlı olabilmesi için testlerin ölçtüğü yapıların eşdeğer olması gerekir. Bu tür çalışmalarda ölçme eşdeğerliğinin sağlanması geçerliğin temel bir sayıtlısıdır (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 (Aktaran: Asil, 2010). Bu kanıtlar ölçme araçlarının geçerlik çalışmalarıdır.

Geçerlik, bir testin ölçmek istediği özelliği başka değişkenlerle karıştırmadan ölçmesidir (Mellor, 1995). Bazı durumlarda bireylerin ölçmek istediğimiz özelliklerine başka değişkenler karışabilir. Bunlar; dil, kültürel farklılıklar, cinsiyet, yaş, okul türü, sosyo-ekonomik düzey, etnik köken gibi değişkenlerdir. Bu değişkenlerin test puanlarına etkisi geçerlik üzerinde bir tehdittir ve test puanlarının yanlı olmasına yol açar (Camilli, 2006; Camilli ve Shepard, 1994). Bu durum eğitimde ve psikolojide ölçme standartlarından 3.10 numaralı standartta açıkça belirtilmiştir:

“Daha önceki araştırmalar, test uygulananlar evrenindeki farklı yaş, etnik köken, kültürel yapı ve cinsiyet grupları için belirli bir testteki madde veya test

performansı farklılıklarını inceleme ihtiyacını işaret ediyorsa mümkün olduğu oranda bu tür çalışmalar yapılmalıdır. Bu tür araştırmalar, belirli bir grup için test puanlarında yanlılığa yol açabilen test tasarımı, kapsamı veya biçimini saptayıp, bunları ortadan kaldırmak için düzenlenmelidir” (Hovardaoğlu ve Sezgin, 1998: 34).

Buna göre aynı yetenek düzeyindeki bireyler için testte kullanılan maddelerin doğru yanıtlanma olasılığı bireylerin cinsiyeti, yaşı, sosyoekonomik durumu, etnik kökeni, okuduğu okul türü ve yaşadığı bölge gibi değişkenlerden bağımsız olarak farklılık göstermemeli yani yanlı olmamalıdır. Yanlılık, ölçme sonuçları üzerinde belirli bir gruba karşı sistematik hata olarak tanımlanmaktadır (Camilli, 2006; Camilli ve Shepard, 1994). Dolayısıyla yanlılık bir geçerlik sorunudur. Madde yanlılığı kavramını test geçerliği içinde ele alan araştırmacılar madde yanlılığını aynı yetenek düzeyinde olan fakat farklı gruplardan gelen iki kişinin bir maddeyi doğru cevaplama olasılıklarının aynı olmaması durumu olarak tanımlamışlardır (Osterlind, 1983; Adam ve Rowe 1988; Tittle, 1988; Cole ve Moss, 1989; Hambleton ve Rogers, 1996; Zumbo, 1999 aktaran: Deveci, 2008). Geçerlik çalışmaları kapsamında ele alınan madde yanlılığı sebebiyle geçerliği düşük olan testlerin sonuçlarına dayanarak alınacak kararların güvenilirliğinin de düşük olacağı bir gerçektir. Kristanjansson, Aylesworth, McDowell ve Zumbo, (2005) bir testin tarafsızlığı ve geçerliliği için en büyük tehdidin yanlılık olduğunu ortaya koymuşlardır. Yanlılık çalışmalarında öncelikle, alt grupların maddeleri yanıtlama yapıları arasında farklılık olup olmadığını belirlemeye yönelik istatistiksel analizler uygulanmalıdır. Bir maddenin yanlı olduğuna karar verebilmek için öncelikle o maddenin farklı fonksiyonlaşması gerekir. Daha sonra, maddenin farklı fonksiyonlaşmasının olası sebepleri tespit edilerek, maddenin ölçülmek istenen yapıdan ilişkisiz olarak bazı alt gruplar için adil olmayan bir avantaj sağlayıp sağlamadığının uzman kişiler tarafından belirlenmesi gerekir (Camilli ve Shepard, 1994; Zumbo, 1999 aktaran: Kalaycıoğlu, 2008). Yanlılığı belirlemek için kullanılan tekniklerden değişen madde fonksiyonları (DMF) tekniği psikometrik yanlılık analizlerinde bir standart haline gelmiştir. Değişen madde fonksiyonu ve madde yanlılığı belirleme ile ilgili yöntemler, yeni bir ölçme aracı geliştirildiğinde ya da var olan bir ölçme aracının farklı diller ve kültürler için uyarlama çalışmalarında madde analizi işlemlerini yapmak için

kullanılır (Zumbo, 2007, aktaran: Atalay,2010). Test ve testteki maddeler için DMF belirleme çalışmaları son yıllarda yaygınlaşmakta, bu çalışmalardan elde edilen bilgilerden madde yanlılığını belirlemede yararlanılmaktadır. DMF belirleme analizleri standart başarı testlerinden, tutumların ölçülmesine değin birçok uygulamada yanlılığı belirlemek adına kullanılan bir yöntem haline gelmiştir (Clauser ve Mazor, 1998; Millsap ve Everson, 1993). TIMSS, PISA ve PIRLS gibi çalışmalarda da öğrenci, öğretmen ya da veli anketi gibi ölçme araçları uygulanmakta ve özellikle öğrenci tutum, davranış ve değer yargıları önemli öğrenme çıktıları olarak görülmektedir. Uluslararası karşılaştırmalarda daha çok bilişsel testler için yanlılık çalışmaları yapılmaktadır. Anketlerin yanlılığı ile ilgili araştırmalar ise yetersiz görünmektedir (Schulz, 2003, 2005, 2008). Bu tür kategorik verilerde DMF belirleme ile ilgili birçok yöntem geliştirilmiştir ancak bu yöntemler henüz yeterince test edilmemiştir (Zumbo ve McDowell, 2005).

Bu çalışma kapsamında çok kategorili verilerin değişen madde fonksiyonunu belirlemede Sıralı Lojistik Regresyon ve Poly-SIBTEST teknikleri karşılaştırılmaktadır. Değişen madde fonksiyonu analizi PISA 2006 öğrenci anketi kapsamında yer alan fen bilimlerine yönelik tutum ölçeğinin bir alt boyutu olan, bilimsel sorgulamaya verilen destek maddeleri üzerinde Avustralya, Yeni Zelanda, ABD ve Türkiye örneklemi karşılaştırılarak yürütülecektir. Maddelerin DMF analizi yapılmadan önce her bir kültürde ve dilde aynı yapının ölçüldüğünün doğrulanması gerekmektedir (Asil, 2010). Ölçülen yapıların benzerliğini göstermek amacı ile her bir örnekleme doğrulayıcı faktör analizi uygulanmış ve daha sonra madde bazında DMF analizi yapılmıştır.

1.1.1. Test ve madde yanlılığı

Madde ve test yanlılığı araştırmaları 1910'lu yıllarda Alfred Binet'in düşük sosyoekonomik tabakadan gelen çocukları test etmesiyle başlamıştır (Camili ve Shepard, 1994). Binet, bazı zeka testi maddelerinin çocukların zeka kapasitelerinden ziyade evde veya okuldaki kültürel eğitim öğretimlerinin etkileriyle ilişkili olduğunu saptamıştır. Bu saptamadan sonra kültürel yanlılık taşıyan; örneğin yedi yaşındaki çocuklar için, parmaklarını sayma, söylenen cümleleri yazma ve paraları adlandırma ile

ilgili bazı maddeleri testten çıkarmıştır. Çünkü böyle maddelerin dil, dikkat, evde alınan eğitim veya eğitsel alıştırmaya bağlı olduğunu düşünmüştür. Sonuçta bu ve benzeri maddelerin testten çıkarılması tüm test puanlarındaki farkı azaltmıştır (Jensen,1980, aktaran: Bekçi, 2007). Benzer olarak, zekâyı kronolojik yaşa göre tanımlayan William Stern (1912), Almanya'daki sınıf farklılıklarını araştırmıştır. Stern'de Binet gibi, belirli farklılıkları gözlemlemiş; farklılıkların kaynağını anlamak için yaptığı çalışmaların doğrultusunda testleri “bir sınıfı diğerine tercih etme” olarak tanımlamış ve testlerin tek bir gruba yönelik hazırlanmasını tavsiye etmiştir (Camilli ve Shepard, 1994, Stern, 1914).

1930'lu yıllarda zekâ testleri sonuçlarının dil ve kültüre; 1940'ların sonu 1950'lerin başlarında ise ırklara göre farklı sonuçlar vermesi test yanlılığı konusunu daha önemli kılmıştır (Tittle, 1988, aktaran: Deveci, 2008).

1951 yılında, Eels, Davis, Havrghurst, Herrick ve Tyler tarafından başlatılan çalışmalar test yanlılığı konusundaki ilk araştırmalardandır. Binet ve Stern'den itibaren birçok araştırmacı zekâ ve sosyal sınıf arasındaki ilişkiyi belirlemiş olmalarına rağmen, bilim adamları yetenekte gözlemlenen farklılıkların sebebi olarak genetik ve çevre etkisi üzerinde yoğunlaşmışlardır (Camilli ve Shepard, 1994). Eells ve arkadaşları, madde analizlerinde doğru yanıt oranlarını grup farklılıklarını belirlemek için kullanmışlar ancak bu, daha sonra önerilen bir yöntem olmamıştır. Bununla birlikte, onlar IQ testlerindeki maddelere daha iç görüsel sorular sormuşlardır: Grup farklılıkları madde boyunca sabit midir ya da içeriğine göre değişmekte midir? Grup farklılığını artırma ya da azaltma eğiliminde olan soru formları var mıdır? Maddelerin olası güçlüklerini açıklayacak yanlış yanıtların seçimi kültürel farklılıklarla mı ilgilidir? Böylece Eells ve arkadaşları ilk kez, sosyal katmanlardaki etnik grupları karşılaştırarak test performansında yer alan etnik ve sosyal sınıf etkisini ortaya çıkarmışlardır (Camilli,1993, aktaran: Bekçi, 2007).

1960'lı yılların ortalarından itibaren yanlılık araştırmalarına verilen değer artmış ve bu konuyla ilgili daha fazla araştırma yapılmaya başlanmıştır (Cole ve Moss 1989). 1970'li yıllarda eğitim eşitliği üzerine yapılan tartışmalar, özel eğitime

gereksinim duyan bireyler hakkında verilen kararlar, birçok alanda kullanılan zekâ testlerinin sorgulanmasına ve testlerin istenilen performans dışında başka etkenlerden etkilenip etkilenmediğini belirlemek amacıyla test yanlılığı çalışmalarına ağırlık verilmiştir. O dönemde test maddelerinin yanlılığına yönelik yapılan araştırmalar testlerin geçerliği konusu kapsamında ele alınmaya başlanmış bu da geçerliğin sınırlarını genişletmek anlamına gelmiştir (Camilli ve Shepard, 1994). 1985 yılında ise Amerikan Psikologlar Derneği tarafından büyük ölçekli ölçme ve değerlendirmelerde test ve madde yanlılığı çalışmaları, Eğitim ve Psikolojik Testlerin Standartları'nda (Standards for Educational and Psychological Testing-AERA, APA NCME, 1985) temel gereklilik olarak kabul edilmiştir (Aktaran: Deveci, 2008).

Cole ve Moss (1989) yaptıkları bir araştırmada son yıllarda Madde Tepki Kuramı'na bağlı yöntemlerin madde yanlılığı belirlemede daha fazla kullanılmakta olduğunu; Narayanan ve Swaminathan (1993) ise son on yılda yetenek ve başarı belirleme sınavlarında madde yanlılığını belirlemek için modern ve klasik test kuramlarına bağlı Mantel-Haeszal, doğrudan madde yanlılığı (Simultaneous Item Bias - SIB) ve lojistik regresyon gibi çeşitli istatistiksel yöntemler geliştirildiğini belirtmişlerdir. Kristjansson ve diğerleri de (1997) madde yanlılığının ölçmelerin geçerliliğini tehdit eden önemli bir etken olduğunu; test ve madde yanlılığını belirleme yöntemlerinin sıklıkla kullanılması gerektiğini dile getirmişlerdir.

Küreselleşen dünyada değişen yaşam koşulları çerçevesinde bireyleri tanımak, onlar hakkında geçerli ve güvenilir kararlar almak yani ölçme işlemini adil yapmak çabası testlerde madde yanlılığını belirleme gerekliliğini doğurmuştur. Böylelikle araştırmacılar yanlılıkla ilgili çalışmalara ağırlık vermeye başlamışlardır.

Yanlılık nedir? Ne değildir?

Test veya ölçek puanlarına test ile ölçülen psikolojik özellikten başka varyans kaynaklarının karışması olağandır. Bundan ötürü test puanlarının etkilendiği varyans kaynaklarının tüm gruplara mı etki ettiği yoksa alt popülasyonları farklı biçimde mi etkilediği yanlılık kavramının ortaya koyulması için önemli bir husustur. Yanlılık, testin

amacına uygun olmayan test koşullarından ya da test maddelerinin özelliklerinden dolayı, bir maddeyi doğru yanıtlama olasılığının aynı yetenek düzeyindeki iki grupta farklı olmasıdır (Zumbo, 1999). Yani sistematiktir; çünkü alt popülasyonları farklı biçimde etkileyen varyans kaynaklarının var olmasından kaynaklanır. Bu durum testi alan belirli bir grubun puanlarında fark yaratır. Camili ve Shepard (1994) test yanlılığını, “geçersizlik ya da testin uygulandığı tüm bireylerden sadece belirli bir grup için “sistemik hata” ya sahip ölçümler içermesi halidir” biçiminde tanımlamışlardır. Tesadüfi hatalar ise kaynağı kesin olarak bilinmeyen ölçmeyi yapanın dikkatsizliği, ölçme ortamı gibi faktörlerin sebep olabileceği bir hata türüdür. Bu hatalar tüm grupların bütün üyelerini eşit bir şekilde etkilediği için yanlılık tesadüfi hataların bir sonucu değildir.

Öte yandan; “test sonuçlarındaki grup farklılıkları başarı veya yetenekteki gerçek farklılıktan mı yoksa başka değişkenlerin yarattığı farklılıktan mı kaynaklanmaktadır?” sorusu madde etkisi ve madde yanlılığı arasındaki farkı gündeme getirmektedir. Madde etkisi, farklı gruplardan gelen cevaplayıcıların bir maddeye doğru cevap verme olasılıklarının ölçülen özellikteki gerçek farklılıktan ileri gelmesi durumudur (Ackerman, 1992; Dorans ve Holland, 1993). Gruplar arasında farklılığın var olması, yanlılığın olduğu anlamına gelmez; çünkü var olan farklılık gerçek yetenek farkı olabilir (Camilli, 2006). Bu farklılık, bir maddenin ölçmeye çalıştığı yetenek düzeyinde var olan farklılıktır. Bu durumda madde etkisi olduğunda yanlılıktan söz edilemez. Görülen o ki; yanlılık araştırmalarının amacı, grup farklılıklarının sebeplerinin gerçek (gerçek performans farklılıklarının doğru yansıtılması testten ziyade diğer faktörlere bağlı olması) veya yapay (ölçme sürecinin kendisinden kaynaklanması) olup olmadığını ortaya çıkarmaya çalışmaktır (Wigdor, 1982; Wigdor & Garner, 1982, aktaran Bekçi, 2007). Yukarıda açıklanan bilgiler doğrultusunda yanlılık testlerin geçerliği, ölçme hataları ve farklı varyans kaynaklarının belirlenmesi kapsamında ele alınabilir.

Testin ölçmeyi amaçladığı özellik bakımından, aynı yetenek düzeyinde olup, farklı gruplardan gelen bireyler tarafından maddelerin aynı olasılıkla doğru cevaplanabilmesi beklenir. Yani; aynı yetenek düzeyinde olup, farklı alt gruplarda

bulunan cevaplayıcılar için bazı maddelerin doğru cevaplanma olasılıklarının daha yüksek veya düşük olması durumu, madde yanlılığı olarak tanımlanmaktadır (Camilli ve Shepard, 1994; Glas,1998; Zumbo, 1999, aktaran: Deveci, 2008). Bu yönüyle madde yanlılığı testlerin geçerlik çalışmaları kapsamında ele alınmaktadır. Madde yanlılığı kavramını ölçme hataları kapsamında ele alan araştırmacılar ise, test maddelerinin gözlenen değerleri ile gerçek değerleri arasındaki farklılığı madde yanlılığı olarak tanımlamışlar ve bu farkın sistematik hata olarak ölçme sonuçlarına karıştığını böylece testin yapı ve yordama geçerliliğini olumsuz yönde etkilediğini belirtmişlerdir (Childs, 1990). Madde yanlılığı kavramını test puanlarının değişik alt gruplardaki farklı varyans kaynaklarından etkilenip etkilenmediğini belirleme olarak ele alan Cromwell (2002) ise farklı varyans kaynaklarının alt gruplardan biri için avantaj olabileceğini belirtmiştir.

Camilli ve Shepar (1994) test yanlılığının daha çok ırksal ve etnik grup farklılıkları ve cinsiyet farklılıkları için kullanılan bir kavram olduğunu; bununla birlikte test ölçümlerindeki kırsal çevreye karşı kentsel çevre, ırk, yaş, sosyal çevre tarafından tanımlanan alt gruplar için bir konu olabileceğini söylemişlerdir. Tittle (1998), ise test yanlılığını, aynı testi alan farklı cinsiyet veya sosyo-kültürel özellik gösteren alt gruplara haksızlık yapmak biçiminde tanımlamıştır. Bu tanımlar doğrultusunda bir ölçme aracının karşılaştırma gruplarından herhangi birine karşı taraf olup olmadığı belirlenmemişse ve bazı alt grup özelliklerinden dolayı test sonuçlarında gruplar arası farklılık çıkıyorsa; bu durum bazı grupların avantajı olurken bazılarının dezavantajı olacaktır. Bu da toplumun bu tür ölçme araçlarının kullanımına ilişkin güvenilirlik ve geçerlik algılarının zedeleyecektir. Bu durumu engellemenin yolu özellikle standart testlerde testin bir gruba karşı üstünlük sağlamadığına dair kanıtlar sunmaktan geçer. Bunun için çeşitli istatistiksel yöntemler geliştirilmektedir. Madde yanlılığını belirlemede değişen madde fonksiyonu tekniği sıklıkla kullanılmaktadır.

1.1.2. Değişen madde fonksiyonu (DMF)

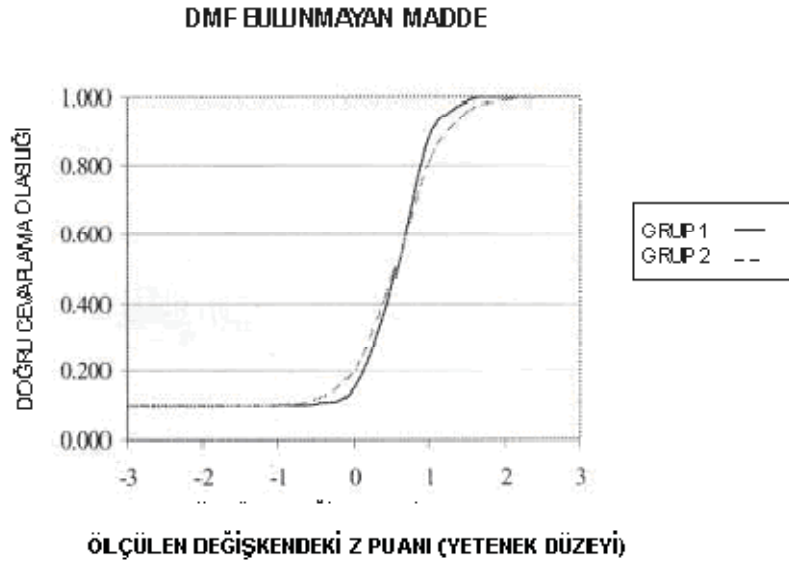
Madde yanlılığı belirlemek için kullanılan istatistiksel yöntemler aynı yetenek düzeyine sahip ancak farklı gruplarda bulunan bireylerin bir maddeye doğru cevap

verme olasılıklarının değişimini ortaya koyar. Ancak bu değişimin madde etkisinden mi yoksa madde yanlılığından mı kaynaklandığı içerik analizi veya uzman kanısı ile ortaya konmalıdır. Bu anlamda madde yanlılığı belirlemek için kullanılan istatistiksel yöntemler tek başına, yanlılığın olduğunu ortaya koymaz. Maddelerin değişen madde fonksiyonunu belirleme, yanlılık için daha objektif bir yaklaşım olarak görüldüğünden sık kullanılan bir tekniktir ve birçok kaynakta benzer bir şekilde tanımlanmıştır. Bu tanımlar ışığında DMF, aynı yetenek düzeyinde olan, fakat cinsiyet, sosyoekonomik düzey, etnik köken, inanç gibi farklı gruplardan gelen bireylerin, test maddelerine doğru cevap verme olasılıklarının değişmesi biçiminde sentezlenebilir (Rodney & Drasgow, 1990; Raju, 1990; Mellenberg, 1989; Tittle, 1988; Adams & Rowe, 1988; Shepard, Camilli, & Williams, 1985; Mellenberg, 1983; Osterlind, 1983; Devine & Raju, 1982, aktaran: Doğan ve Öğretmen, 2008: 102).

Zumbo (1999) değişen madde fonksiyonunu, maddenin ölçmek istediği yetenek düzeyinde yapılacak bir karşılaştırmada farklı gruplardaki yanıtlayıcıların ilgili maddeyi doğru yanıtlama olasılıklarındaki farklılık olarak ifade etmiştir. Yani yetenek düzeyleri sabit tutulduktan sonra iki demografik grup üyesinin bir maddede gösterdikleri performans arasında farklılık bulunuyorsa bu durumda DMF ortaya çıkar (Dorans ve Holland, 1993). Bu farklılığın iki temel kaynağı vardır. Bunlar madde yanlılığı veya gerçek bilgi, beceri farklılığıdır. Camili ve Shepard, (1994) DMF'nin ortaya çıkmasının iki olası nedeni olduğunu ve bunların alt gruplar arasındaki gerçek farklılık ve madde yanlılığı olduğunu ortaya koymuşlardır. Bir maddenin DMF göstermesi o maddenin yanlı olduğunu belirtmek için yeterli değildir çünkü Camili (2006), var olan farklılığın gerçek yetenek farkı olabileceğini söylemiştir. Ancak yanlı olan bir madde DMF gösterecektir. Görüldüğü üzere madde yanlılığı kavramsalken DMF istatistiksel bir yöntemdir ve madde yanlılığının bir indeksidir. Dolayısıyla DMF gösteren maddeler incelenerek maddelerin yanlı olup olmadığına karar verilir (Camili ve Shepard 1994; Clauser ve Mazor, 1998).

Bir madde DMF içeriyorsa bu durumun madde yanlılığından mı yoksa madde etkisinden mi kaynaklandığı araştırılmalıdır. Ancak ilgili madde DMF içermiyorsa yansız bir maddedir ve bu durum ölçülmek istenen yetenek konusunda gruplar arasında

fark olmadığını gösterir. Şekil 1-1' de DMF içermeyen bir madde örneği verilmiştir. Görüldüğü üzere eğriler arasındaki alan çok küçük ve her bir eğrinin parametreleri neredeyse birbirine eşittir (Zumbo, 1999).

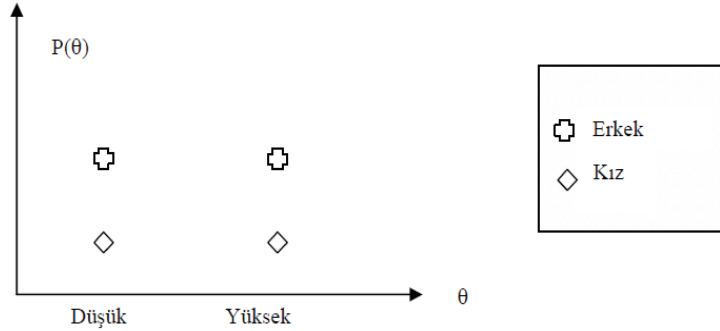


Şekil 1-1. DMF bulunmayan madde (Zumbo, 1999:19).

DMF aynı yeteneğe sahip ancak farklı gruplarda yer alan bireylerin ilgili maddeyi doğru yanıtlama olasılıklarındaki farklılığı işaret etmektedir. Bu farklılık, öğrencilerin yetenek, test puanları gibi içsel bir ölçüt ile karşılaştırılarak ortaya çıkarılmaktadır (Yurdugül, <http://yunus.hacettepe.edu.tr>, 21 Aralık 2011'de erişildi). Zumbo'ya (1999) göre yanlılığın varlığı yargı ve istatistiksel olmak üzere iki yöntemle araştırılabilir. İstatistiksel yöntemlerde, bireyler genellikle referans ve odak grup olarak ikiye ayrılır. Daha sonra odak ve referans grupta bulunan bireyler ölçülen yetenek düzeyinde eşleştirilir ve maddeler üzerindeki grup farklılıklarını belirlemek için istatistiksel süreçler uygulanır (Aktaran: Atalay, 2010). Bu grup farklılıkları iki farklı şekilde ortaya çıkar. Bu anlamda DMF Tek biçimli DMF (TB-DMF) ve Tek Biçimli Olmayan DMF (TBO-DMF) olmak üzere iki türüdür.

Tek biçimli DMF (TB-DMF)

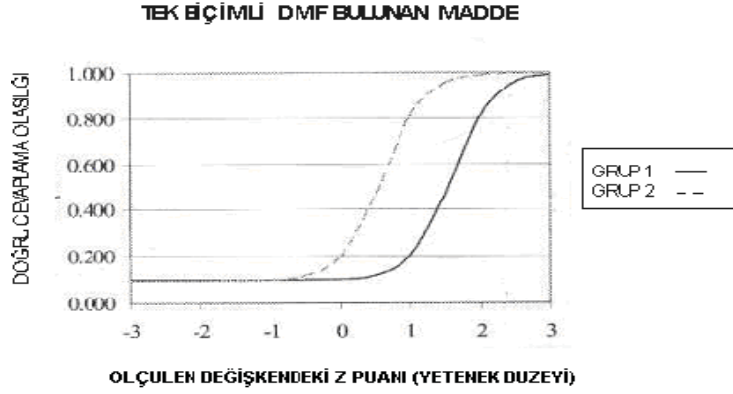
Bir maddenin doğru cevaplanma olasılığı, tüm yetenek düzeyleri için bir grubun lehine işlediğinde tek biçimli DMF söz konusu olur (Zumbo, 1999).



Şekil 1-2. Erkek ve kız öğrencilerin farklı yetenek düzeyindeki maddeyi doğru yanıtlama olasılıkları (Yurdugül, <http://yunus.hacettepe.edu.tr>, 21 Aralık 2011’de erişildi).

Şekil 1-2’ de görüldüğü gibi alt grup olarak cinsiyet alındığında, erkeklerin maddeyi doğru cevaplama olasılığı farklı yetenek düzeylerinin her noktasında kızlara göre daha yüksek ise, yani düzlemde bir kesişim söz konusu değilse bu şekildeki DMF’ ye “tek biçimli DMF” adı verilir (Yurdugül, <http://yunus.hacettepe.edu.tr>, 21 Aralık 2011’de erişildi).

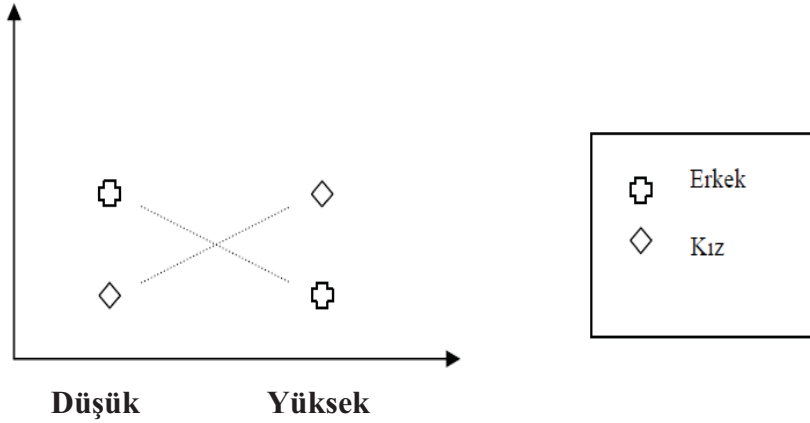
Tek biçimli DMF, bireylerin yetenek düzeyleri değiştikçe maddeyi doğru cevaplama olasılığındaki farklılığın da aynı düzeyde değişmesi durumudur. Şekil 1-3’te tek biçimli DMF gösteren bir maddenin madde karakteristik eğrisi verilmiştir. Madde karakteristik eğrileri arasında kalan alanın genişliği önemli derecede DMF’ nin varlığını göstermektedir. Aynı zamanda madde karakteristik eğrileri birbirlerini kesmedikleri için maddenin tek biçimli DMF’ye sahip olduğu söylenir (Zumbo, 1999).



Şekil 1-3. Tek biçimli DMF bulunan madde grafiği (Zumbo, 1999: 20)

Tek biçimli olmayan DMF (TBO-DMF)

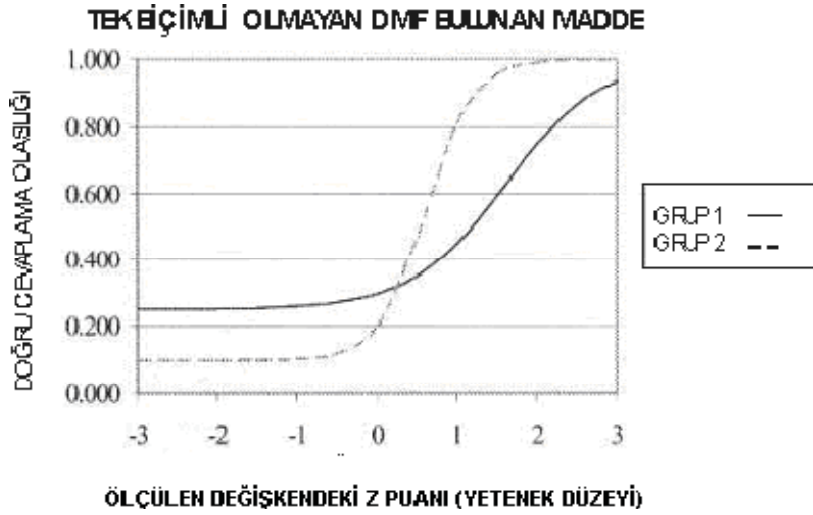
Bir maddenin doğru cevaplanma olasılığı, farklı yetenek düzeylerinde farklı gruplar lehine işlediğinde tek biçimli olmayan DMF'den söz edilmektedir (Zumbo, 1999).



Şekil 1-4. Erkek ve kız öğrencilerin farklı yetenek düzeyindeki maddeyi doğru yanıtlama olasılıkları (Yurdugül, <http://yunus.hacettepe.edu.tr>, 21 Aralık 2011'de erişildi).

Şekil 1-4'de görüldüğü gibi alt grup olarak cinsiyet alındığında, erkeklerin ve kızların maddeyi doğru cevaplanma olasılığı farklı yetenek düzeylerinde sabit kalmamıştır. Bireylerin yetenek düzeyi değiştikçe maddeyi doğru cevaplanma olasılıkları da değişmiştir.

Tek biçimli olmayan DMF, aynı yetenek düzeyindeki iki grubun ilgili maddeyi doğru cevaplama olasılığındaki farklılığın, tüm yetenek düzeylerinde tutarlı olmadığına görülür (Camili ve Shephard, 1994). Şekil 1-5'te tek biçimli olmayan DMF gösteren bir maddenin madde karakteristik eğrisi verilmiştir. Madde karakteristik eğrileri arasında kalan alan geniştir bu durum DMF'nin varlığını gösterirken eğrilerin birbirlerini çapraz kesmiş olmaları tek biçimli olmayan DMF'nin varlığını göstermektedir (Zumbo, 1999). DMF belirleme yöntemlerinin birçoğu tek biçimli olmayan DMF'yi belirlemeye duyarlı değildir (Jodoin ve Gierl, 2003, aktaran: Atalay, 2010).



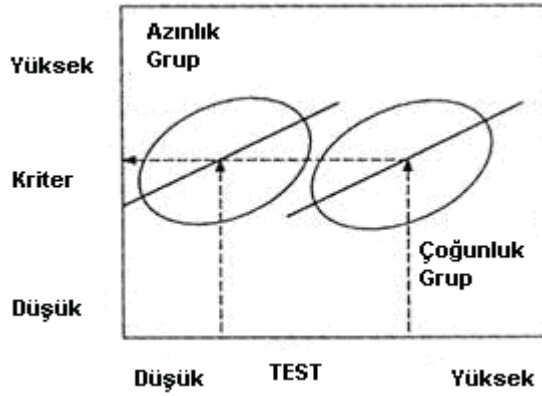
Şekil 1-5. Tek biçimli olmayan DMF bulunan madde grafiği (Zumbo, 1999: 20).

Camile ve Shepard (1994) ise yanlılığı ortaya çıkarmaya yönelik yöntemleri dışsal ve içsel yöntemler olarak tanımlamışlardır.

Dışsal yöntemler

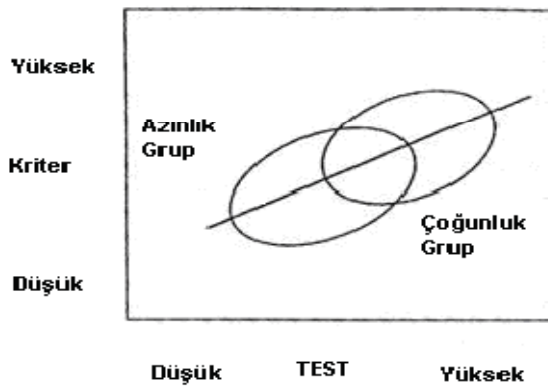
Dışsal yöntemlere göre yanlılık incelemeleri daha çok geleneksel yöntemler olarak bilinmekte ve test yanlılığı olarak adlandırılmakta olup testin tek tek maddelerinden ziyade toplam test puanı düzeyinde testin yordama geçerliğine odaklanmaktadır. Dış ölçüt dayanaklı yanlılık incelemeleri karşılaştırma gruplarının

testten alınan toplam puan ortalamaları arasındaki farklılığa ilişkin olarak ölçüt puan üzerindeki karşılaştırmaları içerir (Camile ve Shepard, 1994). Yani, dışsal yöntemde, test ve ölçülen yetenekle ilişkili olduğu düşünülen dış ölçüt puanları elde edilir ve maddelerin gruplara göre dış ölçüt puanları ile ilişkisi incelenir. Dışsal yöntemde, ölçüt ve test puanları arasında alt gruplara göre farklı regresyon çizgileri elde edilebiliyorsa ilgili madde yanlı iken (şekil 1-6); ölçüt ve test puanları arasındaki regresyon çizgisi alt gruplara rağmen sürekli ise ilgili madde yansızdır (Yurdugül, <http://yunus.hacettepe.edu.tr>, 21 Aralık 2011’de erişildi).



Şekil 1-6. Yanlılık Gösteren Madde (Bekçi, 2007: 10).

Şekil 1-6, regresyon sistemlerinin iki grup için çok farklı olduğunu gösterir. Bundan dolayı test azınlık grup için, belirlenen ölçüte göre yanlılık göstermektedir (Bekçi, 2007).



Şekil 1-7. Yansız Madde (Bekçi, 2007: 10).

Şekil 1-7’de grup ortalamalarındaki farka rağmen test azınlık ve çoğunluk grup için eşit bir tahmin edici geçerliğe sahiptir (Bekçi, 2007).

İçsel yöntemler

İçsel yöntemlere dayalı yanlılık incelemeleri literatürde madde yanlılığı olarak bilinmektedir. Testin içsel yanlılık incelemeleri bir dış ölçüt bulunmadığı durumda tüm test ile testin her bir maddesinin arasındaki yapı geçerliği ilişkilerini ve psikometrik incelemelerini kapsar. İç ölçüte göre uygulanan madde yanlılığı incelemelerinin en temel amacı, herhangi bir dış ölçüt alınmaksızın aynı popülasyonlardan seçilen farklı alt popülasyonlarda maddelerin aynı tarzda işlev görmesidir (Hulin, Drasgow ve Parsons, 1983). Buna göre bir test maddesi bağlı olduğu alt grup ile etkileşim gösteriyorsa o maddenin yanlı olduğu söylenir.

Değişen madde fonksiyonunun olası sebepleri

DMF durumunda maddeler, testi alan alt gruplardan en az biri için daha az geçerliğe sahiptir. DMF gösteren maddelerin testten çıkarılması, testi alan tüm gruplar için testin güvenilirliğini artırmaktadır (Thissen, 2001, aktaran: Kalaycıoğlu, 2008). Dolayısıyla, DMF’nin olası sebeplerinin bilinmesi ölçme aracı geliştirme çalışmalarında oldukça büyük önem taşımaktadır. Araştırmacılar cinsiyet, kültür, dil, ırk, sosyal sınıf veya bölge farklılıklarını göre DMF çalışmaları yapmışlardır (Doolittle ve Cleary 1987; Allalouf, Hambleton ve Sireci 1999; Harris ve Carlton, 1993; Azen, Bronner ve Gafni 2002).

DMF’ nin olası sebepleri şöyle sıralanabilir;

1. Sosyo-ekonomik farklılıklar,
2. Eğitim sistemindeki farklılıklar,
3. Yetersiz veya yanlış çeviriler,
4. Orijinal maddedeki anlam belirsizliği
5. Madde içeriğinin farklı kültürlerle uygunluğu,
6. Pek kullanılmayan veya soyut kelimelerin kullanımı,

7. Kelimelerin deęişken içeriksel anlamlara sahip olması,
8. Negatif, edilgen ve mecazi anlam içeren cümlelerin kullanımı,
9. Deyimsel ifadelerin kullanımı,
10. Ezber gerektiren ve karmaşık referanslı metinlerin kullanımı,
11. Gramer hatalarının tespitini gerektiren maddeler, (Van de Vijver ve Tanzer, 1997; Uiterwijk ve Valen, 2005; Ercikan, 1998; Grisay ve dięerleri, 2007, aktaran: Kalaycıoęlu, 2008: 20).

PISA, TIMSS ve PIRLS gibi uygulamalar farklı dilleri konuşan ve farklı kültürel özelliklere sahip olan öğrencileri bilgi, yetenek ve tutum gibi özellikleri ölçerek karşılaştırmaktadır. Dolayısıyla bu uygulamaları oluşturan maddelere yapılan DMF çalışmalarında, genellikle dilsel ve kültürel farklılıkların, müfredat deęişikliklerinin ve çeviri problemlerinin DMF'ye neden olduęu ortaya çıkmıştır (Yıldırım ve Berberoęlu, 2009; Le, 2009, aktaran: Atalay, 2010).

Çeviri testlerde DMF görülmesinin nedenlerinin belirlenmesi; test geliştirmede, eşitlemede ve puanlamada uygulayıcılara önemli yararlar sağlayacaktır (Allaouf ve dięerleri, 1999). Zumbo da (2007), uluslararası araştırmalarda kullanılan çeviri ve uyarlama yapılan ölçme araçlarında DMF belirlemenin önemine dikkat çekmiştir.

DMF ye etki eden bir dięer etken de içerik ve madde puanlama yöntemlerindeki farklılıklardır. Çoęu iki kategorili veriler üzerinde yürütölen DMF çalışmasında erkekler kızlardan daha iyi bir performans gösterirken, çok kategorili maddelerde ise kızlar erkeklerden daha iyi performans göstermişlerdir. Bunun nedeninin, çok kategorili maddelerin yazma, okuma ve yorumlama becerilerinde sözel ifadede akıcılığı gerektirmesi olduęu öne sürölmüşür (Willingham ve Cole, 1997, aktaran: Atalay, 2010)

1.1.3. Pisa ve Pisa öğrenci anketi

Ekonomik İşbirliği ve Kalkınma Örgütü (OECD)'nin, Uluslararası Öğrenci Başarılarını Deęerlendirme Programı (PISA); 15 yaşı grubu öğrencilerine uyguladıęı

temel beceri arařtırmaları ile öğrenciler yaşam boyu karşılaşacakları zorlukların üstesinden gelmeye hazırlar mı? Düşüncelerini etkili bir şekilde ifade edebiliyorlar mı? Analiz yapıp doğru sonuçlara ulaşabiliyorlar mı? Toplumun ve ekonomi dünyasının üretken bireyleri olarak hayatlarını devam ettirecekleri ilgi alanları var mı? sorularına yanıt bulmaya çalışmaktadır ve OECD üyesi ülkeler ile üye olmayan ülkelerin katılımıyla 3 yılda bir gerçekleştirilmektedir (MEB, 2007).

PISA; okuma becerileri, matematik ve fen bilimleri konularında temel becerilere odaklanarak, zorunlu eğitimin sonunda öğrencilerin topluma tam olarak katılması için bu bilgi ve becerileri ne derece kazandıklarını değerlendirmekte ayrıca uyguladığı anketler ile öğrencilerin ev, aile ve okul ortamları; kendileri hakkındaki görüşleri, öğrenme stilleri ve tutumları ile ilgili bilgiler toplamaktadır. (MEB, 2007).

PISA uygulamasında 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. Türkiye ilk olarak 2003 yılında eğitimde hangi seviyede bulunduğunu görmek ve eğitim düzeyini arttırmak amacıyla PISA projesine katılmıştır ve katılmaya devam etmektedir.

PISA; OECD eğitim direktörlüğüne bağı olan PISA yönetim kurulu tarafından yürütölmektedir. Ölçme araçlarının geliştirilmesi, örneklem seçimi, verilerin analizi, yorumlanması ve uluslararası raporların hazırlanması işlemleri liderliğini Avustralya Eğitim Arařtırmaları Konseyi (ACER)’ın oluşturduğu bir konsorsiyum tarafından yapılmaktadır. Ulusal düzeyde çeviri ve uyarlama işlemleri, projenin uygulanması, analizlerin yapılması ve ulusal raporların yazılması ise katılımcı ölkelerde belirlenen merkezler tarafından yapılmaktadır (MEB, 2007). Orijinal ölçme araçları öncelikle İngilizce ve Fransızca olarak hazırlanmakta daha sonra ulusal dillere çevrilmesi ve uyarlanması için katılımcı ölkelere gönderilmektedir. Her bir ölkö bu iki versiyonu kullanarak PISA çeviri ve uyarlama yönergelerine bağı olacak şekilde iki farklı çeviri yapmakta daha sonra bunları tek bir son versiyona dönüştörmektedir. 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, aktaran: Asil, 2010).

PISA 2006 OECD'ye üye 30 ülke ile üye olmayan 27 ülkeye uygulanmıştır. Bu ülkelerdeki 15 yaş grubu öğrencileri temsil edecek şekilde yaklaşık 400 000 öğrenci rastgele seçilmiştir. PISA 2006 kapsamında öğrencilere; yazılı sınavların yanı sıra öğrenci anketleri uygulanmıştır. Bu anket kapsamında öğrencilerin özgeçmişleri, öğrenme alışkanlıkları, motivasyonları ve fen bilimlerine yönelik tutumları ölçülmeye çalışılmıştır. PISA 2006, öğrencilerin tutumları ve fen bilimleri ile ne kadar iç içe oldukları hakkında veri toplamaya oldukça önem vermiştir. Çünkü PISA araştırmalarında tutumlar öğrencilerin fen bilimlerindeki yeterliklerinin anahtar bir parçası olarak görülmüştür (OECD, 2007a).

PISA 2006'da öğrencilerin fen bilimlerine yönelik tutumları sadece öğrenci anketinde yer alan ölçek maddeleri ile değil aynı zamanda öğrenci performansının ölçüldüğü fen bilimleri testi içine entegre edilmiş bazı sorularla da ölçülmüştür. Tutumları ölçen maddeler 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). Öğrencilerin fen bilimlerine yönelik tutumları, “bilimsel sorgulamaya verilen destek”, “fen bilimlerine yönelik öz inanç”, “fen bilimlerine ilgi”, “çevre ve kaynaklara yönelik sorumluluk” başlıkları altında ölçülmeye çalışılmıştır. PISA 2006'da ölçülen, öğrencilerin fen bilimlerine yönelik tutumları ile ölçülmek istenilen bu özellikler Tablo1-1'de özetlenmiştir (OECD, 2007a).

Tablo 1-1. PISA 2006 fen bilimlerine yönelik tutum

Bilimsel Sorgulamaya Verilen Destek; • 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 İnancı; • Bilimsel görevleri etkili bir şekilde yerine getirir. • Bilimsel problemlerin çözümünde zorlukların üstesinden gelir. • Güçlü bilimsel beceriler gösterir.
Fen Bilimlerine İlgi, • 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 • 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 ortaya koymaktadır (OECD, 2007a).

PISA araştırmaları sonucu elde edilen, öğrencilerin derslere karşı tutum, ilgi ve değer yargıları önemli öğrenme ürünleri olarak görülmekte; öğrenci performanslarındaki farklılıklar bu bilgiler yardımıyla açıklanmaya çalışılmakta ve milletler arası karşılaştırmalar yapılmaktadır. Literatür incelendiğinde daha çok bilişsel

testler üzerinde yanlılık, eşdeğerlik gibi çalışmaların yapıldığı öğrenci anketleri üzerinde bu çalışmaların yeterli olmadığı görülmektedir. Bu çalışmada PISA 2006 öğrenci anketinde yer alan fen bilimlerine yönelik tutum ölçeğinin “bilimsel sorgulamaya verilen destek” alt başlığındaki maddelerin Avustralya, Yeni Zelanda, Amerika Birleşik Devletleri ve Türkiye örneklemelerinde DMF gösterip göstermediği incelenmiştir.

1.1.4. Değişen madde fonksiyonunu belirlemede kullanılan istatistiksel yöntemler

DMF belirlemek için birçok yöntem kullanılmaktadır. (Camilli ve Shepard, 1994; Clauser ve Mazor, 1998; Zumbo, 1999). Bu yöntemler farklı kişiler tarafından farklı biçimlerde sınıflandırılmışlardır. Genel olarak DMF belirleme yöntemleri klasik test kuramına, madde tepki kuramına veya bazı istatistiksel tekniklere dayanarak açıklanabilir.

Klasik test kuramı, gözlenen puan, gerçek puan ve hata puanı kavramlarına dayanır. Klasik test kuramında gerçek puan, sayılabilir sonsuz çokluktaki ölçme sonuçlarının aritmetik ortalaması olarak ifade edilmektedir. Bu ölçme sonuçlarında bulunan hataların evrendeki dağılımının aritmetik ortalamasının sıfıra eşit olduğu sayılına dayanılarak gözlenen puanın, sonsuz çokluktaki ölçme sonuçlarının ortalaması alındığında gerçek puana eşit olduğu söylenir. Bu anlamda gerçek puan teorik bir kavramdır. Gözlenen puan gerçek puan ile hata puanı cinsinden, $X=T+E$ eşitliği ile ifade edilir ve bu ifade klasik test kuramının temel denklemi olarak adlandırılır (Lord ve Novick, 1968). Klasik test kuramına dayalı yöntemler madde indekslerinin (madde güçlük indeksi (p) ve madde ayırıcılık indeksi (r)) alt gruplara göre karşılaştırılması üzerine kurulmuştur ve KTK yöntemleri kullanılarak yapılan DMF araştırmalarında alt gruplara ait ortalama puanlar arasındaki farklılıklar test edilir.

Madde tepki kuramı, bir maddenin doğru cevaplandırılma olasılığı ile o maddenin ölçtüğü yetenek arasındaki ilişkiyi belirleyen matematiksel bir model olarak tanımlanmaktadır. MTK’ ya göre, uygulayıcının bir test maddesini cevaplama davranışı, örtük özellik veya yetenek olarak ortaya konan bir faktör yardımıyla tahmin edilebilir.

Bir kişinin bir maddeyi cevaplama davranışı ile o maddeye cevap verme davranışını belirleyen özellik yani yeteneği arasında bir ilişki vardır ve bu ilişki madde karakteristik eğrisi (MKE) ile gösterilebilir. Buna göre, yüksek yetenek düzeyindeki yanıtlayıcıların maddeye doğru cevap verme olasılıkları daha yüksektir (Hambleton ve Swaminathan, 1985). MTK yöntemlerinde aynı yetenek düzeyine sahip odak ve referans grubun bir maddeye doğru cevap verme olasılıkları arasındaki farklar incelenir. Başka bir ifade ile MTK ile yapılan DMF belirleme yöntemlerinde alt gruplardan (odak ve referans grup) elde edilen madde parametreleri veya MKE'leri karşılaştırılır (Deveci, 2008).

Benito ve Navas-Ara (2000, 17), DMF belirleme yöntemlerini KTK, faktör analizi, Ki-Kare, ve MTK' ya bağlı yöntemler olarak sınıflandırmışlardır (Aktaran: Deveci, 2008: 6). Bu sınıflandırmada KTK'ya bağlı yöntemleri delta plot ve madde ayırt edicilik yöntemleri olarak ele almışlardır. Faktör analizi yöntemlerini, sınırlandırılmış ve sınırlandırılmamış faktör analizi yöntemleri; ki-kareye bağlı yöntemleri ise doğru cevapları dikkate alan Scheuneman'ın ki-karesi, doğru ve yanlış cevapları birlikte ele alan Camilli'nin ki-karesi, Mantel-Haenszel ve lojistik regresyon yöntemleri olarak sınıflandırmışlardır. MTK'ya dayalı yöntemleri ise madde karakteristik eğrileri arasında kalan alan hesabı ve odak - referans grubun madde parametrelerini karşılaştırma yöntemleri olarak ele almışlardır (Aktaran: Deveci, 2008).

Ellis ve Raju (2003), Mantel-Haenszel, lojistik regresyon ve delta plot yöntemlerini KTK; Lord'un ki-karesi, Raju'nun alan ölçümleri ve olabilirlik oran ölçümü yöntemlerini ise MTK içinde ele almışlardır. Böylece madde yanlılığı belirleme yöntemlerini, Klasik Test Kuramı ve Madde Tepki Kuramı başlıkları altında toplamışlardır.

Verilerin parametrik ve non-parametrik olma durumuna göre yapılan sınıflamada ise parametrik veriler için Lord'un ki-kare analizi, olasılık oran testi ve Raju'nun alan indeksleri, parametrik olmayan veriler için ise, SIBTEST ve Mantel-Haenszel yöntemlerinin kullanılabileceği belirtilmektedir (Benito ve Navas-Ara 2000, aktaran: Şencan, 2005).

Schimizu ve Zumbo (2005) DMF belirleme yöntemlerini; olasılık tabloları ve regresyon modelleri yolu ile madde cevapları, madde tepki kuramı ve çok boyutlu model, olmak üzere üç grup altında toplamışlardır. Dorans ve Potenza (1995) bu yöntemleri eşleştirme değişkenine ve modellerin parametrik olup olmama durumuna göre sınıflandırmışlardır (Aktaran: Atalay, 2010). Tablo 1-2’de bu sınıflamaya göre çok kategorili veriler için DMF belirleme yöntemleri gösterilmiştir.

Tablo 1-2. Çok kategorili verilerde DMF belirleme yöntemleri (Atalay, 2010: 8)

	Parametrik	Parametrik olmayan
Gözlenen Puan (Observed Score)	Sıralı lojistik regresyon Lojistik diskriminant fonksiyon analizi	Genelleştirilmiş Mantel Haenszel HW1 ve HW3 Standartlaştırma yöntemi (SD) Mantel
Örtük Değişken (Latent Variable)	MTK	Poly-SIBTEST

DMF belirlemek amacıyla önerilen yöntemlere ilişkin yapılan sınıflandırmalar incelendiğinde, DMF belirleme teknikleri klasik test kuramı kapsamında; madde ayıricılık gücü, madde güçlüğü, faktör analizi, varyans analizi, madde güçlük dönüşümü; madde tepki kuramı kapsamında işaretli ve işaretli alan indeksleri, Lord’un Ki-kare’si, madde parametreleri ya da en çok olabilirlik oranları farklarının karşılaştırılması; bazı istatistiksel tekniklere dayananlar arasında ise Ki-kare, Mantel–Haenszel test istatistiği ve lojistik regresyon gibi teknikler olarak sınıflandırılabilir (Zumbo, 1999; Hambleton, Swaminathan, & Rogers, 1991; Rodney & Drasgow, 1990; Raju, 1990; Mellenberg, 1989; Adams & Rowe, 1988; Seong & Subkoviak, 1987; Holland & Thayer, 1986; Shepard, Camilli, & Williams, 1985; Hills, 1984; Devine & Raju, 1982; Rudner, Getson & Knight, 1980, aktaran: Doğan ve Öğretmen, 2008).

Çok kategorili verilerde değişen madde fonksiyonu belirleme yöntemleri

İki kategorili veriler için DMF belirleme yöntemlerinden oldukça fazla geliştirildiği söylenebilir. Bunlardan bazıları, SIBTEST (Sheally ve Stout, 1993), lojistik regresyon (Swaminathan ve Rogers, 1990), madde tepki kuramı (Thissen, Steinberg ve Wainer, 1988), Mantel-Haenszel (Holland ve Tayler, 1988), standartlaştırma yöntemi (Dorans ve Kulick, 1986), standart ortalama farkları (Dorans ve Schmit, 1993), dönüştürülmüş madde güçlüğü, varyans analizi ve ki-kare yöntemleridir (Akratan: Atalay, 2010). Bu yöntemlerden en sık kullanılanlar ise MH, LR ve SIBTEST'dir. Bunlar MTK'ya dayalı olmamasına rağmen, MTK gibi bireyleri yetenek düzeyinde eşleştirerek DMF belirlerler ve böylece MTK yöntemleri ile karşılaştırma yapılmasına olanak sağlarlar. Bunun yanı sıra, bu yöntemler MTK yöntemlerine göre uygulaması ve yorumlaması daha kolay olan ve büyük örneklem gerektirmeyen yöntemlerdir (Camili ve Shephard, 1994; Henderson, 1999).

Çok kategorili veriler için kullanılan DMF belirleme yöntemleri ise genellikle iki kategorili veriler için kullanılan yöntemlerin bir uzantısıdır. Bu yöntemler arasında MTK yöntemleri (Wainer, Sireci ve Thissen, 1991), Mantel (Mantel, 1963), Genelleştirilmiş Mantel-Haenszel (Zwick, Donoghue ve Grima, 1993), sıralı lojistik regresyon (Wilson, Spray ve Miller, 1993), Poly-SIBTEST (Chang, Mazzeo ve Raussos, 1993), lojistik diskriminant fonksiyon analizi (Miller ve Sprey, 1993) ve standartlaştırma yöntemi (Dorans ve Kulick, 1986) yer alır (Mellor, 1995; Henderson, 1996, aktaran: Atalay, 2010).

İki kategorili verilerde DMF belirleme ile ilgili oldukça fazla sayıda çalışma olmasına rağmen, çok kategorili verilerde DMF belirleme ve DMF belirleme yöntemlerinin karşılaştırılması ile ilgili oldukça az sayıda çalışma yapılmıştır (Clauser ve Mazor, 1998; Downing ve Haladyna, 1997; Potenza ve Dorans, 1995, aktaran: Atalay, 2010). Bu araştırmada çok kategorili verilerde değişen madde fonksiyonunu belirleme tekniklerinden sıralı lojistik regresyon ve Poly-SIBTEST teknikleri kullanılacak olduğundan ayrıntılı olarak bu yöntemler tanıtılacaktır.

Sıralı lojistik regresyon (SLR)

Lojistik regresyon yöntemi kullanılarak DMF belirlemede, maddeye doğru cevap verme olasılığı, toplam puan ve toplam puan ile grup ilişkisinden yararlanılarak tahmin edilir. Lojistik regresyon model temelli bir yöntem olduğundan dolayı tek boyutlu ve tek boyutlu olmayan DMF aynı denklemde modellenenebilir ve katsayıları ayrı ayrı test edilebilir. Lojistik regresyon, iki kategorili verilerde DMF belirlemek için oldukça sık kullanılan bir yöntemdir (French ve Miller, 1996; Swaminathan ve Rogers, 1990; Zumbo, 1999).

French ve Miller (1996) lojistik regresyonun bir uzantısının çok kategorili maddelerde DMF belirlemede uygun olabileceğini önermişlerdir. Sıralı lojistik regresyon, Miller ve Spray (1993) tarafından çok kategorili verilerde DMF belirlemek için önerilirken Zumbo (1999) tarafından da çok kategorili veriler için DMF etki büyüklüğü geliştirilmiştir. Sıralı lojistik regresyon yöntemi iki kategorili ve çok kategorili verilerde DMF belirlemede kullanılabilir. Zumbo (1999) Sıralı lojistik regresyon yönetiminde madde cevaplarını bağımlı değişken, grup değişkeninin her birey için toplam puanla beraber grup toplam puan etkileşimini bağımsız değişken olarak kullanıldığını belirtmiştir.

Sıralı lojistik regresyonun varsayımı, ölçeğin tek boyutlu olmasıdır (Zumbo ve Gelin, 2003). Zumbo'nun (1999) önerdiği sıralı lojistik regresyon yaklaşımında, grup etkisi ve alt-üst ölçek puanı etkileşimi ile DMF testi yapılır. Zumbo (1999) oluşturduğu modele değişkenleri sırasıyla; ilk olarak toplam puan değişkenini, ikinci olarak grup değişkenini ve en son olarak toplam puan ve grup etkileşimi değişkenini girerek DMF analizi yapmıştır.

Lojistik regresyon eşitliği şöyledir:

$$y = b_0 + b_1TOP + b_2GRUP + b_3TOP*GRUP + \epsilon_i,$$

Model üç aşamadan oluştuğu için her aşamada bir ki-kare (χ^2) değeri elde edilir (Gelin ve Zumbo, 2003). Üçüncü aşamada elde edilen χ^2 değerinin birinci aşamada elde edilen χ^2 değerinden farkı, serbestlik derecesinin 2 olduğu χ^2 dağılımı ile karşılaştırılır ve DMF testinin anlamlılığı bu şekilde test edilir (Zumbo, 1999). Karşılaştırmada serbestlik derecesinin 2 alınması, 3. aşamada serbestlik derecesinin 3, birinci aşamada 1 olmasından kaynaklanmaktadır (Swaminathan ve Rogers, 1990) .

Sıralı lojistik regresyon analizi ile aynı zamanda DMF'nin türü hakkında da bilgi sahibi olunur. Swaminathan ve Rogers, (1990) üçüncü aşamadan elde edilen χ^2 değeri ile ikinci aşamadan elde edilen χ^2 değeri farkını serbestlik derecesi 2 olan χ^2 dağılımı ile karşılaştırarak tek biçimli olmayan veya ikinci aşamadan elde edilen χ^2 değeri ile birinci aşamadan elde edilen χ^2 değeri farkını serbestlik derecesi 2 olan χ^2 dağılımı ile karşılaştırarak tek biçimli olan DMF hakkında bilgi elde edilebileceğini vurgulamışlardır.

Sıralı lojistik regresyon analizi ile χ^2 değerinin yanı sıra, etki büyüklüğü kestiricisi de elde edilir. Etki büyüklüğü R^2 değeri üçüncü aşamadan elde edilen R^2 değerinin, birinci aşamadan elde edilen R^2 değerinden çıkartılması ile hesaplanır ve bu değer iki ve çok kategorili maddeler için kullanılabilir. Zumbo (1999) üçüncü model ile birinci modelden elde edilen R^2 değerleri farkının χ^2 testine karşılık geldiğini ortaya koymuştur. Zumbo ve Thomas (1996) tarafından önerilen ΔR^2 'nin DMF'nin yorumlanmasında kullanılan ölçütleri ise Tablo 1-3'de verilmiştir.

Tablo 1-3. ΔR^2 Değerini yorumlama ölçütleri (Zumbo ve Thomas, 1996: 92)

A düzey DMF	$\Delta R^2 < 0,13$	DMF yoktur ya da ihmal edilebilir düzeydedir.
B düzey DMF	$0,13 \leq \Delta R^2 < 0,26$	Orta düzeyde DMF vardır
C düzey DMF	$\Delta R^2 \geq 0,26$	Yüksek düzeyde DMF vardır.

Zumbo (1999) ve Jodoin ve Gierl (2001), bir maddenin DMF gösterdiğinin söylenebilmesi için serbestlik derecesi 2 (grup ve etkileşim etkileri testi) olan χ^2 testinin anlamlı olması (p değerinin .01'den küçük ya da eşit olması) ve etki büyüklüğünün en az .13 olması gerektiğini ortaya koymuşlardır.

Bir maddede DMF olduğu belirlendikten sonra sıralı lojistik regresyondaki aşamalar ile DMF'nin tek biçimli olup olmadığı da incelenebilir. Bu χ^2 testi ile belirlenebileceği gibi tek biçimli DMF, ikinci model ile birinci modelden elde edilen R^2 değerleri farkı ile, tek biçimli olmayan DMF ise üçüncü model ile ikinci modelden elde edilen R^2 değerleri farkı ile de tespit edilebilir.

Poly- SIBTEST

SIBTEST, Shealy ve Stout (1993) tarafından iki kategorili verilerde DMF belirlemek için geliştirilmiş bir yöntemdir. SIBTEST, DMF belirlemede, standartlaştırma yöntemi ile benzerdir; ancak SIBTEST'in standartlaştırma yönteminden farklı 1. tip hatayı kontrol etmede kullanışlı olan regresyon temelli düzeltmeyi kullanmasıdır (Clauser ve Mazor, 1998).

SIBTEST parametrik olmayan örtük değişken modelidir (Potenza ve Dorans, 1995, aktaran: Atalay, 2010: 12). Chang, Mazzeo ve Roussos (1996), SIBTEST yönteminin çok kategorili veriler için farklı bir formu olan Poly-SIBTEST yöntemini geliştirmişlerdir. Poly- SIBTEST örtük değişkenler için tanımlanmıştır ancak MTK yetenek ve madde parametrelerini kullanmaz. Poly-SIBTEST, tek biçimli DMF' yi belirlemek için tasarlanmış olmakla beraber iki kategorili verilerde tek biçimli olmayan DMF'yi belirlemek üzere Li ve Stout (1993) tarafından crossing-SIBTEST yöntemi geliştirilmiştir ancak çok kategorili veriler için crossing-SIBTEST yöntemi henüz geliştirilememiştir (Fang, 1999).

Poly-SIBTEST ile H_0 hipotezi, beklenen sonuçların θ yetenek düzeyinde referans ve odak grupta aynı olması durumu olarak tanımlanır. θ yetenek düzeyinde DMF'nin miktarı aşağıdaki denklemle hesaplanır.

$$B_0 = E_R[Y|\theta] - E_F[Y|\theta]$$

Bir maddede DMF olmadığı durumda $E_R[Y|\theta] = E_F[Y|\theta]$ eşitliği sağlanır.

SIBTEST yöntemi ile bir maddenin DMF gösterip göstermediğinin yanı sıra, maddenin ne kadar DMF gösterdiği de belirlenir. SIBTEST etki büyüklüğü, her bir madde için DMF miktarı olarak yorumlanır. Etki büyüklüğünün işareti ise bize DMF'nin hangi grup lehine çalıştığını gösterir. Eğer β değeri pozitifse o maddenin referans grubun lehine, β değeri negatifse ise o maddenin odak grubun lehine çalıştığı belirtilmiştir. β değeri aşağıda verilen denklemden elde edilir (Atalay, 2010).

$$\beta = \int \beta_0 f_F(\theta) d\theta$$

$f_F(\theta)$: odak gruptaki yoğunluğu,

β : DMF miktarı

Tablo 1-4. β Değerlerini yorumlama ölçütleri (Atalay, 2010: 13)

A düzey DMF	$ \beta < 0,059$	DMF yoktur ya da ihmal edilebilir düzeydedir
B düzey DMF	$0,059 \leq \beta < 0,088$	Orta düzeyde DMF vardır
C düzey DMF	$ \beta \geq 0,088$	Yüksek düzeyde DMF vardır

Poly-SIBTEST yöntemi ile incelenecek olan maddeler çalışan test ve eşleştirme testi olmak üzere iki alt teste ayrılır. Çalışan test olası DMF içeren maddelerden; eşleştirme testi ise testin geri kalan maddelerinden oluşur. Bireylerin

yetenek düzeyini temsil eden toplam puan SIBTEST ve Poly-SIBTEST bilgisayar programlarında eşleştirme testinden elde edilir (Henderson, 1999). Referans ve odak grubu oluşturan bireyler, toplam puana ve kategori sayısına göre eşleştirilir. Her bir gruptaki bireylerin, testle ölçülen özellik bakımından aynı yetenekte olduğu varsayılarak referans ve odak gruptaki bireylerin performansları her bir “çalışan test” için K grupta karşılaştırılır (Henderson, 1999).

1.2 Problem İfadesi

PISA 2006 öğrenci anketinde yer alan “bilimsel sorgulamaya verilen destek” maddeleri sıralı lojistik regresyon ve Poly – SIBTEST yöntemlerine göre farklı kültürler ve diller arasında değişen madde fonksiyonu göstermekte midir?

1.3. Araştırmanın Önemi

Ölçme araçlarında aranan en önemli özelliklerden biri geçerliktir ve DMF geçerliği düşüren bir faktördür. Yapılan sınavlarda kullanılan ölçme araçlarında DMF içeren madde olması bu sınavların sonuçlarının geçerliğini olumsuz bir şekilde etkileyecek ve sınav sonuçlarına bağlı olarak yanlış yorumlar ve düzeltmeler yapılmasına neden olacaktır.

DMF belirleme yöntemleri özellikle iki kategorili veriler için çok iyi yapılandırılmıştır. Kategorik maddeler için DMF belirleme ile ilgili birçok yöntem geliştirilmiş olmasına rağmen, bu yöntemler yeterince test edilmemiştir (Zumbo ve McDowell, 2005).

Türkiye’de DMF çalışmalarının son yıllarda yoğunluk kazandığı söylenebilir. DMF çalışmalarında genellikle klasik test kuramına dayanan teknikler, lojistik regresyon, Ki-kare MH veya örtük özellikler kuramına dayanan teknikler kullanılarak DMF veren maddeler belirlenmeye çalışılmaktadır (Doğan ve Öğretmen 2006; Öğretmen ve Doğan, 2004; Yurdugül, 2003; Özdemir, 2003; Öğretmen, 1995). Ancak, örneklem büyüklüğü,

verilerin yapısı, maddelerin puanlanış biçimi, DMF belirlemede kullanılan teknikler gibi etkenler maddelerin DMF düzeylerinin farklı araştırmalarda farklı şekilde belirlenmesine neden olabilmektedir. Bu nedenle de hangi tekniğin kullanılması gerektiğine karar verebilmek için DMF belirleme tekniklerinin karşılaştırılmasına ihtiyaç duyulmaktadır. Çalışma çok kategorili puanlanabilen testlerin değişen madde fonksiyonunu belirlemek için kullanılabilecek istatistiksel yöntemleri tanıtmayı, bu yöntemlerin benzerlik ve farklılıklarını ortaya koyması ve ileriki araştırmalara ışık tutması açısından önemlidir. Ayrıca SLR ve poly-SIBTEST yöntemleri ile DMF analizini yaparken farklı kültürel ve dilsel özelliklere sahip örneklemeler kullanılmış ve buna göre kültürden ve dilden kaynaklanan yanlışlık sorunlarının olup olmadığı da araştırılmıştır. Bu da ölçme eşdeğerliği olmayan uygulamalardan elde edilen verilere dayanarak ortaya konan sonuçların ne derece geçerli ve güvenilir olduğunun sorgulanması açısından önem arz etmektedir.

1.4. Araştırmanın Amacı

Çalışmanın amacı hem PISA 2006 öğrenci anketinin kültürel ve dilsel (İngilizce- Türkçe) yanlışlığını test etmek hem de farklı yöntemleri karşılaştırılarak uygulamada ortaya çıkan benzerlik ve farklılıkları belirlemektir. Bu amaçlar doğrultusunda PISA 2006 öğrenci anketindeki “bilimsel sorgulamaya verilen destek” ölçme modelinin her bir kültürdeki ve dildeki faktör yapısı DFA ile incelenmiş daha sonra farklı kültürel ve dilsel yapıdaki ülkelerde sıralı lojistik regresyon ve poly-SIBTEST yöntemleri ile karşılaştırmalı olarak DMF’nin varlığı tespit edilmiştir.

Araştırmanın amacı doğrultusunda aşağıdaki sorulara cevap aranmıştır.

1.5. Araştırma Soruları

1. PISA 2006 öğrenci anketi kapsamında ölçülen “bilimsel sorgulamaya verilen destek” ölçme modeli Avustralya, Yeni Zelanda, ABD ve Türkiye kültürlerinde ve dillerinde aynı faktör yapısına sahip midir?

2. PISA 2006 öğrenci anketi kapsamında ölçülen “bilimsel sorgulamaya verilen destek” ölçme modelinin orijinal ve uyarlanan maddeleri;

a) Avustralya ve Yeni Zelanda örneklemeleri arasında sıralı lojistik regresyon ve Poly-SIBTEST yöntemleri ile yapılan analizlerde değişen madde fonksiyonu göstermekte midir? Yöntemler arasında uyum var mıdır?

b) Avustralya ve ABD örneklemeleri arasında sıralı lojistik regresyon ve Poly-SIBTEST yöntemleri ile yapılan analizlerde değişen madde fonksiyonu göstermekte midir? Yöntemler arasında uyum var mıdır?

c) Avustralya ve Türkiye örneklemeleri arasında sıralı lojistik regresyon ve Poly-SIBTEST yöntemleri ile yapılan analizlerde değişen madde fonksiyonu göstermekte midir? Yöntemler arasında uyum var mıdır?

1.6. Sayıtlılar

Öğrencilerin anket maddelerine içtenlikle cevap verdikleri varsayılmıştır.

1.7. Sınırlılıklar

1. Araştırmada elde edilen veriler PISA-2006 verileri ile sınırlıdır.

2. Araştırmadaki diğer bir sınırlılık ülkelere ait örneklem büyüklüklerinin farklı olmasıdır. Analizlerde kullanılan ki-kare değerleri eşit olmayan bu örneklem büyüklüklerinden farklı oranlarda etkilenmiş olabilir.

3. Araştırmada DMF analizi, dört ülke verisi ve iki yöntemle sınırlandırılmıştır.

4. Referans değişken ülke olarak Avustralya ile sınırlandırılmıştır.

1.8. Tanımlar

Test Yanlılığı (Test Bias): Bir gruba üye olan bireylerin test puanlarının yordama ve yapı geçerliğindeki sistematik hatalardır (Jensen,1980).

Madde Yanlılığı (İtem Bias): Testin amacına uygun olmayan test koşullarından ya da test maddelerinin bazı karakteristik özelliklerinden dolayı bir maddeyi bir grubun doğru yanıtlamasının diğer gruba göre daha az ya da çok olmasıdır (Zumbo,1999).

Madde Etkisi: Gruplar arasında var olan gerçek farklılıktır. Farklı gruplardaki yanıtlayıcıların bir maddeyi doğru yanıtlama olasılıklarındaki farklılıkları ifade eder (Zumbo,1999).

Değişen Madde Fonksiyonu (DMF): Maddenin ölçmek istediği yetenek düzeyinde yapılacak bir karşılaştırmada farklı gruplardaki yanıtlayıcıların ilgili maddeyi doğru yanıtlama olasılıklarındaki farklılıklardır (Zumbo,1999).

BÖLÜM II

2. Literatür Taraması

Mellenberg (1982), Hambleton ve Rogers (1989), yaptıkları simülasyon çalışmasında 250 ve 500 kişilik iki ayrı grup ve her bir grup için 40, 60 ve 80 maddelik testler uygulamışlardır. Her testin %10'u tek biçimli, %10'u da tek biçimli olmayan DMF içermektedir. LR ve MH'nin her ikisi de tek biçimli DMF'yi belirleme açısından benzer sonuçlar vermiş olup her iki yöntem de tek biçimli DMF'yi belirlemede küçük grup için %75'lik doğruluk, büyük grup için de %100'lük doğruluk oranı elde edilmiştir. Tek biçimli olmayan DMF'nin belirlenmesinde ise durum oldukça farklıdır. MH bu durumda tamamıyla yetersiz kalmaktadır. LR ise küçük grup için %50'lik, büyük grup için %75'lik doğruluk sağlamıştır. Sonuç olarak iki yöntem de tek biçimli DMF'yi belirlemede eşit ve oldukça iyi performans gösterirken sadece LR tek biçimli olmayan DMF'yi tutarlı olarak belirleyebilmiştir.

Baghi, H.; Ferrara, S. F. (1989), çalışmalarında üç alanda (anayasal hükümet, politika, politik davranış ve haklar, ilkeler ve sorumluluklar) öğrencilerin bilgi ve becerilerini değerlendiren ve çoktan seçmeli 45 maddeden oluşan Maryland Test of Citizenship Skills (MTCS) testinde cinsiyet ve ırksal gruplar üzerinde DMF olup olmadığını MH, delta plot ve MTK tekniklerini kullanarak araştırmışlardır. Araştırmanın verileri 50.000 9. sınıf öğrencisi üzerinden elde edilmiştir. Analizler 1000, 750, 500 ve 200 kişilik temsil edici bir örneklem üzerinde yapılmış ve sonuçlar, MTCS maddelerinin ırka ve cinsiyete göre DMF içermediğini ortaya koymuştur. Irk ve cinsiyet karşılaştırmalarında tüm örneklem büyüklükleri için delta plot ve Rasch göstergeleri arasındaki benzerlik çok yüksek çıkmıştır (Aktaran: Bekçi, 2007).

Chang ve Mazzeo (1993), poly-SIBTEST yöntemini geliştirmek amacıyla simülasyon veri üzerinden yaptıkları çalışmada poly-SIBTEST yöntemi ile GMH ve Mantel yöntemlerini karşılaştırmışlardır. Sonuç olarak, modifiye edilmiş SIBTEST'in DMF belirlemede Mantel ve GMH yöntemi kadar başarılı olduğunu saptamışlardır (Aktaran: Atalay, 2010).

Miller ve Spray (1993), LDFA yönteminin çok kategorili maddelerde DMF belirlemede ne derece kullanışlı olduğunu göstermeye çalışmışlardır. Çalışmada açık uçlu 27 maddeden oluşan bir matematik testi kullanmışlardır. Sonuç olarak çok kategorili maddelerde, LDFA yönteminin sıralı lojistik regresyona göre daha pratik ve GMH yöntemine göre özellikle tek biçimli olmayan DMF'yi belirlemede daha güçlü olduğunu ortaya koymuşlardır.

Mellor (1995), çok kategorili maddelerde DMF belirlemek için GMH, poly-SIBTEST, SLR, LDFA yöntemlerini karşılaştırmıştır. Çalışmada 4 yöntemin farklı yetenek dağılımlarında tek biçimli olan ve olmayan DMF'yi belirleme gücü ve 1. Tip hata oranları incelenmiştir. Araştırma, hem simülasyon veri hem de Texas Eğitim şubesinden alınan gerçek veri üzerinden yürütülmüştür. Sonuç olarak, iki grubun yetenek dağılımlarının aynı olduğu durumlarda dört yöntemin de her kategoride tek boyutlu DMF'yi belirlemede iyi, ancak GMH ve poly-SIBTEST'in daha hassas olduğunu ortaya koymuşlardır. Yetenek dağılımlarının eşit olmadığı durumlarda ise GMH yönteminin en iyi çalışan yöntem olduğunu ve 1. Tip hatayı kontrol gücünün mükemmel olduğunu göstermişlerdir. Gerçek veri çalışmasında da etnik kökene göre DMF belirlemişlerdir. Hispanic ve Beyaz karşılaştırmasında 4 yöntemde aynı düzeyde tek biçimli DMF görüldüğünü, beyaz ve siyah karşılaştırmasında ise LDFA yöntemi ile tek biçimli olmayan DMF görülürken, SLR ve GMH yöntemleri ile de tek biçimli DMF görüldüğünü ortaya koymuşlardır.

Berberoğlu (1995b), 1992 yılında uygulanan ÖSS matematik alt test maddeleri için, sosyo-ekonomik düzeye göre DMF araştırması yapmıştır. Aynı zamanda düşük ve yüksek sosyo-ekonomik statüdeki öğrencilerin cinsiyetlerini de DMF araştırması kapsamında incelemiştir. Araştırmanın sonucunda genellikle üst sosyo-ekonomik

düzeye mensup öğrencilerin sözel problem formatındaki maddelerde daha avantajlı olduklarını, ayrıca cinsiyete göre kız öğrencilerin sözel ve uzamsal yeteneğe yönelik maddelere ilişkin performanslarının erkek öğrencilere göre yüksek olduğunu, hesaplamaya yönelik becerilerde ise erkek öğrencilerin performansının daha yüksek olduğunu ortaya koymuşlardır.

Öğretmen (1995), yapmış olduğu araştırmada 1993 yılı Öğrenci Seçme Sınavı'nın sözel yetenek bölümündeki maddelerin madde yanlılığını araştırmıştır. 1474 kız ve 1326 erkek öğrenciden oluşan toplam 2800 öğrenci verisi üzerinde çalışılmasını yürütmüştür. Cinsiyet grupları üzerinde MTK ile yapmış olduğu madde yanlılığı analizleri sonucunda, Türkçe ve Sosyal Bilgiler testlerindeki çoğu maddenin madde karakteristik eğrilerinde farklı yetenek seviyelerinde cinsiyete göre yanlılık gözlemlenmiştir. Konu alanlarında ise, cinsiyete göre madde yanlılığı gözlemlenmemiştir.

Narayanan P. & Swaminathan H. (1996), çalışmalarında tek biçimli olmayan DMF gösteren maddeleri belirlemeyi amaçlamışlardır. Bunun için araştırmada 1. tip hata oranı ve güce göre MH, SIBTEST ve LR yöntemlerini karşılaştırmışlardır. Veri seti, örneklem genişliği, odak ve referans gruplar arasındaki yetenek dağılımı farklılığı, testte yer alan DMF içeren madde oranı, DMF etki büyüklüğü ve madde tipine göre manipüle edilmiş ve 384 durum çalışılmıştır. Çoğu durumda tek biçimli olmayan DMF'nin belirlenmesinde hem SIB hem de LR eşit olarak güçlü bulunmuştur. Sırasız (dissıralı) etkileşim içeren tek biçimli olmayan DMF'yi belirlemede MH çok etkili bulunmamıştır. 1. tip hata büyüklüğü, MH için beklenen sınırlar içinde, SIB ve LR için beklenenden daha üst sınırlarda bulunmuştur. SIB sonuçları tüm olarak LR sonuçlarının yaklaşık %1 üzerinde çıkmıştır.

Chang, Mazzeo ve Roussos (1996) iki farklı simülasyon çalışmasıyla SMD ve Mantel yöntemleri ile poly-SIBTEST yöntemini karşılaştırmıştır. Birinci çalışmada, Mantel ve SMD' nin 1. Tip hatayı kontrol etmede ve tek boyutlu DMF' yi belirlemede güçlü olduğu belirlenmiş olan veri seti kullanılmış ve SIBTEST yönteminin iyi çalıştığı; ancak SMD ve Mantel yöntemlerinin daha iyi olduğu sonucuna varılmıştır. İkinci çalışmada ise maddelerin zorluk derecelerinin aynı; fakat ayırt ediciliklerinin farklı

olduğu durumlarda bu üç yöntemin gücü karşılaştırılmıştır. Sonuç olarak SIBTEST' in 1. tip hatayı daha iyi kontrol ettiği ve tek biçimli DMF belirlemede diğer yöntemlere göre daha güçlü olduğunu ortaya koymuşlardır.

Stage (1997), İsveç'te yüksek öğretime öğrenci seçmede kullanılan SweSAT (The Swedish Scholastic Aptitude Test) testi üzerinde yaptığı çalışmada, matematiksel düşünme yeteneğini ve şekil, tablo, grafik okuma yeteneğini ölçmeye yönelik maddelerin erkek öğrenciler lehine; kelime bilgisini ölçmeye ve okuduğunu anlamaya yönelik maddelerin ise kız öğrencilerin lehine çalıştığını gözlemlemiştir. Stage (1997) ayrıca bu çalışmanın devamı niteliğinde yapmış olduğu başka bir çalışmada KTK madde karakteristiklerindeki farklılıklar (madde güçlük ve ayıricılık indeksi) ve MH istatistikleri kullanmıştır (Aktaran: Bekçi, 2007).

Fang (1999), çalışmasında Mantel, GMH, poly-SIBTEST, SD ve LDFA yöntemlerinin 1. tip hatayı kontrol ve DMF belirleme güçlerini incelemiştir. İncelemeye alınan faktörler arasında grup yetenek dağılımı farkları, test uzunluğu, örneklem büyüklüğü ve madde ayırt ediciliği yer almaktadır. Grup yetenek düzeylerinin benzer olduğu ve madde ayırt ediciliğin yüksek olmadığı, ya da grup yetenek düzeylerinin çok farklı olduğu ve madde ayırt ediciliğinin düşük olduğu durumlarda uygulanan bütün yöntemlerde 1. tip hatanın kontrol edildiği ve yöntemlerin birbirleriyle karşılaştırılabilir olduğu sonucuna ulaşılmıştır. Ancak grup yetenek düzeylerinde az fark olduğu ve madde ayırt ediciliğinin yüksek olduğu, ya da grup yetenek düzeyleri çok farklı ve madde ayırt ediciliğin yüksek olduğu durumlarda bu yöntemlerin kesin olmayan sonuçlar verdiği görülmüştür. Çalışmada kullanılan dört yöntemle de tek biçimli DMF belirlenebilmiş ve en iyisinin LDFA olduğu gözlemlenirken; tek biçimli olmayan DMF belirlenmesinde ise LDFA ve GMH yöntemlerinin diğer yöntemlere göre daha güçlü olduğu sonucuna varılmıştır.

Gierl, Khaliq, Boughton (1999), Kanada'da her yıl uygulanan Alberta Matematik ve Fen Başarı Testindeki maddelerin DMF analizini 12000 öğrenciye MH, LR ve SIBTETST yöntemleri ile yapmışlardır. Sonuç olarak matematik testinde 469 maddeden MH ile 34 maddede, SIBTEST ile 40 maddede, LR ile 49 maddede DMF

bulmuşlardır. Fen testinde ise toplam 428 maddeden MH ile 36 maddede, SIBTEST ile 63 maddede ve LR ile 74 maddede DMF bulmuşlardır (Aktaran: Bekçi, 2007).

Henderson (1999) , çok kategorili verilerde DMF belirlemek için geliştirilmiş olan GMH, poly-SIBTEST ve LDFA yöntemlerini gerçek veri kullanarak cinsiyet grupları üzerinde karşılaştırmıştır. Çalışma sonucunda, Poly – SIBTEST’ in en çok sayıda DMF gösteren maddeyi belirlediği ve bu maddelerin büyük bir çoğunluğunun da GMH ve LDFA analizlerinde DMF gösterdiği ve çoktan seçmeli maddelerin erkeklerin lehine, yapılandırılmış maddelerin ise kızların lehine çalıştığı sonucuna varmıştır.

Penfield (2001), çoklu alt gruplarda DMF’ yi belirlemede kullanılan üç farklı MH test istatistiğini karşılaştırmıştır. Bu yöntemler; alfa düzeyi ayarlanmadan hesaplanan MH χ^2 istatistiği, alfa düzeyi ayarlanarak bonferroni ile hesaplanan MH χ^2 istatistiği ve genelleştirilmiş MH test istatistiği (GMH)’dir. Tek bir referans grup ve dört farklı odak grup kullanılmıştır. Sonuçlar, 0,05 düzeyinde birinci tip hatayı azaltmak için GMH istatistiğinin daha uygun olduğunu göstermiştir.

Gelin ve Zumbo (2003) yaptıkları çalışmada DMF sonuçlarının madde puanlama yöntemine bağlı olup olmadığını iki ve çok kategorili verilerde kullanılabilen SLR yöntemi ile incelemişlerdir. Sonuç olarak puanlama yöntemin DMF üzerinde etkisi olduğu bulunmuştur.

Hidalgo ve Lopez-Pina (2004), LR, MH ve Mazor, Clauser, Hambleton tarafından yeniden düzenlenen MH yöntemlerini karşılaştırmışlardır. Çalışmada veri seti madde parametrelerinin farklı durumları ve DMF büyüklüğü manipüle edilmiştir. Ayrıca tek biçimli olan ve olmayan DMF durumları da simule edilmiştir. Yapılan simulasyon çalışmasında öğrencilere 59 DMF içermeyen, 16 DMF içeren toplam 75 maddelik 25 ayrı test uygulanmıştır. 1000 referans grup, 1000 odak grup üzerinden toplanan veriler üç ayrı yöntemle analiz edilmiştir. LR, DMF bulunan maddeleri belirlemede daha duyarlı bulunmuştur.

Yurdugöl ve Aşkar (2004), Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı'nda (OÖK-ÖSYS) yer alan maddelerin, sınavı alan öğrencilerin yaşadığı yerleşim yeri alt gruplarına göre değişen madde fonksiyonlarını incelemişlerdir. DMF araştırmaları için MH ve LR istatistiklerinden yararlanmışlar ancak her iki analiz de yaklaşık aynı sonucu verdiği için MH bulgularına çalışmada yer vermişlerdir. Araştırmacılar, Devlet Planlama Teşkilatı çalışmaları doğrultusunda Türkiye'deki 83 ili, sosyoekonomik gelişmişlik göstergelerine göre temel bileşenler analizi yardımı ile 5 ayrı bölgeye ayırarak gruplandırmışlar ve bu bölgelere göre DMF araştırması yapmışlardır. Araştırmada elde edilen bulgular doğrultusunda, türkçe ve sosyal bilimler alt testlerindeki maddelerin farklı fonksiyonlaşmadığı veya fonksiyonlaşmanın ihmal edilebilir düzeyde olduğu, matematik alt testinde yer alan maddelerin büyük çoğunluğunun ikili karşılaştırmalarda, daha çok gelişmiş bölgelerde yaşayan öğrenciler lehine çalıştığı sonucuna varmışlardır.

Öğretmen T. ve Doğan N. (2004), Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı Matematik alt testindeki maddelerin madde yanlılığı analizlerini Madde-Tepki Kuramına göre işaretli ve işaretsiz alan indekslerini kullanarak yapmışlardır. Hesaplanan alan indekslerinin büyüklüklerini matematik alt testinin konu alanlarına göre karşılaştırmışlardır. Araştırmaya göre matematik alt testindeki maddelerin cinsiyete göre yanlılık gösterdiği saptanmıştır. Ayrıca işaretli alan indeksleri sonuçlarına göre, testteki çoğu maddenin kızların lehine yanlılık gösterdiği ve matematik konu alanlarına göre işaretli ve işaretsiz alan indekslerinin büyüklükleri arasında anlamlı bir farkın bulunmadığını ortaya koymuşlardır.

Fidalgo, A. M.; Ferreres, D.; Muniz, J. (2004) yaptıkları çalışmada, küçük örneklemelerde DMF'yi tespit etmek için tip 1 hata oranı ve gücünü araştırmışlardır. Bunun için örneklem büyüklükleri odak ve referans gruplar için sırasıyla 50/50, 100/50, 200/50, 100/100, yetenek dağılımları eşit ve eşit olmayan ve DMF'nin miktarı ise yüksek ve düşük düzeyde manipüle edilmiştir. Sonuçlar, $\alpha = 0,20$ seviyesinde MH ki-kare istatistiği kullanmanın avantajlarını göstermiştir.

Kristjansson, Aylesworth, McDowell ve Zumbo (2005) yaptıkları çalışmada Mantel, GMH, LDFA ve SLR yöntemlerini karşılaştırmışlardır. DMF'nin grup yetenek farkları, madde ayırt ediciliği ve örneklem büyüklüğünden etkilendiğini gözlemlemişlerdir. Tüm yöntemlerde 1. tip hatanın kontrol altına alındığını ve tek biçimli DMF belirlemede uygun olduklarını ortaya koymuşlardır. Madde ayırt ediciliğinin yüksek olduğu durumlarda Mantel yönteminin tek biçimli olmayan DMF'yi belirleyemediğini, LDFA yönteminin ise zayıf olduğunu, SLR ve GMH yöntemlerinin ise benzer sonuçlar verdiğini göstermişlerdir.

Korkmaz (2005), yaptığı çalışmada madde tepki kuramına dayalı olarak kişilik ve tutum gibi çok kategorili maddelerde cinsiyete göre değişen madde fonksiyonu olup olmadığını incelemiştir. Araştırmada $n=500$, $n=1000$ ve $n=1500$ olmak üzere farklı örneklem büyüklükleri ve olabilirlik oranı testi, model karşılaştırma yöntemi, madde test işlev farklılık yöntemi ve parametre karşılaştırma yöntemi olmak üzere farklı DMF yöntemleri kullanmıştır. Araştırmanın sonucuna göre, 500 örneklem büyüklüğünde, olabilirlik oran testi ve madde parametre karşılaştırma yöntemlerinin işlev farklılığı saptamada güçsüz olduğu, örneklem sayısının artması ile her iki yöntemin benzer sonuçlar verdiğini gözlemlemiş olup; buna karşın madde test işlev farklılık yönteminin tüm örneklem genişliklerinde daha tutarlı olduğu sonucuna varmıştır. Araştırma sonucunda biri erkekler lehine, diğeri kızlar lehine olmak üzere iki maddede önemli düzeyde DMF görülmüştür (Aktaran: Atalay, 2010).

Stoneberg, B. D.(2006), çalışmasında 4, 8 ve 10. sınıflarda okuma, dil ve matematikten oluşan 3 standart başarı testi üzerinde cinsiyete ve etnik farklılıklara göre DMF belirleme üzerinde durmuştur. Bunun için SIBTEST ve MH istatistiklerini kullanmıştır. Ayrıca tip 1 hata oranını kontrol etmek için Bonferroni düzeltmesini uygulamıştır. Araştırmanın sonucunda cinsiyete dayalı DMF'nin olduğu maddelerin oranı 4.sınıf okuma testi için %7'den 10.sınıf matematik testi için %37'ye; etnik farklılıklara dayalı DMF içeren maddelerin oranı 8. sınıf matematik testi için %7'den 4. sınıf okuma ve dil testi için %19'a kadar uzandığı görülmüştür. Stoneberg (2006) araştırmasında, öğrencilerin sınavdan elde ettikleri puanları doğru bir şekilde

yorumlamak için DMF gösteren maddelerin program uzmanları tarafından incelenmesi gerektiğini önermiştir (Aktaran: Bekçi, 2007).

Ross ve Okabe (2006), İngilizce'yi yabancı dil olarak öğrenenlere uygulanan dört pasaj, yirmi okuma sorusundan oluşan bir dil testinin cinsiyet yanlılığı içerip içermediğine bakan araştırmalarını beş Japon enstitüsünden, 468 kız öğrenciyi odak, 357 erkek öğrenciyi referans grup olarak belirleyerek yürütmüşlerdir. Madde yanlılığına, objektif ve subjektif yöntemler olarak bakılmışlardır. Objektif yöntemler olarak, Mantel-Haenszel (MH) delta-plot, Lojistik regresyon ve Simultane yöntemleri, Subjektif yöntem olarak ise uzman görüşlerini kullanmışlardır. Sonuç olarak, testte konu ve cinsiyet yanlılığı olduğunu, sübjektif yöntemle madde yanlılığı belirleme yönteminin, daha fazla maddeyi yanlı gösterdiğini saptamışlardır.

Le (2006) PISA 2006 ön uygulaması fen bilimleri maddelerinin farklı ülke, dil ve cinsiyet grupları üzerinde değişen madde fonksiyonu 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 olabileceğini ortaya koymuştur.

Yıldırım (2006), TIMSS-1999 ve PISA 2003 sınavlarındaki matematik okuryazarlığını ölçen testlerde kültürler arası denkliğin ve matematik başarısında kültürlerle ait yapıların olup olmadığını, bu testlerin Türkçe ve İngilizce versiyonlarındaki maddelerinin, psikometrik özellikleri ve maddelerin farklı dildeki testlerinde yanlı çalışmasının olası sebeplerini araştırmıştır. Yanlı çalışan maddelerin tespitinde, Sınırlandırılmış Faktör Çözümlemesi, Mantel-Haenszel, ve Madde Tepki Kuramı En Çok Olabilirlik Oran Analizi yöntemlerini kullanmıştır. Araştırma sonuçlarına göre, PISA verilerinde madde yanlılığı tespitinde kullanılan farklı metotların sonuçları arasında yüksek derecede uyum tespit edilirken, TIMSS verilerinde, bu uyum görülmemiştir. Ayrıca, bilgi düzeyinde ve rutin işlem becerisi gerektiren maddelerde, aynı yetenek düzeyindeki Türk ve Amerikalı öğrenciler için yanlılık bulunmamıştır.

Öğretmen (2006), uluslararası okuma becerilerinde gelişim projesi (PIRLS) 2001 testi kapsamında uygulanan okuma parçalarına ait testlerin Türkiye ve ABD örnekleminde, kültürler arası DMF gösterip göstermediğini incelemiştir. Uygulanan okuma parçalarına ait testler üzerinde MTK parametre karşılaştırma ve olabilirlik oranı testi model karşılaştırma yöntemleriyle DMF analizleri yapmıştır. Her iki yöneme göre de PIRLS testlerine ait maddelerin çoğunun kültürel karşılaştırmada farklı fonksiyonlaştığı ortaya çıkmıştır.

Grisay ve Monseur (2007) PISA 2000 ve 2001 uygulaması ile okuma becerileri maddelerinin değişik dil grupları için farklı fonksiyonlaşıp fonksiyonlaşmadığını araştırmışlardır. Araştırma sonucu, kültürel ve coğrafi farklılıklar minimize edilse bile çeviri testlerin DMF’li madde sayılarının aynı dili konuşan ülkeler arasındaki DMF’li madde sayısından fazla olmasıdır.

Wu, Li ve Zumbo (2007) TIMSS 1999 verisiyle matematik testinin kültürler arasındaki ölçme değişmezliğini incelemişlerdir. Sonuç olarak matematik testini oluşturan maddelerin benzer kültürlerde farklı fonksiyonlaşmadığını; farklı kültürlerde ise farklı fonksiyonlaştığını dolayısıyla bu ülkeler için testin ölçme değişmezliğini sağlamadığını bulmuşlardır (Aktaran: Asil, 2010).

Bekçi (2007), çalışmasında MEB’in 2005 yılında yapmış olduğu Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı’nın değişen madde fonksiyonlarını cinsiyete ve okul türüne göre incelemiştir. Araştırma, OKS’nin fen bilgisi, matematik, türkçe ve sosyal bilgiler alt testine cevap veren 777137 kişi arasından rastgele seçilen ve evreni temsil ettiği düşünülen örneklemelerden elde edilen veriler üzerinden yürütülmüştür. DMF belirlemek için MH ve LR yöntemleri kullanılmıştır. Elde edilen bulgular doğrultusunda, sosyal bilgiler alt testinde DMF bulunmadığı; türkçe alt testinde ise, okul türü alt grubuna göre DMF içeren 2 maddenin bulunduğu; aynı şekilde fen bilgisi alt testinde okul türü alt grubuna göre, 1 maddede DMF bulunduğu, matematik alt testinde de 4 maddede DMF olduğu gözlenmiştir. Elde edilen DMF’lerin madde etkisinden mi, yoksa madde yanlılığından mı kaynaklandığı araştırılmıştır. Bunun için

konu alanı uzmanlarından görüşler alınmıştır. Sonuçta OKS’de yer alan maddelerin yanlışlık içermediği ortaya çıkmıştır.

Feng (2008), Alberta’ da 12. sınıflara uygulanan iki ve çok kategorili maddelerden oluşan diploma sınavını konu alanı ve cinsiyete göre DMF içerip içermediğini Poly-SIBTEST yöntemiyle incelemiştir. Çalışmanın sonunda, yapılandırılmış soruların kızların lehine çalıştığını ortaya koymuştur.

Acar (2008), çalışmasında 2006 yılında uygulanmış olan Ortaöğretim Kurumları Sınavı’nın, alt test maddelerine cinsiyete ve bölge alt gruplarına göre genelleştirilmiş aşamalı doğrusal modelleme, lojistik regresyon, madde tepki kuramı olabilirlik oranı tekniklerini karşılaştırarak değişen madde fonksiyonu analizi yapmıştır. Tüm alt testlerde cinsiyete ve bölge alt gruplarına göre LR yöntemi ile ihmal edilebilir düzeyde, MTK-OO yöntemi ile de bazılarında orta ve yüksek düzeyde DMF çıktığı sonucuna varmıştır. Tekniklerin DMF belirleme düzeyleri bakımından aralarındaki ilişkiye baktığında her bir alt test için cinsiyete göre MTK-OO ile LR arasında anlamlı bir ilişki olduğunu görmüştür. Bölge alt gruplarına göre GDAM ile LR, MTK-OO ile LR ve GDAM ile MTK-OO arasında anlamlı ilişkilerin farklı testler arasında olduğunu ortaya koymuştur.

Kalaycıoğlu (2008), 2005 yılı Öğrenci Seçme Sınavında (ÖSS) yer alan maddelerin cinsiyet ve okul türü değişkenlerine göre farklı fonksiyonlaşıp fonksiyonlaşmadığını incelemiştir. Araştırmada, Mantel-Haenszel (MH) ve Lojistik Regresyon (LR) yöntemlerini kullanmış ve farklı fonksiyonlaştığı belirlenen maddelerin yanlış olup olmadığına karar vermek için uzman görüşüne başvurmuştur. Araştırmada ÖSS’ye giren 599,330 lise son sınıf öğrencisinin her biri 45 maddeden oluşan Türkçe, sosyal bilimler, matematik ve fen bilimleri alt testlerine verdikleri cevaplar kullanılmıştır. İndeks değerleri göz önüne alındığında, MH ve LR yöntemlerinin birbiriyle uyumlu sonuçlar vermekte olduğu gözlenirken; literatürdeki kriter puanlar göz önüne alındığında, bu değerlerin LR yöntemi için yüksek olduğu saptanmıştır. Bu sebeple, çalışmada, yöntemler birbirini tamamlayacak şekilde farklı bir kriter belirlenmiş ve bu kritere uyan maddelerin farklı fonksiyonlaştığı kabul edilmiştir.

Araştırmada elde edilen bulgular doğrultusunda, cinsiyete göre alt testler göz önüne alındığında, türkçe alt testini oluşturan maddelerin farklı fonksiyonlaşmadığı, sosyal bilimler alt testinde 7, matematik ve fen bilimleri alt testlerinde de 3 maddenin farklı fonksiyonlaştığı ve bu maddelerden de 1 fen bilimleri maddesinin yanlılık gösterdiği belirlenmiştir. Okul türlerine göre karşılaştırma yapıldığında, matematik alt testini oluşturan maddelerin farklı fonksiyonlaşmadığı, türkçe alt testinde; genel lise Anadolu lisesi karşılaştırmasında 14, genel lise özel lise karşılaştırmasında 9 ve Anadolu lisesi özel lise karşılaştırmasında 9 maddenin; sosyal bilimler alt testinde, genel lise Anadolu lisesi karşılaştırmasında 16, genel lise özel lise karşılaştırmasında 1 ve Anadolu lisesi özel lise karşılaştırmasında 6 maddenin; fen bilimleri alt testinde genel lise Anadolu lisesi karşılaştırmasında ise 1 maddenin farklı fonksiyonlaştığı ve bu maddelerden de sadece 1 fen bilimleri maddesinin yanlılık gösterdiği belirlenmiştir.

Doğan N. Ve Öğretmen T. (2008), yapmış oldukları çalışmada DMF belirleme tekniklerinden ki-kare, Mantel Haenszel (MH) ve lojistik regresyon (LR) tekniklerini karşılaştırarak uygulamada ortaya çıkan benzerlik ve farklılıkları incelemişlerdir. Çalışma 2003 yılında Ortaöğretim Kurumları Seçme ve Yerleştirme Sınav'ına katılan yaklaşık 600.000 öğrenci arasından yansız olarak seçilen 3345 öğrenciden elde edilen fen bilgisi alt testine verilen cevaplar üzerinden yürütülmüştür. Cinsiyete göre yapılan DMF analizleri sonucunda, tekniklere göre DMF veren madde sayısı 0 ile 15 arasında değişmektedir. Söz konusu tekniklere göre elde edilen yanlı madde oranları arasında manidar bir fark bulunduğu ortaya koyulmuştur. DMF'li madde sayısı en fazla MH tekniğinde elde edilmiştir. İşaretsiz Ki-kare tekniği 9 madde ile ikinci sıradadır. İşaretli Ki-kare ve LR tekniğinde ise hiçbir madde için DMF ortaya çıkmamıştır. LR ve işaretli Ki-kare tekniklerinin DMF'li madde kararı bağlamında benzer sonuçları ürettiği gösterilmiştir. Doğan ve Öğretmen sonuç olarak, MH, Ki-kare ve lojistik regresyon tekniklerinin Ki-kare değerlerinin büyüklüğü bakımından benzer; DMF verdiği belirlenen madde sayısı bakımından farklı sonuçlar ürettiğini ortaya koymuşlardır.

Smith (2009), WASL matematik testinin İngilizce öğrenen öğrencilere yanlı olup olmadığını belirlemeye çalışmıştır. İki yıl boyunca 4. ve 7. sınıflardan toplanan veriler üzerinde madde ve küme düzeyinde DMF'yi MH ve SIBTEST yöntemleri ile

incelemiştir. Araştırmada, 4. sınıflarda ana dili İngilizce olan ve İngilizce öğrenen öğrenciler arasında matematik puanlarının madde formatında DMF gösterdiği, 7. sınıflarda ise var olan farklılığın matematik yeteneğindeki gerçek farktan kaynaklandığı sonucuna varmıştır.

Atalay (2010), PISA 2006 öğrenci anketinde yer alan tutum maddelerinin cinsiyete ve farklı kültürlere karşı değişen madde fonksiyonu içerip içermediğini incelemiştir. Araştırmanın çalışma grubunu, PISA uygulamasına katılan 57 ülke içinden 5611 öğrenci ile Amerika Birleşik Devletleri ve 4942 öğrenci ile Türkiye örneklemi oluşturmaktadır. Çalışmada DMF belirlemek için SLR ve Poly-SIBTEST yöntemleri kullanılmıştır. Araştırmanın sonucunda, Türkiye verisinde cinsiyete göre yapılan analizlerde SLR yöntemi ile 12 maddede, poly-SIBTEST yöntemi ile 13 maddede DMF gözlenmiştir. ABD verisinde cinsiyete göre yapılan analizlerde ise her iki yöntemde 15 maddede DMF gözlenmiştir. Her iki yöntemde ortak ve önemli düzeyde DMF gösteren madde sayısı Türkiye’de 7, ABD’ de 13’tür. İki ülke arasında kültüre göre her iki yöntemde yapılan analizlerde 30 maddede önemli düzeyde DMF görülmüştür. Türkiye ve ABD verisinde cinsiyet ve bu iki ülke arasında kültüre göre yapılan analizlerde iki yöntem karşılaştırıldığında, yöntemlerin DMF belirlemede bir uyum gösterdiği; ancak düzeylerinde farklılıklar olabileceği sonucuna varılmıştır.

Asil (2010), yaptığı çalışmada PISA 2006 kapsamında yer alan öğrenci anketinin kültürler ve diller arası eşdeğerliğini, Avusturya, Yeni Zelanda, Amerika ve Türkiye örneklemi üzerinde karşılaştırmalı olarak incelemiştir. Araştırmada çoklu grup doğrulayıcı faktör analizi kullanarak anket maddelerinin farklı kültür ve diller arasında DMF gösterip göstermediğini incelenmiştir ve DMF görülen maddelerin DMF göstermesinin nedenleri uzman görüşlerine başvurularak belirlenmeye çalışılmıştır. Çalışmada, ülkeler arasında dilsel ve kültürel farklılıklar artıkça DMF gösteren madde sayısının arttığı gözlenmiştir. Ayrıca maddelerin DMF göstermesinin nedenlerinin çeviri problemleri ve kültürel farklılıklar olduğu sonucuna varılmıştır.

BÖLÜM III

3. Yöntem

Bu bölümde, araştırmanın türü, çalışma grubu, verilerin elde edilmesi ve verilerin analizine ilişkin bilgilere yer verilmiştir.

3.1 Araştırmanın Türü

Araştırma PISA 2006 öğrenci anketinin farklı kültürel ve dilsel özelliklere göre değişen madde fonksiyonlarının incelenmesi yönüyle betimsel; DMF belirleme yöntemlerinden sıralı lojistik regresyon ve Poly-SIBTEST tekniklerini karşılaştırması yönüyle de kuramsal özellik taşımaktadır.

3.2. Evren ve Örneklem

PISA 2006 uygulamasına 30'u OECD'ye üye olan 57 ülke katılmıştır ve bu ülkelerdeki yaklaşık 20 milyon 15 yaş grubu öğrenciyi temsil eden 400.000 öğrenci tesadüfi olarak seçilmiştir. PISA 2006 uygulamasında ülkelere ait örneklemeler iki basamaklı tabakalı örnekleme yöntemiyle belirlenmiştir.

Çalışmada ülkeler arasında maddelerin değişen madde fonksiyonu göstermesinin kültürel farklılıklardan mı yoksa uyarlama problemlerinden mi kaynaklandığını ortaya koymak için Avustralya, Yeni Zelanda, ABD ve Türkiye örneklemeleri amaçlı olarak ele alınmıştır. Dil farklılıkları açısından DMF analizi İngilizce ve Türkçe dilleri üzerinden yapılmıştır. Bu dört ülke örnekleminin ele alınmasının nedeni hem kültürel ve dilsel benzerlik ve farklılıkları açısından karşılaştırılabilir olması hem de daha önceki yıllarda bu ülkeler üzerinde benzer

çalışmaların yapılmış olması ve çıkan sonuçların önceki araştırmalarla karşılaştırılmasına olanak sağlamasıdır. Bu doğrultuda DMF analizlerinde aynı kültür-aynı dil (AUS-NZL), farklı kültür-aynı dil (AUS-ABD) ve farklı kültür farklı-dil (AUS-TUR) karşılaştırmaları yapılmıştır. Avustralya hem örneklem büyüklüğünün fazla olması hem de PISA ölçme araçlarının geliştirilmesinden sorumlu olan konsorsiyumun liderliğini yapmasından dolayı referans grup olarak; diğerleri ise odak grup olarak belirlenmiştir.

PISA 2006 kapsamında Avustralya örneklemini 14170, Yeni Zelanda örneklemini 4823, ABD örneklemini 5605 ve Türkiye örneklemini 4942 öğrenciden oluşmaktadır. Öğrencilerin kültürel ve dilsel özelliklerini tam yansıtmaları açısından testin uygulandığı ülkede doğan ve evlerinde konuşulan dilin testin uygulandığı ülke ile aynı olan öğrencilerin verileri analize alınmış diğerleri analiz dışı bırakılmıştır. Analizde kullanılacak yöntemler kayıp ve aykırı verilerle çalışmayacağından dolayı bu verilerin ayıklanması işlemi sonucunda çalışmaya katılacak olan öğrenci grubuna ait frekans ve yüzdeler Tablo 3-1. 'de verilmiştir.

Tablo 3.1: Çalışma grubuna ait frekans ve yüzdeler

Gruplar	F	%
Avustralya	11644	48.1
Yeni Zelanda	3503	14.5
ABD	4519	18.7
Türkiye	4537	18.7

3.3. Veri Toplama Araçları ve Verilerin Elde Edilmesi

Veri toplama aracı olarak PISA 2006 öğrenci anketi kapsamında ölçülmeye çalışılan, öğrencilerin fen bilimlerine yönelik tutumlarından birisi olan “bilimsel sorgulamaya verilen destek” ölçme modeli kullanılmıştır.

“Bilimsel sorgulamaya verilen destek” ölçme modeli her biri beşer maddeli olan fen bilimlerine verilen genel değer ve fen bilimlerine verilen kişisel değer alt boyutlarından oluşmaktadır. “Bilimsel sorgulamaya verilen destek” modelindeki ST18Q01, ST18Q02, ST18Q04, ST18Q06, ST18Q09 kodlu maddeler “fen bilimlerine verilen genel değer” boyutunu; ST18Q03, ST18Q05, ST18Q07, ST18Q08, ST18Q10 kodlu maddeler ise “fen bilimlerine verilen kişisel değer” boyutunu oluşturmaktadır. Modeli oluşturan tüm maddeler dörtlü likert tipindedir ve “tümüyle katılıyorum” 1, “katılıyorum” 2, “katılmıyorum” 3, ve “hiç katılmıyorum” 4 biçiminde puanlandırılmıştır.

PISA 2006 öğrenci anketinde, bilimsel sorgulamaya verilen destek tutumunu ölçen maddelerin Avustralya, ABD ve Yeni Zelanda örneklemelerinde kullanılan İngilizce versiyonu Ek 1’de, Türkiye örnekleminde kullanılan Türkçe versiyonu ise Ek 2’de verilmiştir.

PISA, öğrenci performansını değerlendiren ve öğrenci performanslarının farklılığını açıklamak için öğrenci, aile ve okul etkenleri üzerinden veri toplayan en kapsamlı ve en detaylı uluslararası programdır. Değerlendirmenin yapısı, kapsamı ve anketlerle toplanacak bilgiyle ilgili kararlar, katılan ülkelerdeki uzmanlar tarafından verilmekte, ülkelerin ortak politikaları doğrultusunda yönlendirilmektedir. PISA uygulamalarında kullanılan ölçme araçlarının tüm ülkelerde geçerli ve güvenilir olmasını sağlamak, kültürel ve dilsel farklılıkları en aza indirmek için oldukça yoğun bir çaba ve kaynak sarf edilmektedir. Çeviri, örneklem oluşturma ve veri toplama işlemlerinde uyulması zorunlu olan kalite güvence mekanizmaları bulunmaktadır. Sonuç olarak, PISA sonuçlarının geçerlik ve güvenilirlik derecesi yüksektir ve bu sonuçlar eğitim çıktılarının anlaşılmasına önemli derecede katkı sağlamaktadır (MEB, 2007).

Araştırmada kullanılan, Ekonomik İşbirliği ve Kalkınma Örgütü’nün (OECD) düzenlemiş olduğu PISA 2006 uygulamasından elde edilen öğrenci anketinin verileri OECD PISA internet sitesinden elde edilmiştir.

3.4. Verilerin Çözümlemesi

Verilerin çözümlemesinde aşağıdaki adımlar izlenmiştir.

1.DFA sayıtlılarının test edilmesi

Doğrulayıcı faktör analizi uygulama sürecine başlamadan önce ortaya çıkabilecek sorunları aza indirmek amacıyla veri seti için bazı sayıtlılar test edilmiştir. Bu amaçla her bir ülke verisine ilişkin kayıp değer, aşırı uç değer, tekli bağlantılılık, çoklu doğrusal bağlantı ve doğrusallık sayıtlıları test edilmiştir.

2. Ölçme modeli ve uyumunun incelenmesi

Doğrulayıcı faktör analizi, belirli bir faktör doğrultusunda ölçme yaptığı düşünülen ve aralarındaki ilişkinin daha önceden belirli bir kurama ya da önceki araştırmalara göre belirlendiği maddelerin bu belirlenmiş ilişkiler doğrultusunda ölçme yapıp yapmadığını belirlemeyi amaçlar. Asil (2010), yapmış olduğu doktora tezinde bilimsel sorgulamaya destek ölçme modelini her biri beşer maddeden oluşan fen bilimlerinin genel değerleri ve fen bilimlerinin kişisel değerleri olmak üzere iki faktörlü bir yapı olarak ortaya koymuştur. PISA 2006 öğrenci anketindeki 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. Modeli oluşturan tüm maddeler dörtlü likert tipindedir ve “tümüyle katılıyorum” 1, “katılıyorum” 2, “katılmıyorum” 3, ve “hiç katılmıyorum” 4 biçiminde puanlandırılmıştır. Ancak DMF analizinde kullanılan yöntemler için puanlamanın 0’dan başlaması gerektiğinden dolayı veriler 0-3 aralığında yeniden kodlanmıştır. Aynı zamanda yorumlama kolaylığı açısından maddelerin kodlaması tersine çevrilerek “hiç katılmıyorum” 0, “katılmıyorum” 1, “katılıyorum” 2 ve “tümüyle katılıyorum” 3 olacak şekilde puanlandırılmış böylelikle yüksek değerlerin daha olumlu tutumu belirtmeleri sağlanmıştır.

Farklı dillere uyarlaması yapılan maddelerin değişen madde fonksiyonlarının belirlenmesinde önce ölçülen faktör yapısının araştırmaya katılan her bir grup için incelenmesi gerekmektedir (Asil,2010). Bu doğrultuda çalışmada bilimsel sorgulamaya verilen destek ölçme modelinin birleştirilmiş veri seti ve her bir ülke veri seti için ayrı ayrı DFA analizi yapılmış modelin uyumu test edilmiştir. Doğrulayıcı faktör analizinde kullanılan komut dosyasının örneği Ek 8’ de verilmiştir.

DFA analizi ile modele ait genel uyum indeksleri, faktörlerin güvenirlikleri, faktörler arası korelasyonlar ve her bir maddenin ilişkili olduğu faktör tarafından açıklanan varyansları incelenmiştir.

Çalışmada PISA 2006 uygulamasına ilişkin teknik raporlar doğrultusunda model veri uyumu için χ^2 (ki-kare), RMSEA, CFI ve NNFI indeksleri rapor edilmiştir (OECD, 2009). Ayrıca χ^2 ’nin örneklem büyüklüğünden etkilendiği göz önünde bulundurularak χ^2 ’ye alternatif olarak model uyumunu örneklem büyüklüğünden bağımsız olarak değerlendiren GFI indeksi de çalışmada rapor edilmiştir (Büyüköztürk, 2010).

Faktörlerin güvenirlikleri Cronbach alpha değerinin yanı sıra Hair, Anderson, Tatham ve Black (1998)’in önerdiği yapı güvenirliği (construct reliability) ve açıklanan varyans oranı (variance extracted) değerleri ile de test edilmiştir. Yapı güvenirliği değerleri maddelerin hatalardan arınık olarak hesaplanan iç tutarlılığını, açıklanan varyans değerleri ise maddelerin, ilişkili oldukları faktörler tarafından açıklanan toplam varyansını ifade etmektedir (Asil, 2010). Nunally (1978) ile Hair ve arkadaşları (1998), ölçek güvenirliği için güvenirlik ölçümüne ilişkin tahminlerin %70’in ve açıklanan varyans ölçümüne ilişkin tahminin ise 0.50’nin üstünde olması gerektiğini vurgulamaktadırlar (Aktaran: Yılmaz ve Çelik, 2009). Çalışmada bu değerleri hesaplamak için 1 ve 2’ deki formüller kullanılmıştır.

$$\text{Yapı Güvenirliği} = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum \varepsilon_j} \quad (1)$$

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

λ_i = her bir maddeye ait standartlaştırılmış faktör yükü

ε_j = her bir maddeye ait hata varyansı

3. Ölçeğin her bir alt boyutu için betimsel istatistikler hesaplanmıştır.

4. Kültürel ve dilsel farklılıklara göre değişen madde fonksiyonu SLR ve poly-SIBTEST yöntemleri ile incelenmiştir. Analizler sonucu maddelerin hangi düzeyde DMF gösterdiği ve hangi grubun lehine çalıştığı belirtilerek tablolastırılmıştır.

Çalışmada betimsel istatistikler ve DFA sayıltıları SPSS 15.0 ile, DFA analizi LISREL 8.80 ile, DMF için yapılan SLR analizi Zumbo'nun (1999) yazdığı betik ile test edilmiştir. Betik Ek 13' de verilmiştir. Poly-SIBTEST için Sheally ve Stout (1993) tarafından hazırlanan bilgisayar programları kullanılmıştır.

BÖLÜM IV

4. Bulgular ve Yorum

Bu bölümde DFA sayıtlarının test edilmesi, ölçme modeli ve uyumunun incelenmesi, alt gruplara ilişkin madde istatistiklerinin incelenmesi ve çalışmanın araştırma sorularına ilişkin bulgular ve yorumlar yer almaktadır. PISA 2006 öğrenci anketinde yer alan ve öğrencilerin fen bilimlerine karşı tutumlarını ölçen aracın bir boyutu olan bilimsel sorgulamaya verilen destek ölçme modelinin maddelerinin kültürel ve dilsel farklılıklara göre değişen madde fonksiyonu gösterip göstermediği sıralı lojistik regresyon ve poly-SIBTEST yöntemleri ile incelenmiştir.

4.1. DFA sayıtlarının test edilmesi

Her bir ülke veri grubundaki değişkenlere ilişkin kayıp veri değerleri Ek 3’de verilmiştir. Her bir değişkendeki kayıp veri oranı %5’ten az olduğu ve sistematik bir yapı göstermediği için analiz dışı bırakılmışlardır (Tabacknick ve Fidell, 2007, aktaran: Asil, 2010).

Aşırı uç değerlerin tespit edilmesi amacıyla her bir ülke verisi için katılımcıların Mahalanobis uzaklık değerleri hesaplanmıştır. Hiç bir veri grubunda ± 3.5 ss üzerinde Mahalanobis uzaklığı değerine sahip olan katılımcı bulunmamıştır. Dolayısıyla veri setinin çok değişkenli aşırı uç değer içermediği sonucuna varılmıştır.

Değişkenler arasındaki korelasyonların çok yüksek olmaması ve LISREL programının analiz yapılırken bir uyarı vermemesi dolayısıyla veri setinde tekli bağlantılılık olmadığı sonucuna varılmıştır (Tabacknick ve Fidel,2007, aktaran: Asil, 2010). Çoklu doğrusal bağlantı, değişkenler arasındaki korelasyonların ($r > .90$ ve üzeri)

yüksek olması durumunda ortaya çıkmaktadır ve değişkenlerin modele katkılarını ya da etkilerini ayırt etmeyi güçleştirir (Büyüköztürk, 2010). Değişkenler arasındaki çoklu bağlantılılığın incelenmesi için CI (koşul indeksi), VIF (varyans şişme oranları ve tolerans değerleri incelenmiştir. Değişkenlere ait CI değerinin 30 ve üzeri, VIF değerinin 10'dan büyük veya tolerans değerlerinin .10 veya üzerinde olması durumunda çoklu bağlantılılığın olduğu söylenebilir (Hair, Anderson, Tahtam ve Black, 1998). Her bir ülke verisine uygulanan çoklu bağlantılılık analizi sonucunda CI değerlerinin 30'dan, VIF değerlerinin 10'dan küçük olduğu ve tolerans değerlerinin de sıfırdan oldukça farklı olduğu ve sonucuna varılmıştır. Her bir ülke verisine ilişkin sonuçlar Ek 4, 5, 6, 7'de verilmiştir.

Doğrusallık sayıtlısı, her bir ülkenin veri setinin standartlaştırılmış artıklarla standartlaştırılmamış tahmini değerlerin dağılımlarının incelenmesi ile test edilmiştir. Böylece doğrusallık sayıtlısı da karşılanmıştır.

4.2. Araştırma Sorusu1' e İlişkin Bulgular ve Yorumlar

PISA 2006 öğrenci anketi kapsamında ölçülen “bilimsel sorgulamaya verilen destek” ölçme modeli Avustralya, Yeni Zelanda, ABD ve Türkiye kültürlerinde ve dillerinde aynı faktör yapısına sahip midir?

Araştırma kapsamında incelenen bilimsel sorgulamaya verilen destek ölçme modelini oluşturan fen bilimlerine verilen genel değer ve fen bilimlerine verilen kişisel değer faktörlerine dair betimsel istatistikler Tablo 4-1. ve Tablo 4-2' de verilmiştir.

Tablo 4-1.Fen bilimlerine verilen genel değer faktörünü oluşturan maddelere ait betimsel istatistikler

	ST18Q01	ST18Q02	ST18Q04	ST18Q06	ST18Q09	TOPLAM
Avustralya						
$\bar{x}$	2.193	2.283	2.036	2.162	1.777	10.453
Ss	0,645	0.612	0.633	0.647	0.739	2.472
Yeni Zelanda						
$\bar{x}$	2.150	2.240	2.048	2.086	1.724	10.249
Ss	0,635	0.613	0.625	0.669	0.730	2.405
ABD						
$\bar{x}$	2.323	2.370	2.153	2.244	1.945	11.037
Ss	0.653	0.619	0.670	0.656	0.746	2.551
Türkiye						
$\bar{x}$	2.501	2.414	2.203	2.460	2.289	11.869
Ss	0.616	0.620	0.734	0.637	0.685	2.422

Not. Maddeler 4 dereceli Likert tipindedir. (0= hiç katılmıyorum, 1= katılmıyorum, 2= katılıyorum, 3= tümüyle katılıyorum).

Tablo 4-1'e göre fen bilimlerine verilen genel değeri ölçen maddelerin aritmetik ortalamalarına bakıldığında tüm ülke öğrencilerinin katılıyorum veya tümüyle katılıyorum düzeyinde görüş bildirdikleri görülmektedir. Buna göre tüm ülkelerde öğrencilerin fen bilimlerine karşı göstermiş oldukları genel değer yargılarının olumlu olduğu söylenebilir. Ülke bazında incelendiğinde fen bilimlerine genel olarak en fazla değer veren öğrenciler Türk öğrencilerdir. Türk öğrencileri sırasıyla ABD'li ve Avustralyalı öğrenciler takip etmektedir. Yeni Zelandalı öğrenciler ise en düşük genel ortalamayla fen bilimlerine karşı genel değerleri en düşük olan öğrencilerdir. ST08Q09 kodlu madde diğer maddelere göre Türkiye dışındaki tüm ülkelerde en düşük aritmetik ortalamaya sahip olan maddedir. Bu durum Türk öğrencilerin diğer ülkelerin öğrencilerine göre fen bilimleri ve teknolojiye ilerlemelerin genellikle toplumsal yararlar sağladığı görüşüne daha çok katıldıklarını göstermektedir. Bunun yanı sıra Türkiye dışındaki diğer ülkelerin öğrencilerinin bilim ve teknolojik ilerlemelerin topluma yarar sağladığı düşüncesinin fen bilimlerine verdikleri genel değer içerisinde en az önemsedikleri durum olduğu söylenebilir.

Tablo 4-2. Fen bilimlerinin kişisel değeri faktörünü oluşturan maddelere ait betimsel istatistikler

	ST18Q03	ST18Q05	ST18Q07	ST18Q08	ST18Q10	TOPLAM
AVUSTRALYA						
$\bar{x}$	1.675	1.737	1.613	1.871	1.676	8.574
Ss	0.720	0,823	0.852	0.767	0.877	3.267
YENİ ZELANDA						
$\bar{x}$	1.635	1.725	1.578	1.894	1.713	8.548
Ss	0.703	0.824	0.852	0.765	0.881	3.172
ABD						
$\bar{x}$	1.865	1.897	1.806	2.022	1.900	9.492
Ss	0.731	0.807	0.817	0.747	0.809	3.098
TÜRKİYE						
$\bar{x}$	2.026	2.083	1.807	2.079	1.672	9.670
Ss	0.728	0.772	0.867	0.750	0.888	3.047

Not. Maddeler 4 dereceli Likert tipindedir. (0= hiç katılmıyorum, 1= katılmıyorum, 2= katılıyorum, 3= tümüyle katılıyorum).

Tablo 4-2'ye göre fen bilimlerine verilen kişisel değeri ölçen maddelerin aritmetik ortalamalarına bakıldığında tüm ülke öğrencilerinin katılıyorum veya katılmıyorum aralığında görüş bildirdikleri görülmektedir. Ülke bazında genel ortalamalara göre Türk öğrencilerin en yüksek ortalama ile fen bilimlerine kişisel olarak daha fazla değer verdiği söylenebilir. Türk öğrencileri ABD'li öğrenciler takip ederken, Avustralyalı ve Yeni Zelandalı öğrencilerin ortalamalarının daha düşük olduğu ve bu ülkelerin fen bilimlerine verdikleri kişisel değerlerin daha olumsuz olduğu sonucuna varılabilir. Ayrıca ST18Q07 kodlu madde diğer maddelere göre Türkiye dışındaki tüm ülkelerde en düşük aritmetik ortalamaya sahip olan maddedir. Bu durum Türkiye dışındaki ülkelerin öğrencilerinin fen bilimlerine verdikleri kişisel değeri ölçen maddeler içinde fen bilimlerine karşı olan ilgilerinin daha az olduğunu göstermektedir.

Tablo 4-1 ve Tablo 4-2 birlikte değerlendirildiğinde tüm ülkelerdeki öğrencilerin fen bilimlerine verdikleri genel değer ortalamalarının fen bilimlerine verdikleri kişisel değer ortalamalarından yüksek olduğu görülmekle beraber her bir

faktör için ortalamalar ve standart sapmalar da öğrencilerin tutumlarının birbirine yakın olduğunu ortaya koymaktadır.

DMF analizinden önce bilimsel sorgulamaya verilen destek ölçme modelinin faktör yapısının her bir kültür ve dilde geçerliliğini ortaya koymak için birleştirilmiş veri ve her bir ülke verisi için DFA analizi yapılmıştır. DFA sonucu elde edilen faktör yükleri Ek 9, 10, 11, 12’de verilmiştir. Faktör yükleri Avustralya örnekleme için 0.54-0.88; Yeni Zelanda örnekleme için 0.47-0.85; ABD örnekleme için 0.58-0.81 ve Türkiye örnekleme için 0.58-0.76 arasında değer almıştır. Ölçme modeline ait uyum indeksleri Tablo 4-3’de verilmiştir.

Tablo 4-3. Bilimsel sorgulamaya verilen destek ölçme modeli uyum değerleri

	χ^2	Sd	RMSEA	CFI	NNFI	GFI	SRMR
Birleşik veri	11306,94*	34	0.090	0.97	0.96	0.92	0.060
Avustralya	2679,89*	34	0.080	0.97	0.96	0.95	0.052
Yeni Zelanda	1160,66*	34	0.082	0.97	0.95	0.94	0.058
ABD	1276,43*	34	0.078	0.97	0.97	0.95	0.046
Türkiye	1014,24*	34	0.076	0.97	0.96	0.96	0.045

* $p < 0,01$

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.

Bilimsel sorgulama verilen destek ölçme modelinin uyumu χ^2 ve indeks değerlerine göre incelenmiştir. Analizler sonucunda modele ilişkin uyum indeksleri Tablo 4-3’de verilmiştir. χ^2 örneklem büyüklüğüne duyarlıdır ve örneklem büyüklüğü arttıkça χ^2 değeri de büyümektedir (Büyüköztürk, 2010). Bu durum göz önünde bulundurularak modelin uyumuna ilişkin kararlar diğer indeks değerlerine bakılarak verilmiştir. Tabachnick ve Fidell’e (2011) göre CFI ve NNFI değerlerinin; Hooper, Coughlan ve Mullen’e (2008) göre de GFI değerinin 0.90’ın üzerinde olması iyi uyuma; 0.95’in üzerinde olması modelin mükemmel uyuma sahip olduğunun göstergesidir (Aktaran: Büyüköztürk, 2010). Tablo 4-3’ deki CFI, NNFI ve GFI değerleri bilimsel sorgulamaya verilen destek ölçme modelinin her bir ülke ve birleşmiş veri seti için iyi

veya mükemmel bir uyuma sahip olduğunu göstermektedir. Brown (2006); Hu ve Bentler (1999) SRMR değerinin ≤ 0.08 olduğu durumlarda modelin iyi uyum gösterdiğini belirtmişlerdir (Aktaran: Büyüköztürk, 2010). Tablo 4-3’de her bir veri seti için SRMR değeri 0.08’in altında kalmıştır. Hooper, Coughlan ve Mullen (2008) RMSEA değerinin ≤ 0.08 olduğu durumda modelin iyi uyum gösterdiğini belirtmişlerdir (Aktaran: Büyüköztürk, 2010). Birleşik veri ve Yeni Zelanda verisi dışındaki diğer veriler için RMSEA indeksine göre model iyi uyuma sahiptir. Bilimsel sorgulamaya verilen destek ölçme modeline ait uyum indekslerinin hepsi birlikte değerlendirildiğinde modelin her bir veri seti için uyumlu olduğu söylenebilir.

Modelin her bir ülkedeki uyumunu değerlendirmek için modeli oluşturan faktörlerin güvenirlikleri ve faktörler arası ilişkiler de incelenmiştir. Bilimsel sorgulamaya verilen destek ölçme modelini oluşturan fen bilimlerine verilen genel değer ve fen bilimlerine verilen kişisel değer faktörlerinin her bir ülkedeki güvenirlikleri ve faktörlerin aralarındaki korelasyonlar Tablo 4-4’de verilmiştir.

Tablo 4-4. Fen bilimlerinin genel değeri ve fen bilimlerinin kişisel değeri faktörlerinin güvenirlikleri

	<u>Yapı güvenirliği</u>		<u>Açıklanan varyans</u>		Faktörler arası korelasyon
	GD	KD	GD	KD	
Avustralya	0.81	0.87	0.46	0.62	0.78
Yeni Zelanda	0.78	0.85	0.42	0.55	0.76
ABD	0.82	0.85	0.48	0.54	0.81
Türkiye	0.80	0.82	0.44	0.48	0.77

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

Tablo 4-4’e göre yapı güvenirliği değerleri fen bilimlerine verilen genel değer faktörü için 0.78 ile 0.82 arasında değer alırken, fen bilimlerine verilen kişisel değer faktörü için 0.82 ile 0.87 arasında değer almıştır. Her bir ülke için yapı güvenirliği istenilen değer olan 0.70’in üstündedir (Hair, Anderson, Tahtam ve Black, 1998). Açıklanan varyans oranları fen bilimlerinin genel değer faktörü için 0.42 ile 0.48 arasında değer alırken, fen bilimlerine verilen kişisel değer faktörü içinse 0.48 ile 0.62 arasında değer almıştır. Açıklanan varyanslar da önerilen düzey olan 0.50’ye yakın

değerler vermiştir (Hair, Anderson, Tahtam ve Black, 1998). Her bir ülke verisi içinde fen bilimlerine verilen kişisel değer faktörünün güvenilirliğinin fen bilimlerine verilen genel değer faktörünün güvenilirliğinden daha yüksek olduğu görülmektedir. Avustralya örnekleminin hem fen bilimlerine verilen genel değer hem de fen bilimlerine verilen kişisel değer faktörlerinin güvenilirlikleri diğer ülkelerden yüksektir. Faktörler arası korelasyonlar incelendiğinde ise fen bilimlerine verilen genel değer ve fen bilimlerine verilen kişisel değer faktörlerinin tüm ülkelerde 0,70'in üzerinde korelasyon değerine sahip olduğu dolayısıyla faktörlerin oldukça ilişkili olduğu gözlenmiştir. Faktörler arasındaki korelasyonların ABD örnekleminde en yüksek düzeyde iken diğer ülkelerde birbirlerine yakın değerler aldıkları görülmektedir.

Schulz'a (2000) göre model uyumunun bir başka göstergesi de madde güvenilirliğidir (Aktaran: Asil, 2010). Madde güvenilirlikleri LISREL çıktısındaki her bir maddeye ait (R^2) değerleridir. Fen bilimlerinin genel değerleri ve fen bilimlerinin kişisel değerleri faktörlerine ait maddelerin güvenilirlikleri sırasıyla Tablo 4-5 ve Tablo 4-6'da verilmiştir.

Tablo 4-5. Fen bilimlerinin genel değeri faktörünü oluşturan maddelerin güvenilirlikleri					
	ST18Q01	ST18Q02	ST18Q04	ST18Q06	ST18Q09
Avustralya	0.42	0.50	0.40	0.58	0.40
Yeni Zelanda	0.39	0.46	0.38	0.53	0.39
ABD	0.43	0.51	0.42	0.61	0.42
Türkiye	0.44	0.45	0.36	0.47	0.45

Tablo 4-5 'e göre fen bilimlerine verilen genel değer faktörünün her bir maddedeki varyansın yaklaşık %50'sini açıkladığı görülmektedir. ST18Q06 ve ST18Q02 kodlu maddelerin bu faktörün en iyi temsilcileri olduğu söylenebilir. ST18Q01, ST18Q04, ST18Q09 maddelerinin açıkladıkları varyans oranlarının ise bazı ülke örnekleminde %40'ın altına düştüğü görülmektedir.

Tablo 4-6. Fen bilimlerinin kişisel değeri faktörünü oluşturan maddelerin güvenirlikleri

	ST18Q03	ST18Q05	ST18Q07	ST18Q08	ST18Q10
Avustralya	0.38	0.69	0.76	0.55	0.64
Yeni Zelanda	0.32	0.68	0.72	0.50	0.60
ABD	0.44	0.61	0.65	0.55	0.55
Türkiye	0.43	0.53	0.58	0.52	0.43

Tablo 4-6'ye göre fen bilimlerinin kişisel değer faktörünün her bir maddedeki varyansın yaklaşık %50'sini açıkladığı görülmektedir. ST18Q07 kodlu maddenin bu faktörün en iyi temsilcisi olduğu söylenebilir. ST18Q03 kodlu maddenin ise açıklanan varyans değerinin Avustralya ve Yeni Zelanda örneklemelerinde %40'ın altına düşmüş olduğu görülmektedir. Bu maddenin güvenirliğinin düşük olması maddenin ifade biçiminde bir problem olabileceğini veya faktörü iyi temsil etmediğini düşündürülebilir.

Bilimsel sorgulamaya verilen destek ölçme modelinin güvenirliği Cronbach Alpha ile de test edilmiştir. Sırasıyla bilimsel sorgulamaya verilen destek ölçme modelinin ve onu oluşturan fen bilimlerine verilen genel değer ile fen bilimlerine verilen kişisel değer faktörlerinin Cronbach alpha değerleri Avustralya örnekleme için .89, .81 ve .86; Yeni Zelanda örnekleme için .88, .78, .84; ABD örnekleme için .89, .82, .85; Türkiye örnekleme için .86, .78, .81'dir. Cronbach alpha değerlerine göre bilimsel sorgulamaya verilen destek ölçme modeli her bir ülke örnekleminde iyi derecede güvenilirdir. Fen bilimlerine verilen genel değer ve fen bilimlerine verilen kişisel değer faktörlerinin de iyi derecede güvenilir olduğu görülmektedir. Ayrıca bu faktörlerin her bir ülke örnekleme için hesaplanan Cronbach alpha değerleri ile yapı güvenirliği değerlerinin de birbirine oldukça yakın olduğu sonucuna varılmıştır.

Tukey toplanabilirlik testi ile de ölçekteki maddelerin toplamsal ölçek oluşturacak şekilde bir niteliğe sahip olup olmadığı test edilmiştir. Bilimsel sorgulamaya verilen destek ölçme modelinin Avustralya, Yeni Zelanda, ABD ve Türkiye örnekleme ait Tukey toplanabilirlik testi sonuçlarına göre ($F_{(9,11643)} = 2380.77$; $F_{(9,3502)} = 615.40$; $F_{(9,4518)} = 648.55$; $F_{(9,4536)} = 1037.06$ $p < .01$)

ölçekte yer alan soruların toplamsal ölçek oluşturacak bir niteliğe sahip olduğu sonucuna varılmıştır.

Fen bilimlerine verilen genel değer ve kişisel değer faktörlerinin Avustralya, Yeni Zelanda, ABD ve Türkiye örneklemelerine ait Tukey toplanabilirlik test sonuçlarına göre (GD: $F_{(4,11643)} = 1913.89$; $F_{(4,3502)} = 541.59$; $F_{(4,4518)} = 538.50$; $F_{(4,4536)} = 277.166$ $p < .01$ KD: $F_{(4,11643)} = 389.85$; $F_{(4,3502)} = 160.35$; $F_{(4,4518)} = 98.41$; $F_{(4,4536)} = 450.31$) her bir alt ölçekte yer alan soruların toplamsal ölçek oluşturacak bir niteliğe sahip olduğu sonucuna varılmıştır.

DMF analizi öncesinde ölçme modelinin faktör yapısının her bir grup için ayrı ayrı incelemesi yapılmış ve betimsel istatistikler; DFA analizi sonucunda elde edilen uyum indeksleri, faktör ve madde güvenirlikleri incelendiğinde bilimsel sorgulamaya verilen destek ölçme modelinin her bir kültür ve dil için iki faktörlü bir yapıya sahip olduğu sonucuna varılmıştır.

4.3. Araştırma Sorusu 2'ye İlişkin Bulgular

4.3.1 Araştırma sorusu 2a'ya ilişkin bulgular

Araştırma Sorusu 2a: Pisa 2006 Öğrenci Anketi kapsamında ölçülen “Bilimsel Sorgulamaya Verilen Destek” ölçme modelinin orijinal ve uyarlanan maddeleri Avustralya ve Yeni Zelanda örneklemeleri arasında sıralı lojistik regresyon ve Poly-SIBTEST yöntemleri ile yapılan analizlerde değişen madde fonksiyonu göstermekte midir? Yöntemler arasında uyum var mıdır?

Sıralı Lojistik Regresyon Sonuçları

Tablo 4-7. Avustralya ve Yeni Zelanda örneklemelerinde SLR yöntemi ile yapılan DMF sonuçları

Boyut	Maddeler	DMF χ^2	P	Model1 R^2	Model2 R^2	Model3 R^2	DMF ΔR^2	Düzyey	DMF Türü
GD	ST18Q01	0.6	0.741	69.15	69.15	69.16			
	ST18Q02	1.37	0.504	70.51	70.52	70.52			
	ST18Q04	35.69*	0.000	66.47	66.62	66.62	0.15	B	TB
	ST18Q06	20.47*	0.000	69.37	69.45	69.45	0.08	A	TB
	ST18Q09	1.04	0.595	64.81	64.81	64.81			
KD	ST18Q03	20.377*	0.000	47.62	47.68	47.70	0.08	A	TBO
	ST18Q05	1.48	0.471	82.23	82.24	82.23			
	ST18Q07	12.53*	0.002	85.60	85.63	85.63	0.03	A	TB
	ST18Q08	9.92	0.007	72.52	72.54	72.54			
	ST18Q10	18.22*	0.000	78.42	78.45	78.45	0.03	A	TB

* $p < 0.005$

Avustralya ve Yeni Zelanda örneklemeleri üzerinde yapılan bilimsel sorgulamaya verilen destek ölçme modeli maddelerine ilişkin SLR yöntemi ile yapılan DMF analiz sonuçları Tablo 4-7’de verilmiştir. Tablo 4-7’de verilen Model1 R^2 Model2 R^2 ve Model3 R^2 değerleri SLR analiz çıktısından elde edilmiştir. DMF χ^2 değeri ise SLR çıktısındaki 3.aşamada elde edilen χ^2 değerinden 1. aşamada elde edilen χ^2 değerinin çıkartılması ile bulunmuştur. Ek 14’ de ST18Q01 maddesine ait SLR analiz çıktısı verilmiştir. DMF χ^2 değerinin anlamlılık düzeyi SLR analizi ile bulunamamaktadır. Bunun için her bir maddeye ait χ^2 farklılıklarının anlamlılığının değerlendirilmesinde Bonferroni düzeltmesi kullanılmıştır. Buna göre 10 maddeden elde edilen düzeltilmiş anlamlılık seviyesi ($\alpha_{\text{düzeltmiş}} = \frac{0,05}{10}$) 0.005’dir. DMF χ^2 değerinin anlamlılığını test etmek için gerekli p değeri ise χ^2 calculator ile hesaplanmış ve ST18Q04, ST18Q06, ST18Q03, ST18Q07 ve ST18Q010 maddelerinin χ^2 değerinin anlamlı olduğu görülmüştür. Dolayısıyla bu maddelerde DMF’nin varlığından söz edilebilir. DMF’nin etki düzeyine bakıldığında ise ST18Q04 kodlu maddenin B düzeyinde yani orta derecede DMF gösterdiği görülmektedir. ST18Q04 dışındaki maddelerin DMF düzeyleri ise A düzeyinde olduğu için bu maddelerde DMF yoktur

veya ihmal edilebilir. Bu maddelerdeki TB ve TBO DMF türleri ise serbestlik derecesi 2 olan χ^2 testi ile tespit edilmiştir (Swaminathan ve Rogers, 1990). TB- DMF, ikinci model ile birinci modelden elde edilen χ^2 değerleri farkı, TBO-DMF ise üçüncü model ile ikinci modelden elde edilen χ^2 değerleri farkı ile belirlenmiştir. Buna göre 4,6,7 ve 10. maddelerde tüm yetenek düzeyinde alt gruplardan birinin devamlı üstünlüğünün olduğunu ortaya koyan tek biçimli (TB); 3. maddede ise farklı yetenek düzeylerinde farklı alt grupların üstünlüğünün olduğunu ortaya koyan tek biçimli olmayan (TBO) DMF'nin varlığı tespit edilmiştir.

Avustralya ve Yeni Zelanda örneklemelerine uygulanan SLR analizinde tek bir maddenin B düzeyinde; diğer 4 maddenin yukarıdaki tabloda da görüldüğü gibi A düzeyinde yani ihmal edilebilir düzeyde DMF içerdiği görülmektedir. Bu ülkelerde aynı dilin konuşulması ve benzer kültürel özelliklere sahip olunması maddelerin her iki ülkedeki öğrenciler tarafından benzer şekilde algılandığını ve bundan ötürü yüksek düzeyde DMF çıkmadığını ortaya koymaktadır.

Poly- SIBTEST sonuçları

Tablo 4-8. Avustralya ve Yeni Zelanda örneklemelerinde poly-SIBTEST yöntemi ile yapılan DMF sonuçları

Boyutlar	Maddeler	Beta estimate	Standart e	p	Düzyey	Ülke
GD	ST18Q01	0.018	0.014	0.212		
	ST18Q02	0.008	0.014	0.540		
	ST18Q04	-0.083	0.015	0.000	B	YENİ Z.
	ST18Q06	0.032	0.015	0.033	A	AUS
	ST18Q09	0.003	0.018	0.879		
KD	ST18Q03	0.020	0.018	0.276		
	ST18Q05	0.006	0.016	0.720		
	ST18Q07	0.037	0.015	0.016	A	AUS
	ST18Q08	-0.023	0.016	0.150		
	ST18Q10	-0.046	0.018	0.010	A	YENİ Z.

Referans grup: Avustralya, Odak grup: Yeni Zelanda

Avustralya ve Yeni Zelanda örneklemeleri üzerinde yapılan bilimsel sorgulamaya verilen destek ölçme modeli maddelerine ilişkin poly-SIBTEST yöntemi ile yapılan DMF analiz sonuçları Tablo 4-8’ de verilmiştir. ST18Q01 maddesine ilişkin poly-SIBTEST analiz çıktısı Ek 15’de verilmiştir. Tablo 4-8’e bakıldığında Bilimsel Sorgulamaya Verilen Destek Ölçme Modeli maddelerinden 4 tanesinin beta değerinin 0.05 düzeyinde anlamlı olması bu maddelerde DMF’nin var olduğunu göstermektedir. Ancak bu maddeler incelendiğinde fen bilimlerine verilen genel değer faktöründeki ST18Q04 kodlu maddenin B yani orta düzeyde, yine aynı faktördeki ST18Q06 ve fen bilimlerine verilen kişisel değer faktöründe yer alan ST18Q07 ile ST18Q10 maddelerinde ise A düzeyi yani ihmal edilebilir düzeyde DMF olduğu görülmektedir. Maddelerin hangi ülke lehine farklı fonksiyonlaştığına bakılacak olursa ST18Q04 ile ST18Q10 kodlu maddelerin odak grup olan Yeni Zelanda lehine, ST18Q06 ile ST18Q07 kodlu maddelerin referans grup olan Avustralya lehine farklı fonksiyonlaştığı görülmektedir.

Avustralya ve Yeni Zelanda örneklemelerine uygulanan poly-SIBTEST analizinde tek bir maddenin B düzeyinde; diğer 3 maddenin A düzeyinde yani ihmal edilebilir düzeyde DMF içerdiği görülmektedir. Bu ülkelerde aynı dilin konuşulması ve benzer kültürel özelliklere sahip olunmasından ötürü maddelerin her iki ülkedeki öğrenciler tarafından benzer şekilde algılandığı ve bundan ötürü yüksek düzeyde DMF çıkmadığı söylenebilir.

Yöntemler arası uyum

Tablo 4-9. Avustralya ve Yeni Zelanda örneklemeleri arasında SLR ve poly-SIBTEST analizlerinin karşılaştırılması

	A	B	C
SLR	ST18Q03	ST18Q04	
	ST18Q06		
	ST18Q07		
	ST18Q10		
Poly-SIBTEST	ST18Q06	ST18Q04	
	ST18Q07		
	ST18Q10		

Tablo 4-9 incelendiğinde Avustralya ve Yeni Zelanda örneklemeleri arasında SLR yöntemiyle 5 maddede poly SIBTEST yöntemiyle 4 maddede DMF olduğu gözlemlenmektedir. Her iki yöntemde de ST18Q04 kodlu madde B yani orta düzeyde, ST18Q06, ST18Q07 ve ST18Q10 kodlu maddeler ise A yani hafif veya ihmal edilebilir düzeyde DMF göstermişlerdir. ST18Q03 kodlu madde ise sadece SLR yöntemi ile A düzeyinde DMF göstermektedir. Bu madde DMF düzeyinin düşük ya da ihmal edilebilir olmasından ötürü Poly-SIBTEST yöntemi ile yapılan analizde DMF’li olarak çıkmamış olabilir ya da SLR’ nin DMF belirlemede daha hassas olduğu söylenebilir.

Sonuç olarak Avustralya ve Yeni Zelanda örneklemelerinde Poly-SIBTEST yöntemi ile DMF’li çıkan her madde SLR yöntemi ile de aynı düzeyde farklı fonksiyonlaşmıştır. Bu durum yöntemlerin uyumlu olduğu sonucunu ortaya koymaktadır.

4.3.2 Araştırma Sorusu 2b’ye ilişkin bulgular

Araştırma Sorusu 2b: Pisa 2006 Öğrenci Anketi kapsamında ölçülen “Bilimsel Sorgulamaya Verilen Destek” ölçme modelinin orijinal ve uyarlanan maddeleri Avustralya ve ABD örneklemeleri arasında sıralı lojistik regresyon ve Poly-SIBTEST yöntemleri ile yapılan analizlerde değişen madde fonksiyonu göstermekte midir? Yöntemler arasında uyum var mıdır?

Sıralı lojistik regresyon sonuçları

Tablo 4-10. Avustralya ve ABD örneklemelerinde SLR yöntemi ile yapılan DMF sonuçları

Boyut	Maddeler	DMF χ^2	p	Model1 R^2	Model2 R^2	Model3 R^2	DMF ΔR^2	Düzyey	DMF Türü
GD	ST18Q01	7.82	0.020	69.52	69.55	69.56			
	ST18Q02	5.1	0.078	71.54	71.55	71.54			
	ST18Q04	11.15*	0.004	67.44	67.43	67.50	0.06	A	TBO
	ST18Q06	17.62*	0.000	70.86	70.92	70.93	0.07	A	TB
	ST18Q09	17.77*	0.000	65.65	65.68	65.68	0.03	A	TB
KD	ST18Q03	78.217*	0.000	50.98	51.11	51.30	0.32	C	TBO
	ST18Q05	27.79*	0.000	81.36	81.42	81.43	0.07	A	TB
	ST18Q07	15.66*	0.000	85.01	85.01	85.04	0.03	A	TBO
	ST18Q08	17.19*	0.000	73.31	73.32	73.38	0.07	A	TBO
	ST18Q10	21.97*	0.000	77.83	77.84	77.87	0.04	A	TBO

* $p < 0.005$

Avustralya ve ABD örneklemeleri üzerinde, bilimsel sorgulamaya verilen destek ölçme modeli maddelerine ilişkin SLR yöntemi ile yapılan DMF analiz sonuçları Tablo 4-10'da verilmiştir. Bilimsel sorgulamaya verilen destek ölçme modelini oluşturan fen bilimlerine verilen genel değer ve fen bilimlerine verilen kişisel değer maddelerine ait χ^2 farklılıkları Bonferroni düzeltmesi uygulanmış anlamlılık düzeyinde test edilmiş ve ST18Q01 ve ST18Q02 dışındaki tüm maddelerin DMF içerdiği sonucuna varılmıştır. Ancak DMF içerdiği sonucuna varılan bu maddelerden sadece ST18Q03 kodlu madde C düzeyinde yani yüksek düzeyde DMF gösterirken diğer maddeler A düzeyinde yani hafif ya da göz ardı edilebilir düzeyde DMF göstermişlerdir. Bu maddelerdeki TB ve TBO DMF türleri ise serbestlik derecesi 2 olan χ^2 testi ile incelendiğinde 4. 3. 7. 8. ve 10. maddelerde farklı yetenek düzeylerinde farklı alt grupların üstünlüğünün olduğunu ortaya koyan tek biçimli olmayan (TBO); 6.9. ve 5. maddelerde ise tüm yetenek düzeyinde alt gruplardan birinin devamlı üstünlüğünün olduğunu ortaya koyan tek biçimli (TB) DMF olduğu görülmektedir.

Avustralya ve ABD örneklemelerine uygulanan SLR analizinde ST18Q03 kodlu tek bir maddenin C düzeyinde; diğer 7 maddenin yukarıdaki tabloda da görüldüğü gibi A düzeyinde yani ihmal edilebilir düzeyde DMF içerdiği görülmektedir. Bu ülkelerde aynı dil konuşulmasına rağmen kültürel özellikler farklı olduğundan dolayı DMF gösteren madde sayısının artmış olabileceği düşünülebilir. Ayrıca DMF gösteren maddelerin Avustralya ve ABD örneklemelerinde aynı biçimde anlaşılmadığı da söylenebilir.

Poly- SIBTEST sonuçları

Tablo 4-11 Avustralya ve ABD örneklemelerinde poly-SIBTEST yöntemi ile yapılan DMF sonuçları

Boyutlar	Maddeler	Beta estimate	Standart	p	Düzyey	Ülke
			e			
GD	ST18Q01	0.003	0.015	0.837		
	ST18Q02	0.045	0.010	0.001	A	AUS
	ST18Q04	0.020	0.015	0.182		
	ST18Q06	0.059	0.014	0.000	B	AUS
	ST18Q09	-0.055	0.018	0.002	A	ABD
KD	ST18Q03	-0.093	0.018	0.000	C	ABD
	ST18Q05	0.071	0.016	0.000	B	AUS
	ST18Q07	0.045	0.015	0.003	A	AUS
	ST18Q08	0.044	0.016	0.006	A	AUS
	ST18Q10	0.015	0.018	0.394		

Referans grup: Avustralya , Odak grup: ABD

Avustralya ve ABD örneklemeleri üzerinde yapılan bilimsel sorgulamaya verilen destek ölçme modeli maddelerine ilişkin poly-SIBTEST yöntemi ile yapılan DMF analiz sonuçları Tablo 4-11’de verilmiştir. Tablo 4-11’e bakıldığında Bilimsel Sorgulamaya Verilen Destek Ölçme Modeli maddelerinden 7 tanesinin beta değerinin 0.05 düzeyinde anlamlı olması bu maddelerde DMF’nin var olduğunu göstermektedir. DMF’li maddeler incelendiğinde fen bilimlerine yönelik kişisel değer faktöründeki ST18Q03 kodlu maddenin C yani yüksek düzeyde, ST18Q05 kodlu maddenin B yani

orta düzeyde, ST18Q07 ve ST18Q08 kodlu maddelerin ise A yani hafif veya ihmal edilebilir düzeyde oldukları; fen bilimlerine yönelik genel değer faktöründeki ST18Q06 kodlu maddenin B yani orta düzeyde, ST18Q02 ve ST18Q09 kodlu maddelerinse A yani hafif veya ihmal edilebilir düzeyde oldukları görülmektedir. Maddelerin hangi ülke lehine farklı fonksiyonlaştığına bakılacak olursa ST18Q03 ile ST18Q09 kodlu maddelerin odak grup olan ABD lehine, diğer DMF'li maddelerin referans grup olan Avustralya lehine farklı fonksiyonlaştığı görülmektedir.

Genel olarak poly-SIBTEST sonucuna bakıldığında 7 maddeden 1 tanesi C düzeyinde, 2 tanesi B düzeyinde ve geri kalan 4 tanesi ise A düzeyinde DMF göstermişlerdir. A düzeyindeki DMF ihmal edilebilmekle beraber, DMF'li madde sayısındaki artışın aynı dilin konuşulduğu Avustralya ve ABD ülkelerindeki kültürel farklılıklardan kaynaklandığı düşünülebilir.

Yöntemler arası uyum

Tablo 4-12. Avustralya ve ABD örneklemi arasında SLR ve poly-SIBTEST analizlerinin karşılaştırılması

	A	B	C
SLR	ST18Q04		ST18Q03
	ST18Q05		
	ST18Q06		
	ST18Q07		
	ST18Q08		
	ST18Q09		
	ST18Q10		
Poly-SIBTEST	ST18Q02	ST18Q05	ST18Q03
	ST18Q07	ST18Q06	
	ST18Q08		
	ST18Q09		

Tablo 4-12 incelendiğinde Avustralya ve ABD örneklemelerinde SLR yöntemiyle 8, poly-SIBTEST yöntemiyle 7 maddede DMF olduğu gözlenmektedir. Her iki yöntemde ortak DMF gösteren madde sayısı 6'dır. Bunlardan ST18Q07, ST18Q08, ve ST18Q09 kodlu üç madde her iki yöntemde de A düzeyinde; ST18Q03 kodlu madde ise her iki yöntemde C düzeyinde farklı fonksiyonlaşmıştır. SLR yöntemiyle A düzeyinde DMF gösteren ST18Q05 ve ST18Q06 kodlu maddeler poly-SIBTEST yönteminde B

düzeyinde DMF göstermişlerdir. Bu sonuca göre poly-SIBTEST tekniğinin DMF belirleme de daha hassas olduğu söylenebilir.

Bununla beraber ST18Q04 ve ST18Q10 kodlu maddeler SLR yöntemiyle A düzeyinde DMF göstermiş ancak poly-SIBTEST yönteminde farklı fonksiyonlaşmamışlardır. Aynı şekilde ST18Q02 kodlu madde poly-SIBTEST yönteminde A düzeyinde DMF gösterirken SLR yönteminde farklı fonksiyonlaşmamıştır. Ancak bu maddelerdeki DMF düzeyinin ihmal edilebilir seviyede olması yöntemlerin uyumu için sorun teşkil etmeyebilir.

Her iki yöntemde ortak DMF gösteren madde sayısı 6 ‘dır. Bu 6 maddeden 4 ü yani %67’sinin aynı düzeyde DMF göstermektedir. Diğer iki madde SLR yönteminde A düzeyinde DMF gösterirken poly-SIBTEST yönteminde B düzeyinde farklı fonksiyonlaşmıştır. Bu durum Avustralya-ABD örnekleme üzerinde poly-SIBTEST’in daha ayrıntılı sonuç verdiğini gösterebilir. Ancak genel olarak her iki yöntem de DMF’li maddeleri bulmakta uyum göstermektedirler.

4.3.3 Araştırma Sorusu 2c’ye ilişkin bulgular

Araştırma Sorusu 2c: Pisa 2006 Öğrenci Anketi kapsamında ölçülen “Bilimsel Sorgulamaya Verilen Destek” ölçme modelinin orijinal ve uyarlanan maddeleri Avustralya ve Türkiye örneklemleri arasında sıralı lojistik regresyon ve Poly-SIBTEST yöntemleri ile yapılan analizlerde değişen madde fonksiyonu göstermekte midir? Yöntemler arasında uyum var mıdır?

Sıralı lojistik regresyon sonuçları

Tablo 4-13. Avustralya ve Türkiye örneklemelerinde SLR yöntemi ile yapılan DMF sonuçları

Boyut	Maddeler	DMF χ^2	p	Model1 R^2	Model2 R^2	Model3 R^2	DMF ΔR^2	Düzyey	DMF Türü
GD	ST18Q01	125.25	0.000	87.06	87.49	87.37	0.31	C	TBO
	ST18Q02	329.2	0.000	87.86	88.61	88.51	0.65	C	TBO
	ST18Q04	380.02	0.000	79.16	79.69	79.68	0.52	C	TB
	ST18Q06	47.33	0.000	86.35	86.47	86.34	0.01	A	TBO
	ST18Q09	1268.24	0.000	74.70	76.68	76.66	1.96	C	TB
KD	ST18Q03	453.61	0.000	56.70	58.27	58.43	1.73	C	TBO
	ST18Q05	239.61	0.000	82.98	83.38	83.41	0.43	C	TBO
	ST18Q07	1513.53	0.000	81.12	84.05	84.05	2.93	C	TB
	ST18Q08	1200.64	0.000	75.99	78.68	78.69	2.7	C	TBO
	ST18Q10	1568.82	0.000	72.42	76.31	76.39	3.97	C	TBO

* $p < 0.005$

Avustralya ve Türkiye örneklemeleri üzerinde yapılan bilimsel sorgulamaya verilen destek ölçme modeli maddelerine ilişkin SLR yöntemi ile yapılan DMF analiz sonuçları Tablo 4-13’de verilmiştir. Bilimsel sorgulamaya verilen destek ölçme modelini oluşturan fen bilimlerine verilen genel değer ve fen bilimlerine verilen kişisel değer maddelerine ait χ^2 farklılıkları, Bonferroni düzeltmesi uygulanmış anlamlılık düzeyinde test edilmiş ve tüm maddelerin DMF içerdiği sonucuna varılmıştır. Aynı zamanda DMF içeren bu maddelerden sadece ST18Q06 kodlu madde A düzeyinde yani hafif ya da göz ardı edilebilir düzeyde DMF gösterirken diğer maddeler C düzeyinde yani yüksek düzeyde DMF göstermişlerdir. Bu maddelerdeki TB ve TBO DMF türleri ise serbestlik derecesi 2 olan χ^2 testi ile incelendiğinde 1.,2.,6.,3.,5.,8. ve 10. maddelerde farklı yetenek düzeylerinde farklı alt grupların üstünlüğünün olduğunu ortaya koyan tek biçimli olmayan (TBO); 4.9. ve 7. maddelerde ise tüm yetenek düzeyinde alt gruplardan birinin devamlı üstünlüğünün olduğunu ortaya koyan tek biçimli (TB) DMF olduğu görülmektedir.

Avustralya ve Türkiye örneklemeleri arasında tüm maddeler SLR yöntemi ile DMF göstermiştir. Bu durumun hem ülkelerdeki kullanılan dilin farklılığından hem de kültürel farklılıklardan ortaya çıkmış olma olasılığı yüksektir. Ve maddelerin Avustralya ve Türkiye'deki öğrenciler için aynı anlama gelmediği sonucuna varılabilir.

Poly- SIBTEST sonuçları

Tablo 4-14. Avustralya ve Türkiye örneklemelerinde poly-SIBTEST yöntemi ile yapılan DMF sonuçları

Boyutlar	Maddeler	Beta estimate	Standart e	p	Düzye	Ülke
GD	ST18Q01	-0.229	0.069	0.001	C	TUR
	ST18Q02	-0.389	0.059	0.000	C	TUR
	ST18Q04	-0.431	0.081	0.000	C	TUR
	ST18Q06	-0.105	0.070	0.131		
	ST18Q09	-0.635	0.071	0.000	C	TUR
KD	ST18Q03	0.273	0.025	0.000	C	AUS
	ST18Q05	0.111	0.022	0.000	C	AUS
	ST18Q07	-0.597	0.020	0.000	C	TUR
	ST18Q08	0.345	0.021	0.000	C	AUS
	ST18Q10	-0.536	0.023	0.000	C	TUR

Referans grup: Avustralya, Odak grup: Türkiye

Tablo 4-14 incelendiğinde Avustralya ve Türkiye örneklemeleri arasında yapılan poly-SIBTEST analizine göre 10 maddelik bilimsel sorgulamaya verilen destek ölçme modelinde ST18Q06 kodlu madde dışındaki tüm maddelerde beta değeri anlamlıdır ve hepsi C yani yüksek düzeyde DMF göstermektedir. Maddelerin hangi ülke lehine farklı fonksiyonlaştığına bakılacak olursa ST18Q03, ST18Q05 ve ST18Q08 kodlu maddelerin referans grup olan Avustralya lehine, diğer DMF'li maddelerin ise odak grup olan Türkiye lehine farklı fonksiyonlaştığı görülmektedir. Bu durumun hem ülkelerdeki kullanılan dilin farklılığından hem de kültürel farklılıklardan ortaya çıkmış olma olasılığı yüksektir. Ve maddelerin Avustralya ve Türkiye'deki öğrenciler için aynı anlama gelmediği sonucuna varılabilir.

Yöntemler Arası Uyum

Tablo 4-15. Avustralya ve Türkiye örneklemeleri arasında SLR ve poly-SIBTEST analizlerinin karşılaştırılması

	A	B	C
SLR	ST18Q06		ST18Q01
			ST18Q02
			ST18Q03
			ST18Q04
			ST18Q05
			ST18Q07
			ST18Q08
			ST18Q09
			ST18Q10
Poly-SIBTEST			ST18Q01
			ST18Q02
			ST18Q03
			ST18Q04
			ST18Q05
			ST18Q07
			ST18Q08
			ST18Q09
			ST18Q10

Tablo 4-15 incelendiğinde Avustralya Türkiye örneklemeleri arasında bilimsel sorgulamaya verilen destek ölçme modeli maddelerinin SLR yöntemine göre hepsi, poly-SIBTEST yöntemine göre ise 9 tanesi DMF göstermektedir. Poly-SIBTEST yönteminde DMF gösteren her madde SLR yöntemi ile de farklı fonksiyonlaşmıştır. Ve bu 9 maddenin hepsi her iki yöntemde de C yani yüksek düzeyde DMF göstermişlerdir. Poly-SIBTEST yönteminden farklı olarak ST18Q06 kodlu madde SLR yöntemi ile A düzeyinde yani ihmal edilebilir düzeyde DMF göstermektedir. Poly-SIBTEST yönteminde bu maddenin DMF göstermemesinin nedeni DMF düzeyinin ihmal edilebilir seviyede olmasından kaynaklanabilir. Avustralya Türkiye örneklemeleri arasında yapılan bu iki analizin sonuçlarının oldukça yüksek derecede uyum gösterdikleri görülmektedir.

BÖLÜM V

5.Sonuç ve Öneriler

Bu bölümde, her bir alt probleme ilişkin sonuçlar özetlenmiş ve bulgular ışığında, yeni çalışmalara yönelik önerilere yer verilmiştir.

5.1 Sonuçlar

Bu araştırmada PISA 2006 öğrenci anketinde yer alan bilimsel sorgulamaya verilen destek ölçme modelini oluşturan maddelerin DMF analizi yapılmıştır. Bilimsel sorgulamaya verilen destek ölçme modeli fen bilimlerine verilen genel değer 5 madde ve fen bilimlerine verilen kişisel değer 5 madde olmak üzere toplam 10 maddeden oluşmaktadır. DMF analizinde dil (İngilizce- Türkçe) ve kültürel farklılıklar temel alınarak 4 ülke üzerinde SLR ve poly-SIBTEST yöntemleri ile analiz yapılmıştır.

5.1.1 Araştırma sorusu 1'e ilişkin sonuçlar

Fen bilimlerine yönelik genel değer ve fen bilimlerine yönelik kişisel değer faktörlerinin betimsel istatistikleri incelenmiş ve her bir ülke örneklemindeki öğrencilerin fen bilimlerine genel ve kişisel olarak değer verdikleri sonucuna varılmıştır. Her bir faktör için ülkelerin ortalamalarının birbirine yakın olduğu görülmüştür. Türkiye her iki faktörde de en yüksek ortalamaya sahip ülkedir.

Bilimsel sorgulamaya verilen destek ölçme modelini oluşturan faktörlerin yapısının her bir ülke verisi için aynı olup olmadığını belirlemek için yapılan doğrulayıcı faktör analizi sonucunda ölçme modelinin genel uyumu RMSEA, CFI, NNFI ve SRMR indeks değerleri ile incelenmiş ve birleştirilmiş veri ve tüm ülke

verileri için deęerlerin kabul edilebilir seviyede olduęu tespit edilmiřtir. Dolayısıyla iki faktörlü ölçme modelinin faktör yapısının her bir ülke örnekleminde geçerli olduęu kabul edilmiřtir. Bilimsel sorgulamaya verilen destek ölçme modelini oluřturan faktörlerin güvenilirlikleri ve açıklanan varyans oranları kabul edilebilir düzeyde çıkmıřtır. Tüm ülke örneklemlerinde her iki faktör arasındaki korelasyon deęerleri incelendięinde, faktörler arası iliřkinin de benzer olduęu görölmüřtür. Her bir maddenin iliřkili oldukları faktörler tarafından açıklanan varyans miktarları deęerlendirildięinde, her bir maddenin varyansın genelde %40 ile %60 arasında açıklandığı ortaya konulmuřtur. Birinci alt problem sonucunda ölçme modelinin her bir ülke için iki faktörlü olduęu ve maddelerin faktörlerle benzer örüntüye sahip olduęu sonucuna varılmıřtır.

5.1.2 Arařtırma sorusu 2'ye iliřkin sonuçlar

Arařtırma sorusu 2a'ya iliřkin sonuçlar

Bilimsel sorgulamaya verilen destek ölçme modeline ait maddeler için Avustralya ve Yeni Zelanda örneklemleri arasında yapılan DMF analizleri sonucunda SLR yöntemiyle 5, poly-SIBTEST yöntemi ile 4 maddenin farklı fonksiyonlařtığı ortaya çıkmıřtır. Poly-SIBTEST yöntemi ile DMF gösteren her madde SLR yöntemi ile de farklı fonksiyonlařmış ve her iki yöntemle de aynı düzeyde DMF göstermiřlerdir. Poly-SIBTEST'den farklı tek bir madde SLR yöntemi ile farklı fonksiyonlařmıřtır o da A düzeyinde yani ihmal edilebilir düzeydedir. Buna göre SLR ve Poly-SIBTEST yöntemlerinin oldukça iyi uyum gösterdikleri sonucuna varılmıřtır. Ayrıca Avustralya Yeni Zelanda arasında yapılan DMF analizinde ST18Q04 kodlu tek bir madde B düzeyinde tek biçimli ve Yeni Zelanda lehine DMF göstermiř olup dięer maddeler A düzeyinde yani ihmal edilebilir düzeyde DMF göstermiřlerdir. Bu iki ülke arasında önemli düzeyde DMF gösteren madde sayısının 1 olması ülkelerin benzer dil ve kültürel özellikler gösterdiklerinden yani öğrencilerin maddelere aynı anlamı yüklediklerinden kaynaklı olabileceęi sonucuna varılmıřtır.

Araştırma sorusu 2b'ye ilişkin sonuçlar

Bilimsel sorgulamaya verilen destek ölçme modeline ait maddeler için Avustralya ve ABD örneklemeleri arasında yapılan DMF analizleri sonucunda SLR yöntemiyle 8, poly-SIBTEST yöntemi ile 7 maddenin farklı fonksiyonlaştığı ortaya çıkmıştır. Her iki analizde de DMF gösteren madde sayısı 6'dır ve bunların 4'ü aynı düzeyde DMF göstermişlerdir. Bu 4 maddeden biri olan ST18Q03 kodlu madde her iki yöntemde de C düzeyinde farklı fonksiyonlaşmıştır; ABD lehine ve tek boyutlu olmayan DMF göstermiştir. Diğer 3 madde her iki yöntemde A yani ihmal edilebilir düzeyde DMF göstermişlerdir. Her iki yöntemde DMF gösteren diğer iki madde ise SLR yöntemiyle A düzeyinde farklı fonksiyonlaşırken Poly-SIBTEST yöntemiyle B düzeyinde farklı fonksiyonlaşmıştır. Dolayısıyla her iki yöntem de DMF'li maddeleri tespit etmiş ancak 2 madde de DMF düzeylerini farklı ortaya koymuşlardır. Ancak genel anlamda yöntemlerin uyum gösterdikleri görülmektedir.

Her iki yönteme ayrı ayrı bakıldığında SLR yöntemiyle DMF gösteren 8 maddenin 5 tanesinin TBO ve 3 tanesinin de TB olduğu; poly-SIBTEST yöntemiyle DMF gösteren 7 maddenin ise 5 tanesinin Avustralya, 2 tanesinin ise ABD lehine çalıştığı sonucuna varılmıştır.

Avustralya ve ABD örneklemeleri arasında önemli düzeyde DMF gösteren madde sayısının 3'dür ve bu durumun aynı dili konuşan iki ülkenin kültürel yapısındaki farklılıklardan kaynaklanmakta olduğu sonucuna varılmıştır.

Araştırma sorusu 2c'ye ilişkin sonuçlar

Bilimsel sorgulamaya verilen destek ölçme modeline ait maddeler için Avustralya ve Türkiye örneklemeleri arasında yapılan DMF analizleri sonucunda SLR yöntemiyle tüm maddelerin, poly-SIBTEST yöntemi ile 9 maddenin farklı fonksiyonlaştığı ortaya çıkmıştır. Her iki yöntemde de maddelere ait ki-kare farklılık değerlerinin diğer karşılaştırmalara oranla oldukça yüksek olduğu tespit edilmiştir. ST18Q06 kodlu madde dışındaki tüm maddelerde her iki yöntemde de C düzeyinde

DMF olduđu sonucuna varılmıştır. Farklı olarak sadece ST18Q06 kodlu madde SLR analizi ile A yani ihmal edilebilir düzeyde farklı fonksiyonlaşmıştır. Bu durumda yöntemlerin oldukça yüksek uyum gösterdikleri sonucuna ulaşılmıştır.

Her iki yönteme ayrı ayrı bakıldığında SLR yöntemiyle DMF gösteren 10 maddenin 7 tanesinin TBO ve 3 tanesinin de TB olduđu; poly-SIBTEST yöntemiyle DMF gösteren 9 maddenin ise 3 tanesinin referans grup olan Avustralya, 6 tanesinin ise odak grup olan Türkiye lehine çalıştığı sonucuna varılmıştır.

Avustralya ve Türkiye arasında 9 madde önemli düzeyde DMF göstermiştir. Bu durum bu iki ülkenin farklı dil ve kültürel yapıya sahip olmalarından kaynaklanmaktadır. Ancak maddelerin çoğu odak grup olan Türkiye örnekleminin lehine farklı fonksiyonlaşmıştır. Bu durum betimsel istatistik sonuçlarına bakılarak da desteklenebilir. Betimsel istatistik sonuçlarına göre bilimsel sorgulamaya verilen destek ölçme modeli aritmetik ortalamasının en yüksek olduđu ülke de Türkiye'dir.

Genel olarak,

Doğrulamayı faktör analizi sonucunda ölçme modelinin faktör yapısının tüm kültürlerde benzer yapıda olduđu ortaya çıkmış olmasına rağmen bu sonuç bilimsel sorgulamaya verilen destek ölçme modelinin çalışmaya dahil edilen 4 ülke için eşdeğer olduđu sonucunu beraberinde getirmemiştir. Avustralya referans, Yeni Zelanda, ABD ve Türkiye odak grup alınarak yapılan SLR ve poly-SIBTEST analizleriyle bilimsel sorgulamaya verilen destek ölçme modelini oluşturan maddelerin DMF gösterip göstermediği araştırılmıştır. Bu çalışmada ortaya çıkan sonuçlar Tablo 5-1' de özetlenmiştir.

Tablo 5-1. Kültür ve dil farklılıklarına göre SLR ve poly-SIBTEST sonuçları

	SLR			Poly-SIBTESTS	
	Madde sayısı	DMF’li madde sayısı ve düzeyi	%DMF	DMF’li madde sayısı ve düzeyi	%DMF
Aynı dil- Benzer kültür (AUS- Yeni.Z)	10	1 tane B 4 tane A	%10 %40	1 tane B 3 tane A	%10 %30
Aynı dil- Farklı kültür (AUS-ABD)	10	1 tane C 7 tane A	%10 %70	1 tane C 2 tane B	%10 %20
Farklı dil-Farklı kültür (AUS-TUR)	10	1 tane A 9 tane C	%10 %90	4 tane A 9 tane C	%40 %90

Tablo 5-1’ de görüldüğü üzere ülkeler arasındaki dil ve kültür farklılıkları arttıkça DMF gösteren madde sayısı artmış ve DMF’nin düzeyi yükselmiştir. Elde edilen bulgular benzer çalışmalar tarafından da desteklenmektedir. Sonuçlar Asil 2010’in dil ve kültür farklılıkları arttıkça maddelerin farklı fonksiyonlaştıkları sonucu ile tutarlıdır. Yine Atalay 2010 çalışmasında kültüre dayalı yaptığı analizlerde kültürler arası farkın maddelerin farklı fonksiyonlaşmasına neden olduğunu ortaya koymuştur. Sonuç olarak kültür ve dil farklılıkları bilimsel sorgulamaya verilen destek ölçme modeli maddelerinin farklı fonksiyonlaşmasına neden olmuştur dolayısıyla ülkeler arası yapılan karşılaştırmalar sadece Avustralya ve Yeni Zelanda arasında anlamlı olacaktır.

Aynı zamanda bilimsel sorgulamaya verilen destek ölçme modelinin kültürel ve dilsel farklılıklara göre değişen madde fonksiyonu gösterip göstermediği SLR ve poly-SIBTEST teknikleri ile test edilmiş bu teknikler karşılaştırılmış ve tekniklerin benzer sonuçlar verdiği ortaya konmuştur. Ancak Poly-SIBTES’in yüksek düzeyde DMF belirlemede SLR’ ye göre biraz daha hassas olduğu görülmektedir. Genel anlamda bakıldığında yöntemlerin uyumlu çalıştığı ve karşılaştırılabilir olduğu sonucuna varılmıştır.

5.2. Öneriler

5.2.1. PISA uygulayıcılarına yönelik öneriler

1) PISA uygulaması ile elde edilen anket verileri üzerinden öğrenci başarısı ile ilgili sonuçlar çıkarılmakta ve eğitim politikalarında bu sonuçlara göre düzenlemeler yapılmaktadır. Dolayısıyla bu anketleri oluşturan maddelerin kültürel ve dilsel eşdeğerlikleri ortaya konmalıdır. Bunun için uygulayıcılara PISA, PIRLS, TIMSS gibi uluslar arası uygulanan sınavların öğrenci, okul, veli ve yönetici anketleri üzerinde düzenli olarak DMF analizi yapmaları ve DMF'nin olası sebeplerini ortaya koymaya çalışmaları önerilmektedir.

2) PISA sınavının sonuçlarına bakarak eğitim sistemi ile ilgili önemli kararlar alınmadan önce MEB Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü tarafından düzenli olarak DMF analizlerinin yapılması ve bu alanda uzmanların eğitilmesi önerilmektedir.

5.2.2. Araştırmacılara yönelik öneriler

1) Çalışmada dil ve kültürel farklılıklara göre DMF gösteren maddeler ortaya çıkarılmıştır. Farklı dillerde farklı kültürlerle uyarlanmış olan anket veya test maddelerinin aynı biçimde cevaplanmasını beklemek pek de doğru bir yaklaşım olmayabilir. Dolayısıyla DMF gösteren maddeler uzmanlar tarafından incelenerek, DMF'nin olası nedenleri belirlenmeli ve ne derecede DMF'nin yanlılığa sebep olduğu araştırılmalıdır.

2) Hambleton ve Patsula (1999), kültürler arası karşılaştırmalarda, testin farklı dil versiyonlarının eşdeğerliğinden emin olabilmek için bir veya birden fazla yöntemle DMF analizlerinin yapılmasını önermektedirler. Çünkü DMF belirleme yöntemleri örneklem büyüklüğü, madde karakteristikleri ve yetenek dağılımı gibi değişkenlere göre farklı çalışmaktadır. Bu araştırmada DMF belirlemede SLR ve poly-SIBTEST

yöntemleri kullanılmıştır. Araştırmacılar, ÇG-DFA, LDFA, GMH gibi yöntemleri kullanarak birden fazla yöntem karşılaştırması ile DMF analizleri yapabilirler.

3) Avustralya hem örneklem büyüklüğünün fazla olması hem de PISA ölçme araçlarının geliştirilmesinden sorumlu olan konsorsiyumun liderliğini yapmasından dolayı referans grup olarak belirlenmiştir. Ancak referans değişkenin gruplar arasında ölçme değişmezliğini sağladığından emin olunmalıdır. Dolayısıyla bundan sonra yapılacak çalışmalarda referans değişkenin belirlenmesi için literatürde önerilen farklı yöntemler kullanılarak bunların DMF analizleri üzerindeki etkileri belirlenebilir.

4) PISA uygulamasına farklı dillerle katılan tüm ülkeler araştırmaya dahil edilerek DMF analizleri yapılabilir. Bu sayede çeviri maddelerde DMF araştırması yapılabilir.

5) SLR yönteminde kullanılan ölçütlerin uygunluğunu belirlemek için simülasyon çalışmaları yapılabilir. Farklı ölçütlere göre madde seçiminin 1. tip hataya etkisi araştırılabilir.

KAYNAKLAR

- Acar, T. (2008). *Maddenin farklı fonksiyonlaşmasını belirlemede kullanılan genelleştirilmiş aşamalı doğrusal modelleme, lojistik regresyon ve olabilirlik oranı tekniklerinin karşılaştırılması*. Doktora Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, 29(1), 67-91.
- Allouf, A., Hambleton, R.K. ve Sireci, S.G. (1999). Identifying the causes of DIF in translated verbal items. *Journal of Educational Measurement*, 36, 185-198.
- Asil, M. (2010). *Uluslararası öğrenci değerlendirme programı (pisa) 2006 öğrenci anketinin kültürler arası eşdeğerliğinin incelenmesi*. Doktora Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Atalay, K. (2010). *PISA 2006 öğrenci anketinde yer alan tutum maddelerinin değişen madde fonksiyonu açısından incelenmesi*. Yüksek Lisans Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Azen, R., Bronner, S., & Gafni, N. (2002). Examination of Gender Bias in University Admissions. *Applied Measurement in Education*, 15, 75-94.
- Bekçi, B. (2007). *Orta öğretim kurumları öğrenci seçme ve yerleştirme sınavının değişen madde fonksiyonlarının cinsiyete ve okul türüne göre incelenmesi*. Yüksek Lisans Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.

- Berberoğlu, Giray. (1995b). Differential item functioning (dif) analysis of computation, word problem and geometry questions across gender and SES, *Studies in Educational Evaluation*, 21, 439-456.
- Büyüköztürk, Ş., Şekercioğlu, G. ve Çokluk, Ö. (2010). *Sosyal bilimler için çok değişkenli istatistik*. (1. Baskı). Ankara: Pegem Akademi
- Camilli, G. (2006). Test fairness. In R. L. Brennan (ed.), *Educational measurement* (4th ed., 4, 221-256). Westport: American Council on Education & Praeger Publishers.
- Camili, G., ve Shepard, L. A. (1994). *Methods for identifying biased test items*. London: Sage publications.
- Chang, H., Mazzeo, J., ve Roussos, L. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *Journal of Educational Measurement*, 33(3), 333-353.
- Childs, R.A., (1990). Gender Bias and Fairness. Washington, *Clearinghouse on Tests Measurement and Evaluation*, ERIC Identifier: ED328610
- Clauser, B. E., ve Mazor, K. M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice*, 17, 31-44.
- Cole, N.S., and Moss, P.A. (1989). Bias in Test Use. In *Educational Measurement* (Ed.. Robert L. Linn), 201-218. Collier MacMillian: London
- Cromwell, S., (2002) A Primer on Ways To Explore Item Bias. February, *Educational Resources information center (ERIC)*

- Deveci, N. (2008). *Üniversiteler arası kurul yabancı dil sınavının madde yanlılığı açısından incelenmesi*. Doktora Tezi, Ankara: Ankara Üniversitesi Eğitim Bilimleri Enstitüsü.
- Doolittle, A. E. ve Cleary, T. A (1987). Gender-based differential item performance in mathematics achievement items. *Journal of Educational Measurement*, 24, 157-166.
- Dorans, N. J., ve Holland, P. W. (1993). DIF *detection and description*: Mantel-Haenszel and standardization. In P. W. Holland ve H. Wainer (Eds.), *Differential item functioning* (35-66). Hillsdale, NJ: Erlbaum.
- EARGED (2011). *Üçüncü Uluslararası Matematik Ve Fen Bilgisi Çalışması: Ulusal Rapor*. MEB, Ankara
- Ellis, B. B., and Raju, N.S.(2003). Test and İtem Bias:What They Are, What They Aren't, and How To Detect Them. *Educational Resources information center(ERIC)*
- Fang, T. (1999). *Detecting DIF in polytomous item responses*. Unpublished doctor dissertation, University of Ottawa, Canada.
- Feng, Y. (2008). *Difference in gender differential item functioning patterns across item format and subject area on diploma examinations after change in administration procedure*. Unpublished doctor dissertation, University of Alberta, Edmonton, Alberta, Canada.
- Fidalgo, A.M., Ferreres, D., Muniz, J. (2004). Utility of the mantel-haenszel procedure for detecting differential item functioning in small samples. *Educational and Psychological Measurement*, 64 (6), 925-936.

- French, A. W., ve Miller, T. R. (1996). Logistic regression and its use in detecting differential item functioning in polytomous items. *Journal of Educational Measurement*, 33(3), 315-332
- Gelin, M. N., ve Zumbo, B. D. (2003). DIF results may change depending on how an item is scored: An illustration with the Center for Epidemiological Studies Depression (CES-D) scale. *Educational and Psychological Measurement*, 63, 65-74.
- Gierl, M. J. (2000). Construct Equivalence on Translated Achievement Tests. *Canadian Journal of Education*, 25(4), 280-296
- Grisay, A. ve 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. and Swaminathan, H.(1985). *Item response theory: Principles and Applications*. Boston, Kluwer - Nijhoff Publishing.
- Hambleton, K. R., Rogers, J. H. (1989). Detecting potentially biased test items: comparison of IRT area and MH methods. *Applied Measurement in Education*, 2 (4), 313–334.
- Harris, A. M. & Carlton, S. T. (1993). Patterns of Gender Differences on Mathematics Items on the Scholastic Aptitude Test. *Applied Measurement In Education*, 6(2), 137-151.

- Henderson, D. L. (1999). *Investigation of differential item functioning in exit Examinations across item format and subject area*. Unpublished doctor dissertation, University of Alberta, Edmonton, Alberta, Canada.
- Hidalgo, M. D., Lopez-Pina, J. (2004). Differential item functioning detection and effect-size: a comparison between LR and MH procedures. *Educational and Psychological Measurement*, 64 (6), 903–915.
- Hovardaoğlu, S. ve Sezgin N. (1998) *Eğitimde ve Psikolojide Ölçme Standartları*. Türk Psikologlar Derneği ve Öğrenci Seçme ve Yerleştirme Merkezi Yayını.
- Hui, C. H. ve Triandis, H. C. (1985). Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology*, 16(2), 131-152.
- Hulin, C.H., Drasgow, F. ve Parsons, C.K. (1983). *Item response theory: Application to psychological measurement*, Dow Jones-Irwin
- Jensen, A. R. (1980). *Bias in mental testing*. Newyork: Free Press.
- Jodoin M. G. & Gierl, M. J. (2001).Evaluating type 1 error and power using an effect size measure with the logistic regression procedure for dif detection. *Applied Measurement in Education*, 14(4), 329-349.
- Kalaycıoğlu, D. (2008). *Öğrenci seçme sınavının madde yanlılığı açısından incelenmesi*. Doktora Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Kristjansson , E., Aylesworth,R., Mcdowell,I., and Zumbo,D A (1997). Comparison of Four Methods for Detecting Differential Item Functioning in Ordered Response Items, ERIC/AE.

- Kristanjanjonn E., R. Aylesworth, I. McDowell & B.D. Zumbo (2005). a comparison of four methods for detecting differeantial item functioning in ordered response model. *Educational and Psychological Measurement*. 6, 935-953.
- 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). Investigation gender differential item functioning across countries abd test languages for PISA science items. *International Journal of Testing*, 9(2), 122-133.
- Lord, F. M. ve Novick, M.R., Birnbaum, A. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- MEB (2007). *PISA 2006 Uluslararası Öğrenci Değerlendirme Programı Ulusal Ön Rapor*. MEB, Ankara
- Mellenberg, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics*, 7, 105-118.
- Mellor, T. L. (1995). *A comparison of four differantial item functioning methods for polytomously scored items*. Unpublished doctor dissertation. The university of texas Austin.
- Miller, T. R., ve Spray, J. A. (1993). Logistic discriminant function analysis for DIF identification of polytomously scored items. *Journal of Educational Measurement*, 30(2), 107-122.
- Millsap, R. E, ve Everson, H. T. (1993). Methodology review: Statistical approaches for assesing measurement bias. *Applied Psychological Measurement*, 20(3), 257-274.

- Narayanan, P. and Swaminathan, H.(1993). Performance of the Mantel-Haenszel and Simultaneous Item Bias Procedures for Detecting Differential Item Functioning. *Educational Resources Information Center (ERIC). ED 364 557*
- Narayanan, P. & Swaminathan H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement*, 20(3), 257-274.
- OECD (2002). *PISA 2000 Technical Report*. Paris: OECD Publications.
- OECD (2007a). *PISA 2006: Science Competencies for Tomorrow's World: Volume 1 and Volume 2*. Paris: OECD Publications.
- OECD (2009). *PISA 2006 Technical Report*. Paris: OECD Publications.
- Öğretmen, T. (1995). *Differential item functioning analysis of the verbal ability section of the university entrance examination in Turkey*, Orta Doğu Teknik Üniversitesi Eğitim Bilimleri Bölümü Yüksek Lisans Tezi, Ankara.
- Öğretmen, T. (2006). *Uluslararası Okuma Becerilerinde Gelişim Projesi (PIRLS) 2001 Testinin psikometrik Özelliklerinin İncelenmesi: Türkiye- Amerika Birleşik*
- Öğretmen, T., Doğan, N. (2004). OKÖSYS matematik alt testine ait maddelerin yanlılık analizi. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 5,8.
- Öğretmen, T., Doğan, N. (2006). Madde yanlılığını belirleme teknikleri arasında bir karşılaştırma. *Eğitim Araştırmaları Dergisi*, 23, 94-105.
- Öğretmen. T. ve Doğan, N. (2008). Değişen madde fonksiyonunu belirlemede mantel-haenszel, ki-kare ve lojistik regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim*, 148,(33), 100-112.

- Özdemir, D. (2003). Çoktan seçmeli testlerde iki kategorili ve önsel ağırlıklı puanlamanın diferansiyel madde fonksiyonuna etkisi ile ilgili bir araştırma. *Eğitim ve Bilim*. 25; 37–44
- Penfield, D. R. (2001). Assessing differential item functioning among multiple groups: A comparison of three mantel-haanszel procedures. *Applied Measurement in Education*, 14 (3), 253–259.
- Ross, J. S. and Okabe, J. (2006). The Subjective and Objective Interface of Bias Detection on Language Tests. *International Journal of Testing*, 6(3), 229-253
- 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
- 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
- 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, 23-27
- Shimizu, Y. ve Zumbo, B. D. (2005). *A Logistic Regression for Differential Item Functioning Primer*. *Japan Language Testing Association Journal*, 7, 110- 124.
- Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. New Jersey: Lawrence erlbaum associates.
- Smith, F. W. (2009). *Language-related differential item functioning in the WASL mathematics test*. Unpublished doctor dissertation. The university of Washington.

- Stern, W. (1914). *The psychological methods of testing intelligence*. Baltimore: Warwick & York.
- Swaminathan, H. & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361-370.
- Şencan, H.(2005). *Sosyal ve Davranışsal Ölçümlerde Güvenirlik ve Geçerlik*. Özkesen Matbaacılık .Ankara
- Tittle, C.K. (1988). *Test Bias In Educational Teearch, Methodology, and measurement: an International Handbook* (Eds. J. Keeves), 392-398.Pergamon Press: UK.
- Törnkvist, B., and Henriksson, W. (2004). Repeated Test Taking Differences Between Social Groups. 47, 1103-2685
- Yıldırım, H.(2006). *The Differential Item Functioning (Dif) Analysis Of Mathematics Items In The International Assessment Programs*. Doktora Tezi, Ankara: Orta Doğu Teknik Üniversitesi, Eğitim Bilimleri Enstitüsü.
- Yıldırım, H.H. ve Berberoğlu, G. (2006). Judgementel and statistical analyses of the PISA 2003 mathematics literacy items. *International Journal of Testing*, 9(2),108-121
- Yılmaz, V. ve H. E. Çelik (2009), *Lisrel ile yapısal eşitlik modellemesi- 1 temel kavramlar, uygulamalar, programlama*, Ankara: PEGEM Akademi
- Yurdagül, H. *Ölçmelerde hata yanlılık ve diferansiyel madde fonksiyonu*, <http://yunus.hacettepe.edu.tr>, 21. 12. 2011 tarihinde erişildi.

- Yurdugöl, H. ve Aşkar, P. (2004). Ortaöğretim kurumları seçme ve yerleştirme sınavının öğrencilerin yerleşim yerine göre diferansiyel madde fonksiyonu açısından incelenmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi* 27, 268-275
- Zumbo, B. D., & Thomas, D. R. (1996). *A measure of DIF effect size using logistic regression procedures*. Paper presented at the National Board of Medical Examiners, Philadelphia.
- Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.
- Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, Where it is now, and Where It Is Going. *Language Assessment Quarterly*, 4(2), 223–233

EKLER

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

Q18	How much do you agree with the statements below? (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

S18	Aşağıdaki ifadeler ne ölçüde katılıyorsunuz? (Her sırada sadece bir kutuyu işaretleyiniz)
ST18Q01	a) Fen bilimleri ve teknolojiadaki 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 teknolojiadaki 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 teknolojiadaki 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	21

Not. Her bir madde için kayıp veri oranı 0.004 – 0.014 arasında değişmektedir.

Ek 4. Avustralya Örnekleme için Çoklu Bağlantılılık Analizi Sonuçları

Dimension	Condition Index	ST18Q01	ST18Q02	ST18Q03	ST18Q04	ST18Q05	ST18Q06	ST18Q07	ST18Q08	ST18Q09	ST18Q10
1	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	6.58	0.02	0.01	0.01	0.02	0.04	0.00	0.07	0.00	0.01	0.09
3	9.94	0.02	0.00	0.84	0.01	0.00	0.01	0.00	0.00	0.00	0.03
4	10.77	0.00	0.02	0.03	0.00	0.02	0.00	0.00	0.00	0.91	0.00
5	11.87	0.00	0.00	0.06	0.02	0.05	0.01	0.15	0.42	0.00	0.33
6	12.91	0.01	0.00	0.00	0.03	0.17	0.00	0.25	0.34	0.01	0.41
7	14.25	0.02	0.00	0.01	0.08	0.67	0.00	0.34	0.08	0.03	0.13
8	14.85	0.03	0.05	0.01	0.82	0.03	0.04	0.03	0.11	0.02	0.00
9	15.71	0.68	0.00	0.00	0.02	0.02	0.25	0.00	0.03	0.00	0.00
10	17.29	0.01	0.05	0.01	0.00	0.00	0.46	0.15	0.00	0.01	0.00
11	18.87	0.21	0.87	0.03	0.01	0.00	0.21	0.01	0.01	0.02	0.00
Tolarance		0.60	0.54	0.68	0.63	0.37	0.50	0.33	0.48	0.63	0.41
VIF		1.66	1.84	1.45	1.57	2.69	1.98	2.97	2.06	1.57	2.42

Ek 5. Yeni Zelanda Örneklemi için Çoklu Bağlantılılık Analizi Sonuçları

Dimension	Condition Index	ST18Q01	ST18Q02	ST18Q03	ST18Q04	ST18Q05	ST18Q06	ST18Q07	ST18Q08	ST18Q09	ST18Q10
1	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	6.43	0.02	0.01	0.01	0.02	0.05	0.00	0.09	0.00	0.01	0.09
3	9.69	0.01	0.00	0.87	0.01	0.00	0.01	0.00	0.01	0.07	0.02
4	10.28	0.01	0.02	0.03	0.01	0.00	0.01	0.00	0.00	0.87	0.00
5	11.45	0.00	0.00	0.05	0.00	0.00	0.02	0.29	0.13	0.00	0.64
6	12.11	0.02	0.00	0.02	0.04	0.12	0.00	0.13	0.66	0.00	0.09
7	13.45	0.04	0.00	0.00	0.02	0.75	0.01	0.38	0.02	0.01	0.14
8	14.33	0.04	0.03	0.00	0.36	0.05	0.45	0.00	0.13	0.00	0.01
9	15.32	0.31	0.09	0.00	0.49	0.02	0.27	0.00	0.00	0.02	0.00
10	16.38	0.41	0.01	0.01	0.00	0.00	0.21	0.09	0.00	0.02	0.01
11	17.79	0.14	0.82	0.00	0.05	0.00	0.02	0.01	0.03	0.00	0.00
		0.63	0.59	0.74	0.66	0.40	0.55	0.38	0.53	0.69	0.46
Tolarance VIF		1.59	1.70	1.34	1.51	2.52	1.81	2.60	1.90	1.44	2.15

Ek 6. ABD Örnekleme için Çoklu Bağlantılılık Analizi Sonuçları

Dimension	Condition Index	ST18Q01	ST18Q02	ST18Q03	ST18Q04	ST18Q05	ST18Q06	ST18Q07	ST18Q08	ST18Q09	ST18Q10
1	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	7.98	0.03	0.01	0.00	0.03	0.08	0.00	0.11	0.01	0.00	0.08
3	10.99	0.01	0.00	0.76	0.00	0.01	0.01	0.01	0.01	0.00	0.09
4	11.67	0.01	0.02	0.00	0.00	0.03	0.01	0.01	0.00	0.84	0.02
5	12.60	0.00	0.00	0.17	0.00	0.03	0.02	0.31	0.11	0.02	0.46
6	13.33	0.00	0.00	0.00	0.05	0.39	0.00	0.02	0.42	0.02	0.22
7	14.57	0.04	0.01	0.01	0.05	0.33	0.00	0.40	0.40	0.00	0.10
8	15.13	0.01	0.02	0.01	0.85	0.09	0.01	0.06	0.00	0.10	0.03
9	16.73	0.43	0.02	0.00	0.01	0.00	0.02	0.03	0.00	0.00	0.00
10	17.48	0.33	0.04	0.01	0.00	0.02	0.68	0.03	0.04	0.00	0.00
11	19.52	0.12	0.88	0.03	0.00	0.01	0.24	0.01	0.00	0.01	0.00
Tolerance		0.60	0.53	0.65	0.62	0.44	0.48	0.43	0.51	0.62	0.51
VIF		1.66	1.87	1.53	1.61	2.23	2.09	2.34	1.98	1.62	1.98

Ek 7. Türkiye Örnekleme için Çoklu Bağlantılılık Analizi Sonuçları

Dimension	Condition Index	ST18Q01	ST18Q02	ST18Q03	ST18Q04	ST18Q05	ST18Q06	ST18Q07	ST18Q08	ST18Q09	ST18Q10
1	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	7.16	0.01	0.01	0.00	0.02	0.01	0.01	0.10	0.00	0.01	0.30
3	10.47	0.00	0.00	0.01	0.02	0.03	0.00	0.40	0.03	0.01	0.66
4	11.73	0.01	0.00	0.77	0.04	0.02	0.02	0.04	0.03	0.03	0.00
5	12.75	0.01	0.01	0.05	0.75	0.00	0.02	0.03	0.14	0.01	0.00
6	13.02	0.00	0.00	0.02	0.07	0.54	0.00	0.05	0.31	0.05	0.00
7	13.69	0.05	0.04	0.02	0.05	0.32	0.00	0.34	0.37	0.00	0.02
8	15.01	0.04	0.08	0.07	0.05	0.00	0.02	0.00	0.10	0.72	0.00
9	17.29	0.15	0.09	0.01	0.00	0.06	0.64	0.00	0.00	0.14	0.01
10	18.91	0.00	0.29	0.01	0.00	0.01	0.26	0.02	0.01	0.01	0.00
11	19.92	0.72	0.47	0.03	0.00	0.00	0.04	0.00	0.01	0.03	0.00
Tolerance		0.62	0.57	0.66	0.71	0.54	0.61	0.50	0.55	0.62	0.61
VIF		1.60	1.73	1.49	1.39	1.82	1.61	1.97	1.79	1.60	1.62

Ek 8. Doğrulayıcı Faktör Analizi LISREL Komut Dosyası

TÜRKİYE DFA

Observed variables

V1-V10

Covariance Matrix from File TUR.COV.

Sample Size: 4537

Latent Variables: GD KD

Relationships:

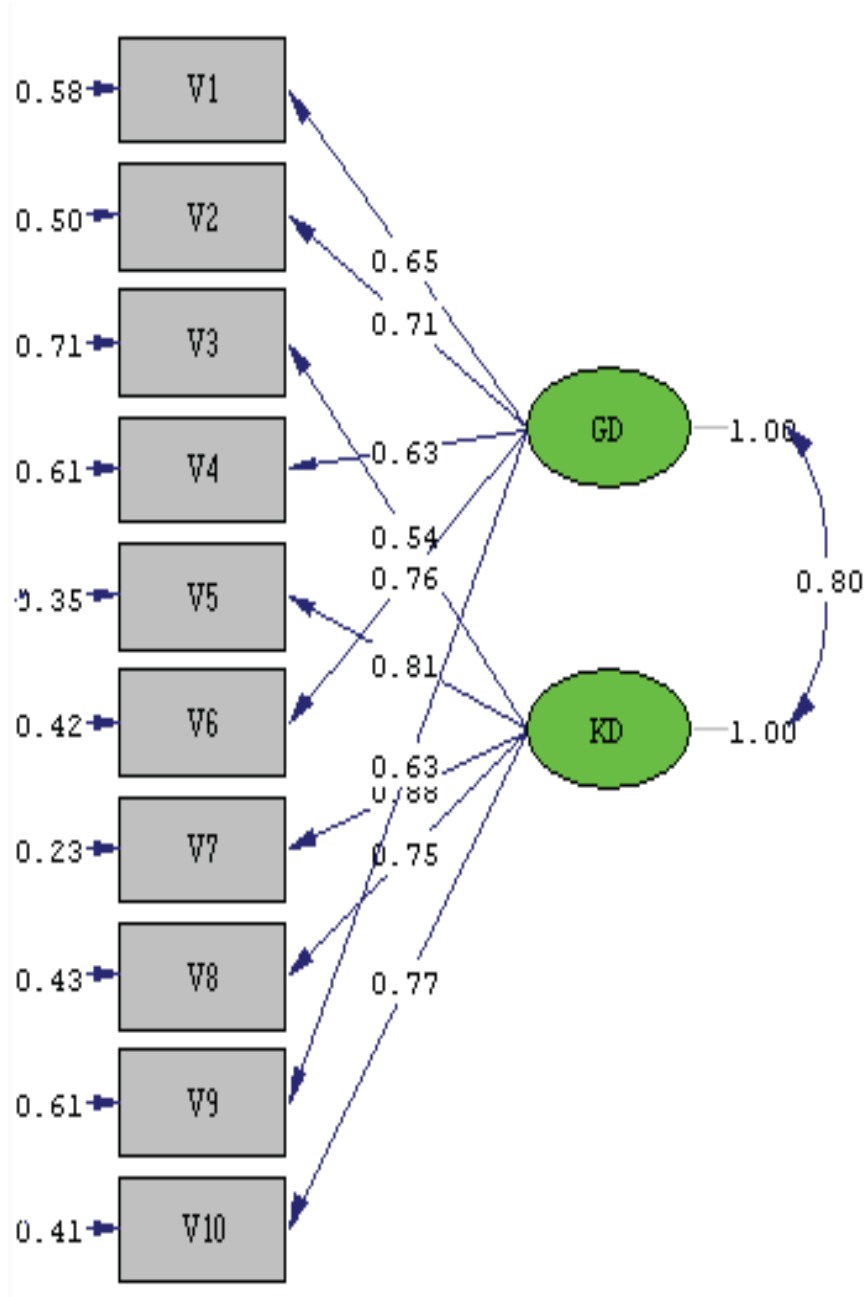
V1 V2 V4 V6 V9 = GD

V3 V5 V7 V8 V10 = KD

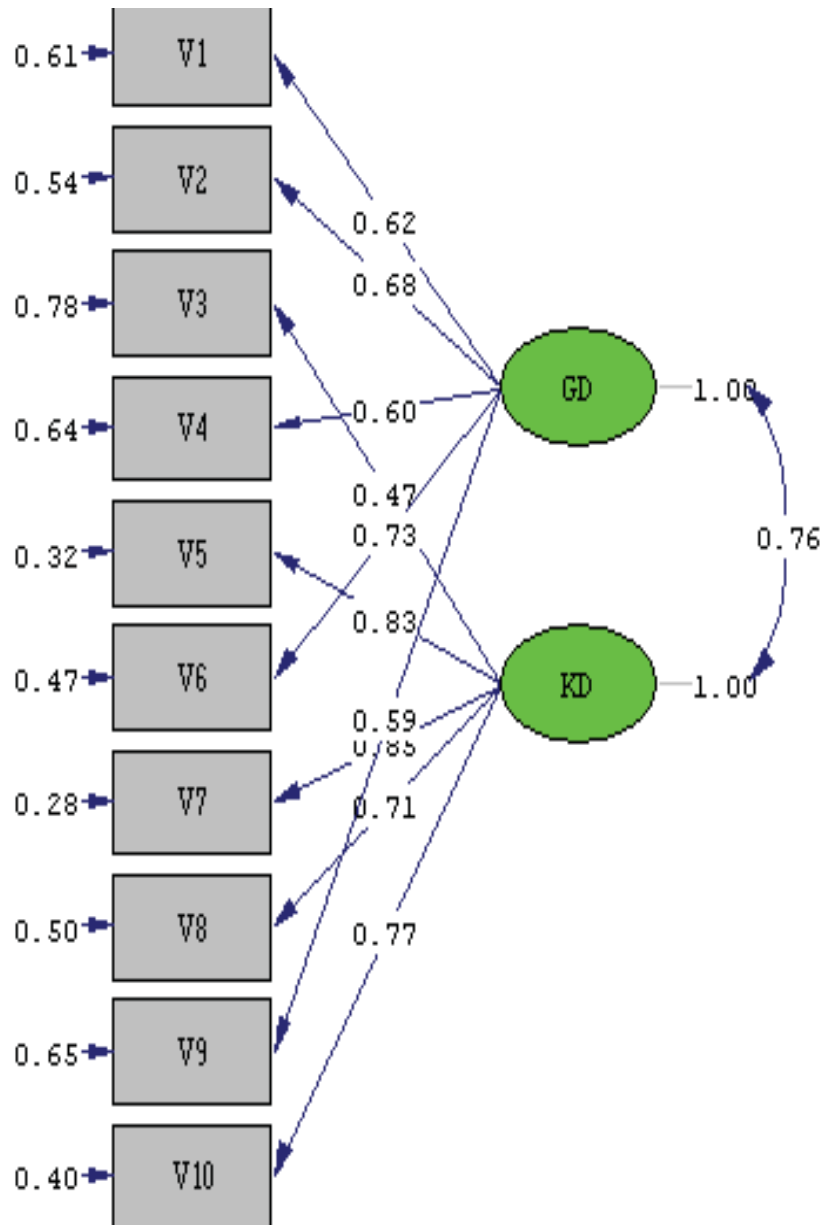
Path Diagram

End of Problem

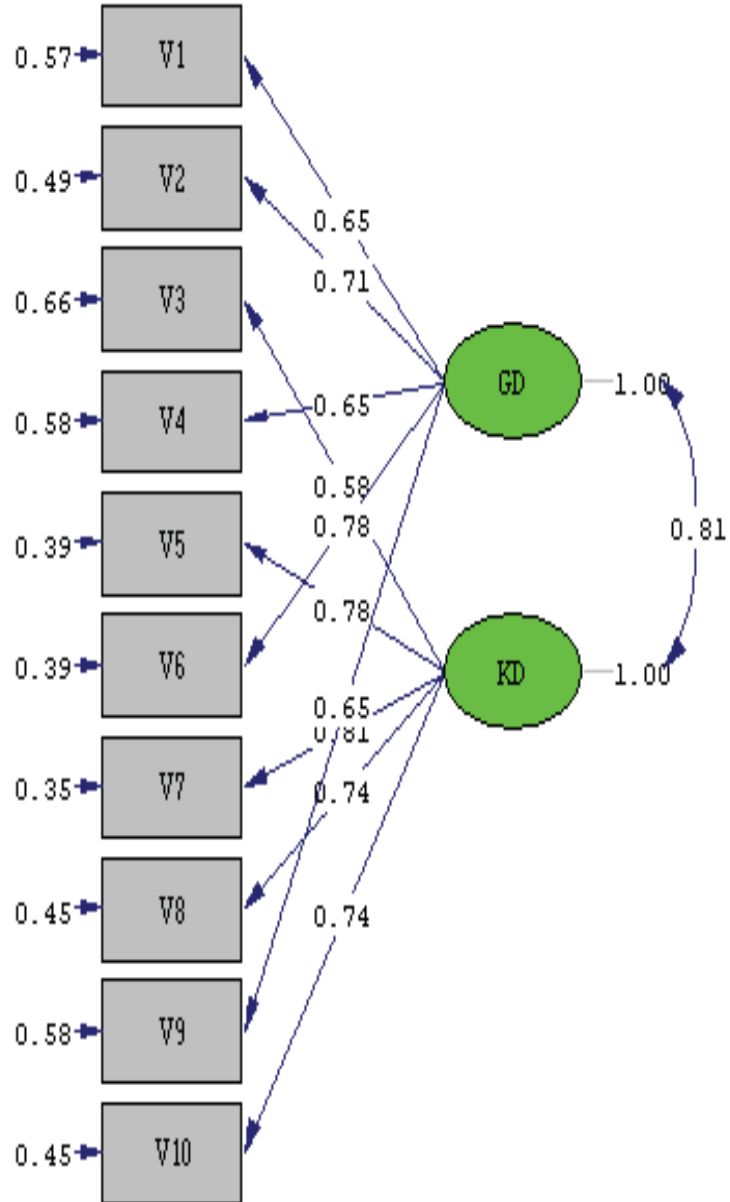
Ek.9 Avustralya Örneklemi için Standartlaştırılmış Faktör Yükleri ve Hata Varyansları



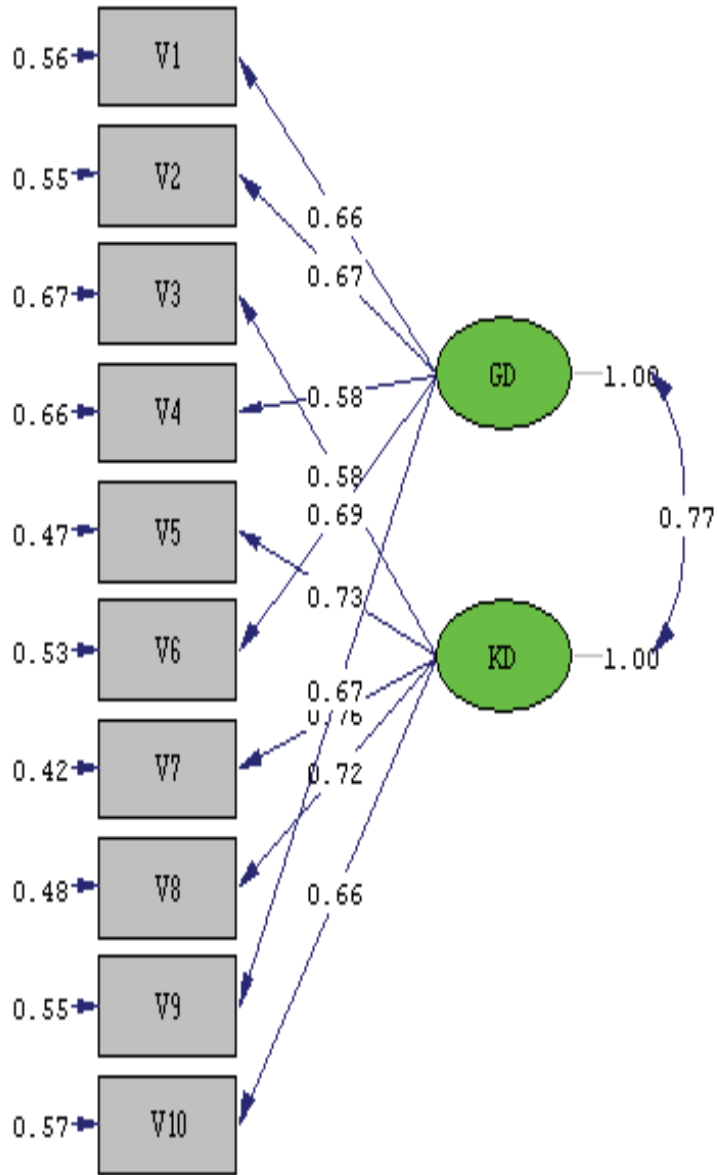
Ek 10. Yeni Zelanda Örneklemi için Standartlaştırılmış Faktör Yükleri ve Hata Varyansları



Ek 11. ABD Örneklemi için Standartlaştırılmış Faktör Yükleri ve Hata Varyansları



Ek 12. Türkiye Örneklemi için Standartlaştırılmış Faktör Yükleri ve Hata Varyansları



Ek 13. Sıralı Lojistik Regresyon Betik

```

* SPSS SYNTAX written by: .
* Bruno D. Zumbo. PhD .
* Professor of Psychology and Mathematics. .
* University of Northern British Columbia .
* e-mail: zumbob@unbc.ca .

* Instructions .
* Copy this file and the file "ologit2.inc". and your SPSS data file
into the same folder .
* Change the filename. currently 'binary.sav' to your file name .
* Change 'item'. 'total'. and 'grp'. to the corresponding variables in
your file.
* Run this entire syntax command file.

include file='ologit2.inc'.
execute.

GET
FILE='gd.sav'.
EXECUTE .

compute item= ST18Q01.
compute total= top.
compute grp= ulke.

* Regression model with the conditioning variable. total score. in
alone.
ologit var = item total
      /output=all.
execute.

* Regression model adding uniform DIF to model.
ologit var = item total grp
      /contrast grp=indicator
      /output=all.
execute.

* Regression model adding non-uniform DIF to the model.
ologit var = item total grp total*grp
      /contrast grp=indicator
      /output=all.
execute.

```

Ek 14. ST18Q01 kodlu maddenin Avustralya- Yeni Zelanda örnekleminde SLR

çıktısı

```
GET
FILE='gd.sav'.
EXECUTE .

compute item= ST18Q01.
compute total= top.
compute grp= ulke.

* Regression model with the conditioning variable, total score, in alone.
ologit var = item total
/output=all.
```

Matrix

Run MATRIX procedure:

LOGISTIC REGRESSION with an ORDINAL DEPENDENT VARIABLE

(by Steffen M. KUEHNEL)

***** Information Section *****

Dependent variable is:

item

Marginal distribution of dependent variable

Value	Frequ.	Percent	%>Value
,00	205,00	1,35	98,65
1,00	1381,00	9,12	89,53
2,00	8985,00	59,32	30,21
3,00	4576,00	30,21	,00

Effective sample size:

15147

Means and standard deviations of independent variables:

	Mean	Std.Dev.
total	10,4064	2,4585

***** Estimation Section *****

Running Iteration No.:
1

Running Iteration No.:
2

Running Iteration No.:
3

Running Iteration No.:
4

Running Iteration No.:
5

➔ Running Iteration No.:
6

..... Optimal solution found.



***** OUTPUT SECTION *****

LR-test that all predictor weights are zero

-2 Log-Likelihood of Model with Constants only:
28718,516

-2 Log-Likelihood of full Model:
16264,013

LR-statistic

Chisqu.	DF	Prob.	%-Reduct
12454,50	1,000	,000	,434

Estimations, standard errors, and effects

	Coeff.=B	Std.Err.	B/Std.E.	Prob.	exp(B)
total	1,104547	,014047	78,633292	,000000	3,017856
Const.1	-3,580141	,133028	-26,912675	,000000	,027872
Const.2	-7,645296	,120987	-63,190976	,000000	,000478
Const.3	-13,110484	,160814	-81,525873	,000000	,000002

Results assuming a latent continuous variable

R-Square (%):
69,15

Standardized regression weights of the latent variable:
total ,8316

----- END MATRIX -----

execute.

* Regression model adding uniform DIF to model.
ologit var = item total grp

Matrix

Run MATRIX procedure:

LOGISTIC REGRESSION with an ORDINAL DEPENDENT VARIABLE

(by Steffen M. KUEHNEL)

***** Information Section *****

Dependent variable is:
item



Marginal distribution of dependent variable

Value	Frequ.	Percent	%>Value
,00	205,00	1,35	98,65
1,00	1381,00	9,12	89,53
2,00	8985,00	59,32	30,21
3,00	4576,00	30,21	,00

Effective sample size:
15147

Means and standard deviations of independent variables:

	Mean	Std.Dev.
total	10,4064	2,4585
grp	1,4625	,8433

***** Estimation Section *****

Running Iteration No.:
1

Running Iteration No.:
2

Running Iteration No.:
3

Running Iteration No.:
4

Running Iteration No.:
5

Running Iteration No.:
6

..... Optimal solution found.

***** OUTPUT SECTION *****

LR-test that all predictor weights are zero

-2 Log-Likelihood of Model with Constants only:
28718,516

-2 Log-Likelihood of full Model:
16263,528



LR-statistic

Chisqu.	DF	Prob.	%-Reduct
12454,99	2,000	,000	,434

Estimations, standard errors, and effects

	Coeff.=B	Std.Err.	B/Std.E.	Prob.	exp(B)
total	1,104360	,014048	78,612667	,000000	3,017294
grp	-,016977	,024367	-,696690	,485997	,983167
Const.1	-3,554229	,138062	-25,743743	,000000	,028603
Const.2	-7,618196	,126986	-59,992264	,000000	,000491
Const.3	-13,083796	,165186	-79,206283	,000000	,000002

Results assuming a latent continuous variable

R-Square (%):
69,15

Standardized regression weights of the latent variable:

total ,8314
grp -,0044

>Error #1. Command name: contrast
>The first word in the line is not recognized as an SPSS command.
>This command not executed.

/output=all.
execute.

* Regression model adding non-uniform DIF to the model.

ologit var = item total grp total*grp
/contrast grp=indicator

Matrix

Run MATRIX procedure:

LOGISTIC REGRESSION with an ORDINAL DEPENDENT VARIABLE

(by Steffen M. KUEHNEL)

Interaction term total*grp
int1.1 total grp

***** Information Section *****

Dependent variable is:
item

Marginal distribution of dependent variable

Value	Frequ.	Percent	%>Value
,00	205,00	1,35	98,65



1,00	1381,00	9,12	89,53
2,00	8985,00	59,32	30,21
3,00	4576,00	30,21	,00

Effective sample size:
15147

Means and standard deviations of independent variables:

	Mean	Std.Dev.
total	10,4064	2,4585
grp	1,4625	,8433
int1.1	15,1471	9,4855

***** Estimation Section *****

Running Iteration No.:
1

Running Iteration No.:
2

Running Iteration No.:
3

Running Iteration No.:
4

Running Iteration No.:
5

Running Iteration No.:
6

..... Optimal solution found.

***** OUTPUT SECTION *****

LR-test that all predictor weights are zero

-2 Log-Likelihood of Model with Constants only:
28718,516

-2 Log-Likelihood of full Model:
16263,100

LR-statistic

Chisqu. DF Prob. %-Reduct



12455,42 3,000 ,000 ,434

Estimations, standard errors, and effects

Estimations, standard errors, and effects

	Coeff.=B	Std.Err.	B/Std.E.	Prob.	exp(B)
total	1,114784	,021275	52,399401	,000000	3,048909
grp	,055318	,113091	,489149	,624736	1,056877
int1.1	-,007160	,010934	-,654877	,512547	,992865
Const.1	-3,659077	,211604	-17,292074	,000000	,025756
Const.2	-7,724299	,206266	-37,448215	,000000	,000442
Const.3	-13,189283	,231197	-57,047868	,000000	,000002

Results assuming a latent continuous variable

R-Square (%):
69,16

Standardized regression weights of the latent variable:

total ,8392
grp ,0143
int1.1 -,0208

----- END MATRIX -----

Ek 15. ST18Q01 kodlu maddenin Avustralya- Yeni Zelanda örnekleminde poly-SIBTEST çıktısı

name of input parameter file = sib.in
 number of items on test = 5
 name of file for Ref. grp. scores = C:\Gözde D\aus abd GD\AUS.dat
 name of file for Focal grp. scores = C:\Gözde D\aus abd GD\NZL.dat
 minimum no. of examinees per matching score cell = 2
 number of runs for this data set = 1
 number of examinees in Reference Group = 2500
 number of examinees in Focal group = 2450

Examinee Test Score Summary Statistics

Reference Group:	Mean = 10.57
	Standard deviation = 2.41
Focal Group:	Mean = 10.23
	Standard deviation = 2.42

Standardized Score Difference = 0.14

Item Statistics

= item number

m = mean score on item

r = point biserial (item score-test score correlation)

#: 1 2 3 4 5

m: 2.183 2.274 2.048 2.138 1.762

r: 0.761 0.756 0.727 0.756 0.720

p-value notation:

R denotes p-value for test of DIF/DBF against Ref. group

F denotes p-value for test of DIF/DBF against Foc. group

E denotes p-value for test of DIF/DBF against either the

Ref. or Foc. Group

SIBTEST-pooled weighting										F
										L
Run	Suspect	Subtest	Beta	standard		p-elim	MS	A		
no.	Item	Numbers	estimate	error	p-value	R	F	SSD	G	
1	1		0.018	0.014	0.212 E	.00	.01	0.13	0	

Ek.16. Özgeçmiş**Kişisel Bilgiler**

Ad Soyad : Gözde Sırgancı

Doğum Yeri ve Yılı: Ordu/ Fatsa-10.05.1986

Eğitim Durumu

Yüksek Lisans : Abant İzzet Baysal Üniversitesi
Eğitimde Ölçme ve Değerlendirme Bölümü (2012)

Lisans : Dokuz Eylül Üniversitesi
Eğitim Fakültesi
Ortaöğretim Matematik Öğretmenliği (2009)

İletişim

E-mail : gozdesirganci@gmail.com

Tel : 05321528040