

**ANKARA ÜNİVERSİTESİ**  
**EĞİTİM BİLİMLERİ ENSTİTÜSÜ**  
**ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI**  
**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI**

**TIMSS 2007 FEN BİLİMLERİ TESTİNDEKİ MADDELERİN DİL VE  
CİNSİYET YANLILIĞI AÇISINDAN İNCELENMESİ**

**YÜKSEK LİSANS TEZİ**

**Hayri Eren Suna**

**Ankara**

**Haziran, 2012**

**ANKARA ÜNİVERSİTESİ**  
**EĞİTİM BİLİMLERİ ENSTİTÜSÜ**  
**ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI**  
**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI**

**TIMSS 2007 FEN BİLİMLERİ TESTİNDEKİ MADDELERİN DİL VE**  
**CİNSİYET YANLILIĞI AÇISINDAN İNCELENMESİ**

**YÜKSEK LİSANS TEZİ**

**Danışman: Prof. Dr. Nükhet Demirtaşlı**

**Hayri Eren Suna**

**Ankara**

**Haziran, 2012**

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼'ne

Bu alıřma j¼rimiz tarafından Eđitim Bilimleri Olme ve Deęerlendirme Anabilim Dalında Y¼KSEK LİSANS TEZİ olarak kabul edilmiřtir.

Başkan Prof. Dr. Ezel Tavřancıl .....

¼ye Prof. Dr. N¼khet Demirtařlı (Danıřman) .....

¼ye Do. Dr. Duygu Anıl .....

Onay

Yukarıdaki imzaların, adı geen oęretim ¼yelerine ait olduęunu onaylıyorum.

.../.../2012

Prof. Dr. Necla KURUL

Enstit¼ M¼d¼r¼

## ÖNSÖZ

Araştırma konumun belirlenmesinde, planlanmasında ve düşünceden bir ürün olarak ortaya çıkmasına kadar geçen süre boyunca yanımda bulunan saygıdeğer hocalarım ve arkadaşlarım bana yön verdiler ve hep yanımda oldular; ailem ise süreç boyunca bana hep destek oldu ve yardımları ile her zaman yanımda olduklarını hissettirdiler.

Öncelikle, süreç henüz başlamadan desteğini aldığım, yardımları ve yönlendirmeleri ile bulunduğum yolda bana daima rehber olan sayın tez danışmanım ve değerli hocam Prof. Dr. Nükhet Demirtaşlı'ya; akademik olarak programa kayıt olduğum günden bu yana desteğini esirgemeyen saygıdeğer ve sevgili hocalarım Prof. Dr. Nizmettin Koç, Prof. Dr. Ezel Tavşancıl ve Yrd. Doç. Dr. Ömer Kutlu'ya; araştırmamla ilgili görüş, dönüt ve düzeltmeleri için sayın Doç. Dr. Ömay Çokluk ve Yrd. Doç. Dr. Deniz Gülleroğlu'na; araştırmam süresince beni yürekte destekleyen ve önerilerini paylaştıkları sayın Prof. Dr. Figen Çok, Yrd. Doç. Dr. Neslihan Güney Karaman, Yrd. Doç. Dr. Türkan Doğan, Yrd. Doç. Dr. Deniz Örucü ve görev yaptığım Başkent Üniversitesi Eğitim Bilimleri Bölümünde bulunan tüm hocalarıma; destekleri, paylaşımları ve yanımda oldukları için Arş. Gör. Tuğçe Özmen ve Arş. Gör. Cansu Ayan'a; süreç boyunca sıklıkla fikir paylaşımında bulunduğumuz Ersoy Karabay'a; tüm yüksek lisans arkadaşlarıma ve farklı şehirlerde olsak da, her koşulda yanımda hissettiğim aileme çok teşekkür ederim.

Bu araştırmayı sizler ve destekleriniz olmadan tamamlayamazdım; nice yıllar boyunca beraber olabilmek dileğiyle.

H. Eren Suna

## ÖZET

### TIMSS 2007 FEN BİLİMLERİ TESTİNDEKİ MADDELERİN DİL VE CİNSİYET YANLILIĞI AÇISINDAN İNCELENMESİ

Suna, Hayri Eren

Yüksek Lisans, Ölçme ve Değerlendirme Ana Bilim Dalı

Tez Danışmanı: Prof. Dr. Nükhet Demirtaşlı

Haziran, 2012, 98 Sayfa

Bu araştırmada, TIMSS 2007 fen bilimleri testinde yer alan maddelerin dil ve cinsiyet yönünden yanlı ölçme yapıp yapmadıklarının belirlenmesi amaçlanmıştır.

Madde yanlılığının olası bir göstergesi olan Madde İşlev Farklılaşması'nın (MİF) belirlenmesi için Mantel-Haenszel (MH), Simultaneous Item Bias Test (SIBTEST) ve lojistik regresyon (LR) yöntemleri kullanılmıştır. Yöntemler kullanılarak yapılan analizler sonucunda MİF gösterdiği belirlenen ve araştırmacıların erişime açılmış olan maddeler için, olası yanlılık kaynaklarının belirlenmesi amacıyla uzman görüşü alınmıştır. Araştırmada dile yönelik işlev farklılığı incelenirken Türkiye-İngiltere örneklemi karşılaştırılmış, cinsiyete yönelik MİF'in belirlenmesi sürecinde Türkiye örnekleme analiz edilmiştir.

Araştırma sürecinde TIMSS 2007 uygulamasında kullanılan 14 kitapçık incelenmiş, fen bilimleri başarı puanına göre, tüm kitapçıklar için ortak olan dokuz yetenek grubu oluşturulmuştur. Analizlerde, üç yöntemin tutarlı olarak en az orta düzey (B düzeyi) MİF gösteren maddelerin tek biçimli MİF, yalnızca SIBTEST ve lojistik regresyon yöntemlerinin en az orta düzeyde MİF gösterdiği maddelerin ise tek biçimli olmayan MİF gösterdiği kabul edilmiştir.

Araştırma bulgularına göre, 14 kitapçıkta dile yönelik MİF gösteren madde sayısı toplam 19'dur. Farklı kitapçıklarda ortak maddeler bulunmaktadır ve 8 madde birden çok kitapçıkta MİF göstermiştir. Sonuç olarak 11 *farklı* maddenin dile göre işlev farklılığı gösterdiği tespit edilmiştir. Bu maddelerin yedisi Türk öğrenciler lehine, dördü İngiliz öğrenciler lehine

MİF göstermiştir. Uzman görüşleri sonucunda, erişime açılmış MİF gösteren maddelerin tercüme hataları içerdiğine fakat yanlılık göstermediğine kanaat getirilmiştir. Türkiye örnekleminde gerçekleştirilen analizler, cinsiyete yönelik MİF gösteren maddenin bulunmadığını ortaya koymuştur.

Anahtar Kelimeler: Madde Yanlılığı, Madde İşlev Farklılaşması (MİF), TIMSS 2007 Fen Bilimleri Testi, Mantel-Haenszel, SIBTEST, Lojistik Regresyon.

## ABSTRACT

### DIFFERENTIAL ITEM FUNCTIONING (DIF) INVESTIGATION OF TIMSS 2007 SCIENCE ITEMS IN TERMS OF LANGUAGE AND GENDER

Suna, Hayri Eren

Master Thesis, Department of Educational Measurement and Evaluation

Thesis Adviser: Professor Dr. N khet Demirtaşı

June, 2012, 98 Pages

The main purpose of the present study was to determine TIMSS 2007 Science Items for potentially biased items in terms of language and gender.

To identify Differential Item Functioning (DIF), which is a potential indicator of item bias, Mantel-Haenszel (MH), Simultaneous Item Bias Test (SIBTEST) and logistic regression (LR) methods were utilized. For the items which include DIF and that were released by IEA, expert opinions were taken to identify possible sources of bias. In the present study, Turkish and English samples were compared to clarify language based DIF and Turkish sample was analyzed to identify gender based DIF.

During analysis process, each item which was presented in 14 different booklets, was investigated. According to science achievement scores of responder, a nine-level ability groups, which are common for all booklets, were formed. In analyses, it was assumed that, items which were signed as moderate DIF (Level B) or large DIF (Level C) by three methods consistently, included uniform DIF and items which were signed as moderate DIF (Level B) or large DIF (Level C) by only SIBTEST and logistic regression consistently, included non-uniform DIF.

Results showed that, total number of the items including DIF in 14 booklets was 19. However, it is known that, different booklets may have common items and 8 of 19 items included DIF in two different booklets. Consequently, it is found that, 11 *different* items included DIF. Seven of these items were signed as including DIF in favor of Turkish students and four of them were showed DIF in favor of English Students. Expert opinions resulted

in that two of the items which showed DIF had translation mistakes and uncertainty, but these were not certain proofs for bias. According to analyses in Turkish sample, there was no item that included DIF in terms of gender.

Keywords: Item Bias, Differential Item Functioning, TIMSS 2007 Science Test, Mantel-Haenszel, SIBTEST, Logistic Regression.



## İÇİNDEKİLER

JÜRİ ÜYELERİNİN İMZA SAYFASI .....	i
ÖNSÖZ.....	ii
ÖZET .....	iii
ABSTRACT .....	v
İÇİNDEKİLER.....	vii
ÇİZELGELER VE ŞEKİLLER LİSTESİ .....	ix
EKLER LİSTESİ .....	xi
BÖLÜM	
1. GİRİŞ	
Problem.....	1
Yanlılık.....	3
Test ve Madde Yanlılığı .....	3
Madde İşlev Farklılaşması (MİF) .....	5
Madde İşlevinin Tek Biçimli Farklılaşması .....	7
Madde İşlevinin Tek Biçimli Olmayan Farklılaşması .....	8
MİF Belirleme Yöntemleri .....	8
Dönüştürülmüş Madde Güçlük İndeksi .....	9
Varyans Analizi (ANOVA) .....	11
Korelasyona Dayalı Yöntemler .....	11
Mantel-Haenszel (MH) Yöntemi.....	12
Simultaneous Item Bias Test (SIBTEST) Yöntemi.....	14
Lojistik Regresyon (LR) Yöntemi .....	16
Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS) .....	17
Amaç .....	19
Önem.....	20
Sayılıtlar.....	21

Sınırlılıklar.....	22
Tanımlar .....	22
Kısaltmalar.....	22
2. İLGİLİ ARAŞTIRMALAR	
Yurtdışında Gerçekleştirilen Araştırmalar.....	23
Yurtiçinde Gerçekleştirilen Araştırmalar.....	30
3. YÖNTEM	
Araştırmanın Modeli .....	33
Evren ve Örneklem.....	33
TIMMS Uygulamasında Kullanılan Ölçme Araçları .....	35
TIMMS Uygulamasında Ölçülen Bilişsel Beceriler.....	37
Verilerin Elde Edilmesi.....	40
Verilerin Analizi.....	40
4. BULGULAR VE YORUMLAR	
Dile Göre MİF Gösteren Maddelerin Bulunması .....	44
Cinsiyete Göre MİF Gösteren Maddelerin Bulunması .....	47
Dile Göre MİF Gösteren Maddelere Yönelik Uzman Görüşleri .....	47
Cinsiyete Göre MİF Gösteren Maddelere Yönelik Uzman Görüşleri.....	53
5. SONUÇ VE ÖNERİLER.....	54
Sonuç .....	54
Öneriler.....	57
KAYNAKÇA.....	59
6. EKLER .....	64
Ek-1 .....	64
Ek-2 .....	73
Ek-3.....	82
Ek-4.....	84

## ÇİZELGE VE ŞEKİLLER LİSTESİ

### Çizelgeler

1. Mantel-Haenszel Analizi İle Elde Edilen $\alpha_i$ Değerleri İçin Kabul Edilen Ölçütler .....	14
2. SIBTEST Analizi İle Elde edilen $\beta_{UNI}$ Değerleri İçin Ölçütler .....	15
3. Lojistik Regresyon Analizi İle Elde Edilen $\Delta R^2$ Değerleri İçin Kabul Edilen Ölçütler .....	16
4. TIMSS 2007 Sekizinci Sınıf Uygulaması Türkiye ve İngiltere Evren ve Örneklemelerindeki Okul, Öğrenci Sayıları ve Cevaplayıcıların Ortalama Yaşları .....	33
5. TIMSS 2007 Türkiye Uygulaması İçin Örnekleme Değerlerinin Bölgelere Göre Dağılımı .....	34
6. TIMSS 2007 Fen Bilimleri Sekizinci Sınıf Testinde Yer Alan Maddelerin Konu Alanları .....	36
7. TIMSS 2007 Sekizinci Sınıf Uygulaması Erişime Açılan Çoktan Seçmeli Fen Bilimleri Sorularının Fen Bilimleri Disiplinlerine Göre Dağılımı .....	40
8. Fen Bilimleri Başarı Puanlarına Göre Oluşturulmuş Eşit Aralıklı Yetenek Düzeyleri .....	42
9. TIMSS 2007 Uygulamasında Kullanılan Kitapçıkları Cevaplayan Öğrencilerin Dağılımı ve Her Bir Kitapçıkta Yer Alan Maddelerin Dile Göre MİF Gösterme Durumu .....	45

### Şekiller

1. Madde İşlevinin Tek Biçimli Farklılaşması .....	7
2. Madde İşlevinin Tek Biçimli Olmayan Farklılaşması .....	8
3. Dönüştürülmüş Madde Güçlüğü İndeksi Yönteminde Kullanılan Çizelge ..	10
4. i Maddesi İçin j Yetenek Düzeyinde Mantel-Haenszel Matrisi.....	13
5. Türk ve İngiliz Öğrencilerin Uluslararası Yeterlik Düzeylerine Ulaşma Durumu .....	39

6. S032115 Maddesinin İngilizce ve Türkçe Kitapçıklardaki Şekli .....	48
7. S032257 Maddesinin İngilizce ve Türkçe Kitapçıklarındaki Şekli .....	49
8. S032385 Maddesinin İngilizce ve Türkçe Kitapçıklarındaki Şekli .....	50
9. S042028 Maddesinin İngilizce ve Türkçe Kitapçıklarındaki Şekli .....	51
10. S042071 Maddesinin İngilizce ve Türkçe Kitapçıklarındaki Şekli .....	52

## EKLER LİSTESİ

EK 1. Birinci Amaç Doğrultusunda Kitapçıklarda Yer Alan Maddelerin Sırası ile MH, SIBTEST, Lojistik Regresyon Çıktılarına Örnekler.....	64
EK 2. İkinci Amaç Doğrultusunda Kitapçıklarda Yer Alan Maddelerin Sırası ile MH, SIBTEST, Lojistik Regresyon Çıktılarına Örnekler.....	73
EK 3. Fen Bilimleri Alan Uzmanlarına Gönderilen Form Örneği .....	82
EK 3. Yabancı Dil Alan Uzmanlarına Gönderilen Form Örneği .....	84

## BÖLÜM I

### GİRİŞ

Bu bölümde araştırmanın problemine, araştırma sorularına, araştırmanın önemine, sayıltılarına ve sınırlılıklarına yer verilmiştir.

#### Problem

Eğitim, “bireyin davranışında, kendi yaşantısı yoluyla istendik davranış değişiklikleri meydana getirme süreci” olarak tanımlanır (Ertürk, 1971). Eğitim sürecinde temel amaç öğrencilere, süreç başlamadan tanımlanmış belli davranışların kazandırılmasıdır. Söz konusu davranışların ne düzeyde kazandırılabilirdiği ve eğitim sürecinin etkinliğinin belirlenmesinde testler kullanılır.

Test sonuçları günümüzde rutin olarak seçme, yerleştirme, terfi işlemleri ve lisans verme süreçlerine temel oluşturmaktadır. Söz konusu sonuçlara göre alınan kararlar kişisel, sosyal ve politik öneme sahiptir. Dolayısı ile bu kararları almak amacıyla kullanılan testlerin geçerli yorumlara izin vermesi kaçınılmaz olmalıdır (Clouser ve Mazor, 1998). Bununla beraber test puanlarının, testin ölçmeyi amaçladığı yapı dışındaki farklı kaynaklardan etkilenmesi kaçınılmazdır. Test sonuçları, test ile ölçülmek istenen yapı dışındaki faktörlerden tamamen yalıtılabilsen, tamamen geçerli ve güvenilir olarak kabul edilebilirdi ancak test sonuçlarını sözü edilen bozucu kaynaklardan tamamen arındırmak mümkün olmadığından, testi cevaplayan çeşitli gruplara haksız bir şekilde avantaj veya dezavantaj sağlanmadığından emin olmak büyük önem taşımaktadır (Crocker ve Algina, 1986). Adil bir test, amacına uygun olarak testi cevaplayan tüm gruplar veya tüm bireylere, bilgi ve becerilerini gösterecek şekilde eşit fırsat sağlar (Roever, 2005). Ayrıca, kullanılan testlerin adil olması hem test geliştiricileri hem de test cevaplayıcıları için ahlaki bir zorunluluktur (Holland ve Wainer, 1993).

Bu doğrultuda çeşitli amaçlar için kullanılacak testlerde, yansızlık maddeler düzeyinde incelenmekle beraber literatür, bir testten elde edilen sonuçların adil olarak kabul edilebilmesi için farklı boyutlarda şu şartların

aranması gerektiğini göstermektedir. (Kunnan, 2000; Shohamy, 2000, akt. Roever, 2005):

- Yapı: İncelenen özelliğin yeterince ölçülmesi gerekmektedir.
- İçerik: Ölçme yapılan alan yeterince temsil edilmelidir.
- Standartlaştırma: Testin tüm cevaplayıcıları aynı şartlara sahip olmalı, aynı materyalleri kullanmalı ve tüm cevaplayıcılar için aynı puanlama şekli kullanılmalıdır.
- *Yansız Maddeler*: Testte yer alan maddeler, testi cevaplayan gruplardan hiçbirine avantaj ya da dezavantaj sağlamamalıdır.
- Test Sonuçları ve Sonuçların Kullanımı: Testten elde edilen sonuçlar, testin amacına uygun kararlar almak için kullanılmalıdır.

Amerika Eğitim Araştırmaları Derneği'nin (AERA) 1999 yılında yayınladığı Eğitim ve Psikolojik Testler İçin Standartlar'da bir test veya testte yer alan maddeden elde edilen sonuçların yorumlanması sürecinde adilliğin önemi vurgulanmış, farklı gruplarda yer alan bireylerin testin tümünden veya testte yer alan maddelerden elde ettikleri puanların farklı anlamlar taşıdığına (karşılaştırılabilir olmadığına) test veya maddeler için adilliğin söz konusu olmadığını belirterek, günümüzde yanlılık durumunun test geliştirme sürecinde, geçerliği tehdit eden en önemli unsurlardan biri olduğu ifade edilmiştir (Osterlind ve Everson, 2009).

Bir testin sonuçlarına dayanarak verilen kararların adil olması, o testin içerdiği maddelerin yansız olmasını gerektirir ve bu durum, daha önce değinildiği gibi, eğitim ve psikolojide kullanılan tüm testlerin sağlaması gereken bir koşuldur. Günümüzde öğrenci başarısını uluslararası düzeyde belirlemeyi ve karşılaştırmayı amaçlayan çalışmalar artmaktadır. Söz konusu çalışmalar neticesinde farklı ülkelerde yer alan öğrencilerden elde edilen puanların karşılaştırılabilirliği, kullanılan testlerin psikometrik özelliklerine yönelik kanıt toplama sürecinde incelenmesi gereken noktalardan birisidir (Wu ve Ercikan, 2006). Bu durum göz önünde bulundurularak “yanlılık” konusunun daha detaylı incelenmesinde yarar görülmektedir.

## **Yanlılık**

Yanlılık kavramının çeşitli alanlarda çalışan birçok araştırmacı tarafından farklı tanımları yapılmıştır. Sözlük tanımı olarak yanlılık “gerçeğin ifadesinden veya göstergelerinden ayrılma” şeklinde ifade edilir. İstatistiksel olarak yanlılık “gerçek değerden bir yöne veya diğerine doğru sapma” olarak tanımlanabilir. Eğitim ve psikolojideki ölçmeler kapsamında ise yanlılık Shephard (1981) tarafından “Bir gruba avantaj veya dezavantaj sağlanmasına yol açan sistematik hata” olarak tanımlanmıştır. Yanlılık sistematik hata oluşturur çünkü test sonuçlarında belirli bir grubun üyelerine yönelik bozucu bir etki yaratır. Farklı bir ifade ile yanlılık, ölçülmek istenen değişken değerinin, sistematik olarak olduğundan fazla veya olduğundan az ölçülmesidir (Reynolds, Livingston ve Willson, 2006).

Diğer taraftan sıklıkla yanlılık ve adillik kavramları beraber kullanılmaktadır ve bu durum söz konusu iki kavramın karışmasına neden olmaktadır. Test yanlılığı ve adillik birbiri ile ilişkili olmakla beraber farklı kavramlardır. Brown, Reynolds ve Whitaker (1999)'in belirttiği gibi, adillik insanların farklı bakış açılarına sahip olabildiği ahlaki, psikolojik ve hukuksal bir durumdur. Diğer taraftan yanlılık bir testin istatistiksel bir özelliğini ifade eder. Yani, yanlılık gerçek bir test verisinden elde edilen, deneysel bir süreci ifade ederken adillik, farklı tartışma ve görüşlerle desteklenen bir ilkedir.

## **Test ve Madde Yanlılığı**

Test yanlılığı, bir testin sonuçlarına dayanılarak alınan bir karar, tüm gruplar için adil değilse ya da bir gruba, diğerine göre eşit olmayan bir etki yapıyorsa ortaya çıkar (Osterlind ve Everson, 2009). Diğer bir tanım ile test yanlılığı, bir testin, bir alt gruba ait bireylerin özelliklerini ölçme sürecindeki sistematik hatadır (Camilli ve Shephard, 1994).

Test yanlılığı sıklıkla ırksal, etnik ve cinsiyet farklılıklarının ortaya konulduğu çalışmalarda bir problem olarak görülür. Bununla beraber test yanlılığı farklı çalışma alanlarında, sosyoekonomik düzey, yaş, din, şehirde ya da kırsal bölgede yaşama gibi kimi grup farklılıklarının incelendiği çalışmalarda da önem teşkil etmektedir (Camilli ve Shephard, 1994).



Hiçbir test ölçmeyi amaçladığı özelliği mükemmel olarak ölçemez ve bu özelliğin gösterildiği alanı tümüyle temsil edemez, dolayısı ile test sonuçlarını ölçme hatalarından tamamı ile yalıtım mümkün değildir. Bununla beraber söz konusu ölçme hatası farklı alt gruplarda bulunan tüm bireyleri eşit şekilde etkiliyorsa, bu durumda testin yanlı olduğu savunulamaz.

Birçok araştırmacı (Ironson, 1982; Linn ve Drasgow, 1987; Linn, Levine ve Hastings, 1981; Shepard, Camilli ve Averill, 1981) madde yanlılığını benzer şekilde tanımlamışlardır; eğer bir maddeyi, aynı yeteneğe sahip fakat farklı gruplarda bulunan bireylerin doğru cevaplama olasılıkları eşit değilse, o madde yanlıdır. Diğer bir tanımla madde yanlılığı, testte yer alan bir maddenin testi cevaplayanların oluşturduğu bir gruba istenmeyen bir şekilde avantaj veya dezavantaj sağlamasıdır (Osterlind ve Everson, 2009).

Madde yanlılığının incelenmesinde temel varsayım, bir testte ölçülmek istenen özellik hakkında aynı bilgiye sahip olan bireyler, yaşları, kültürleri, etnik kökenleri veya ırksal özelliklerinden bağımsız olarak, testte yer alan her bir maddeyi cevaplama aynı performansı göstermelidirler (Subkoviak, Mack ve Ironson, 1984, akt. Perrone, 2006). Bu varsayım doğrultusunda, madde ile ölçülen özelliğe sahip olan bir cevaplayıcı, herhangi bir demografik grubun (cinsiyet, yaş, etnik köken vb.) üyesi olması nedeni ile o maddeyi doğru cevaplayamıyor ise o madde yanlılık göstermektedir.

Camilli ve Shepard (1994)'a göre bazı maddeler test geliştiriciler tarafından, bir demografik grupta bulunan cevaplayıcıların, söz konusu maddeyi doğru yanıtlayacak özelliğe sahip olsa dahi yanıtlayamayacağı şekilde sorulmaktadır. Testin ölçmek istediği yapı ile ilgili olmayan çeşitli bozucu kaynaklara sahip olan böyle maddeler, söz konusu grup için geçerli sonuçlar sağlamazlar. Bu durumda sözü geçen grupta bulunan cevaplayıcılar, madde ile ölçülmek istenen özelliğe sahip olmalarına rağmen maddelerin içerdiği ilgisiz bir özellik nedeniyle gerçek performanslarını gösterememişlerdir; bu durumda maddeler bu gruba, diğer gruplara göre dezavantaj sağlamıştır.

Madde yanlılığının incelenmesi sürecinde, madde işlevinin farklılaşması ve madde etkisi sıklıkla birbirinin yerine kullanılan ve karıştırılan

iki kavramdır. Madde etkisi, farklı gruplarda bulunan bireylerin bir maddeyi doğru yanıtlama olasılıklarının, söz konusu madde ile ölçülmek istenen özellik boyutunda farklılaştığında oluşur (Zumbo, 1999). Madde etkisinde gruplar arasındaki fark, grupların ölçülen özelliğe yönelik yetenek düzeylerinin farklılaşmasından kaynaklanır. Oysa madde işlevinin farklılaşması sürecinde, gruplarda yer alan bireylerin puan ortalamaları hesaplanmadan önce grupların yetenek düzeyleri eşitlenir, dolayısı ile grupların puan ortalamaları arasındaki fark, grupların yetenek düzeylerindeki farklılaşmadan değil, maddenin işlevindeki farklılaşmadan kaynaklanır. Dolayısıyla, farklı grupların testten ve maddeden aldıkları puanlar ortalaması arasındaki fark, tek başına yanlılığın bir kanıtı olarak görülmemelidir (Camilli ve Shepard, 1994).

### **Madde İşlevinin Farklılaşması (MİF)**

Madde işlevinin farklılaşması basitçe, bir demografik grupta yer alan bireylerin testteki bir maddeyi, aynı bilgi düzeyine sahip olan ikinci bir demografik grupta yer alan bireylere oranla daha sıklıkla doğru cevapladığı durumları açıklamayı sağlayan istatistiksel bir kavramdır (Zumbo, 2007). Madde işlevinin farklılaşması, bir evrende bulunan çeşitli demografik grupların performanslarındaki sistematik farklılaşmayı madde ve test düzeyinde belirleyen istatistiksel teknikleri ifade eder (Osterlind ve Everson, 2009). Diğer bir anlatım ile MİF, bir maddenin iki farklı grupta psikometrik olarak nasıl farklılaştığını ortaya koyar. Buna göre, bir madde kendisi ile ölçülmesi amaçlanan özellik boyutunda eşleştirilmiş gruplarda yer alan bireylerin performanslarında farklılaşmaya neden oluyorsa, o maddenin işlevinin farklılaştığından söz edilir (Holland ve Wainer, 1993).

Lord'a (1980) göre tüm test maddeleri, testi alan tüm gruplarda aynı yanıt işlevine sahip olsaydı; aynı yetenek veya beceriye sahip olan tüm bireyler, hangi grubun üyesi olduklarından bağımsız olarak, maddelere doğru cevap vermek için aynı şansa sahip olurlardı. Böyle maddelerden oluşan test ise kesinlikle yansız ve adil olurdu. Diğer taraftan, bir maddenin cevap işlevi bir gruptan diğerine değişiyorsa, söz konusu maddenin tüm cevaplayıcı gruplarına doğru yanıt vermeleri için eşit şans tanıdığını iddia etmek mümkün değildir (Holland ve Wainer, 1993).

MİF analizlerinin temel amacı işlev farklılığı gösteren maddeleri belirlemek olmakla beraber, analiz sonuçlarının yapı geçerliğine yönelik kanıt sunduğu da günümüzde tartışılmaktadır (Walker 2011; Walker ve Beretvas, 2001). Sıklıkla işlev farklılığı gösteren maddeler, maddenin cevaplanması için gerekli kazanımla ilgili olmayan özellikler barındırdığından; ölçülmek istenilen yapı ile ölçülen yapı arasındaki farkı ortaya koyma sürecinde MİF analizlerinin önemli sonuçlar ortaya koyduğu ifade edilmektedir. Dahası, MİF analizlerinden elde edilen sonuçlar, karşılaştırılan grupların birbirlerine göre zayıflıklarının ve güçlü yönlerinin nedenlerini belirlemede öne çıkmaktadır (Klieme ve Baumert, 2001).

Madde işlevinin farklılaşıp farklılaşmadığının belirlenmesi sürecinde yetenek düzeyleri eşitlenerek karşılaştırılan iki grup, referans ve odak grup olarak adlandırılırlar. İstatistiksel olarak, karşılaştırılan demografik gruplardan hangisinin referans grubu, hangisinin odak grup olduğu bir fark yaratmamasına rağmen, literatürde referans grup genellikle test veya maddenin lehine işlev gösterdiğinden şüphe edilen grup (örneğin çoğunluğu oluşturan grup veya erkekler), odak grup ise test veya maddenin özelliği bakımından dezavantajlı durumda olabileceği düşünülen grup (örneğin azınlığı oluşturan gruplar veya kadınlar) olarak seçilmektedir (Osterlind ve Everson, 2009).

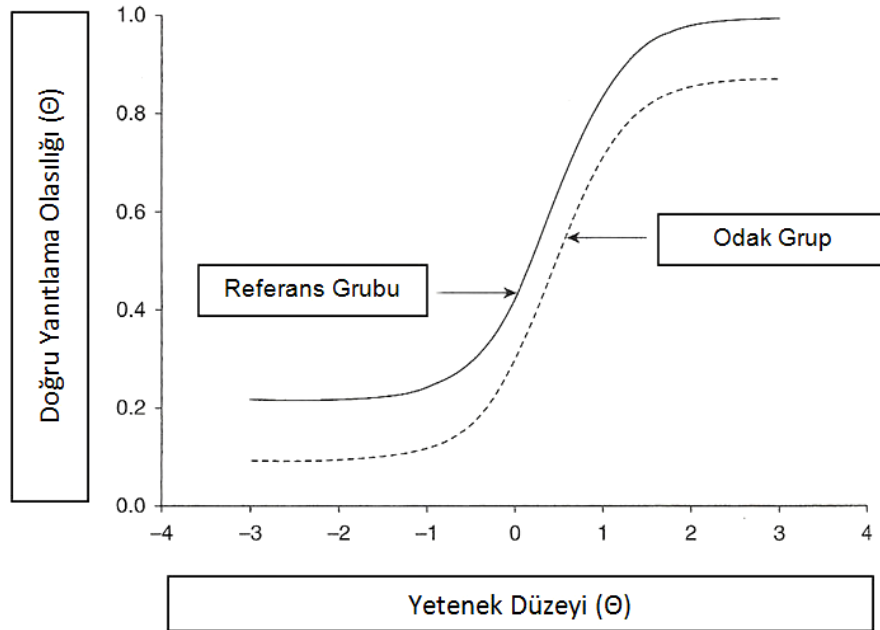
Referans ve odak gruplarının uygun şekilde belirlenmesinin ardından, söz konusu iki grubun bir değişkene göre, gösterdikleri performans anlamında *eşitlenmesi* gerekmektedir; bu süreç sıklıkla yetenek düzeylerinin eşitlenmesi olarak isimlendirilir. MİF analizleri ancak yetenek düzeyleri eşit olan iki grup arasında gerçekleştirilir. Yetenek düzeylerini eşitlemek için iki farklı yol bulunmaktadır (Karami ve Nodoushan, 2011) :

1. İçsel Eşitleme (Internal Matching): Bu eşitleme türünde eşitleme değişkeni, incelenen maddeleri içeren testten alınan toplam puandır.
2. Dışsal Eşitleme (External Matching): Bu eşitleme türünde ise eşitleme değişkeni, incelenen maddeleri içeren test dışında farklı bir testten alınan toplam puandır.

Madde işlevinin farklılaşması iki farklı şekilde gerçekleşmektedir. Söz konusu iki işlev farklılaşması türünün açıklanması araştırma için yararlı olacaktır.

### Madde İşlevinin Tek Biçimli Farklılaşması

Madde işlevinin tek biçimli farklılaşması, işlev farklılaşmasının en basit şeklidir. Bu durumda incelenen madde, yetenek düzeyinin her noktasında, bir gruba diğerine göre sürekli bir avantaj sağlamaktadır. Diğer bir ifade ile tek biçimli işlev farklılaşmasında, bir grubun maddeyi doğru cevaplama yüzdesi, diğer gruptan her yetenek düzeyinde sürekli ve tutarlı olarak daha yüksektir (Osterlind ve Everson, 2009). Madde işlevinin tek biçimli farklılaşmasında grupların maddeyi doğru cevaplama olasılıklarının arasında, tüm yetenek düzeylerinde sürekli bir bağımlılık söz konusudur ve bu durum Şekil 1'de gösterilmiştir.

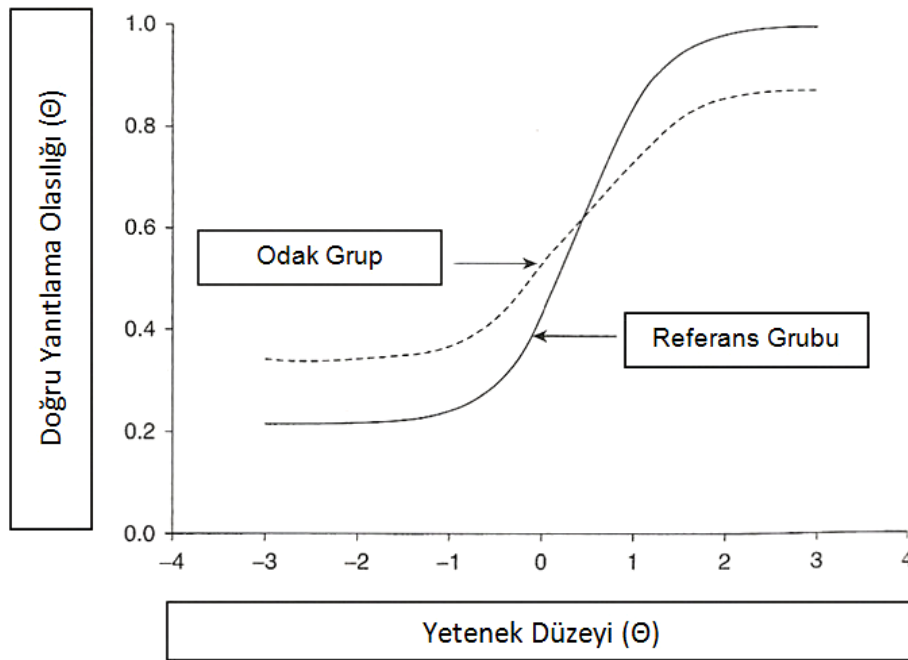


Şekil 1. Madde İşlevinin Tek Biçimli Farklılaşması (Osterlind ve Everson, 2009, sf. 12).

## Madde İşlevinin Tek Biçimli Olmayan Farklılaşması

Tek biçimli olmayan farklılaşma, farklı yetenek düzeylerinde maddeye doğru cevap verme olasılıklarının gruplar arasında düzey ve yön olarak farklılaştığı zamanlarda ortaya çıkar (Osterlind ve Everson, 2009). Yetenek düzeyi değiştikçe, grupların maddeye cevap verme olasılıkları da farklı düzey ve yönlerde değişmektedir.

Tek biçimli işlev farklılaşmasından ayrıldığı noktalar, maddenin her yetenek düzeyinde bir gruba diğerine göre aynı düzeyde ve yönde avantaj sağlaması durumunun geçerli olamaması ve gruplar arasında sürekli bir bağımlılığın söz konusu olmamasıdır. Maddenin işlevinin tek biçimli olmayan farklılaşması Şekil 2’de gösterilmiştir (Osterlind ve Everson, 2009).



Şekil 2: Madde İşlevinin Tek Biçimli Olmayan Farklılaşması (Osterlind ve Everson, 2009, sf.12).

## MİF Belirleme Yöntemleri

MİF belirlemek yöntemlerinin formal olarak kullanıldığı ilk uygulamalardan biri, 1963'te gerçekleştirilen Scholastic Aptitude Test'te (SAT) yer alan maddeler üzerinde Cardall ve Coffman tarafından gerçekleştirilmiştir.

Bu yıllarda başlayan formal uygulamalarla beraber günümüze dek MİF belirleme yöntemleri çeşitlilik kazanmış ve farklı bilim insanlarınca sınıflamalar geliştirilmiştir. Bu yöntemler sıklıkla Klasik Test Kuramı'na dayalı yöntemler ve Madde Tepki Kuramı'na dayalı olmak üzere iki grupta incelenmektedir. Araştırmanın amacı doğrultusunda Klasik Test Kuramı'na dayalı MİF belirleme yöntemleri aşağıdaki şekilde sıralanabilir: (Holland ve Wainer, 1993; Camilli ve Shephard, 1994).

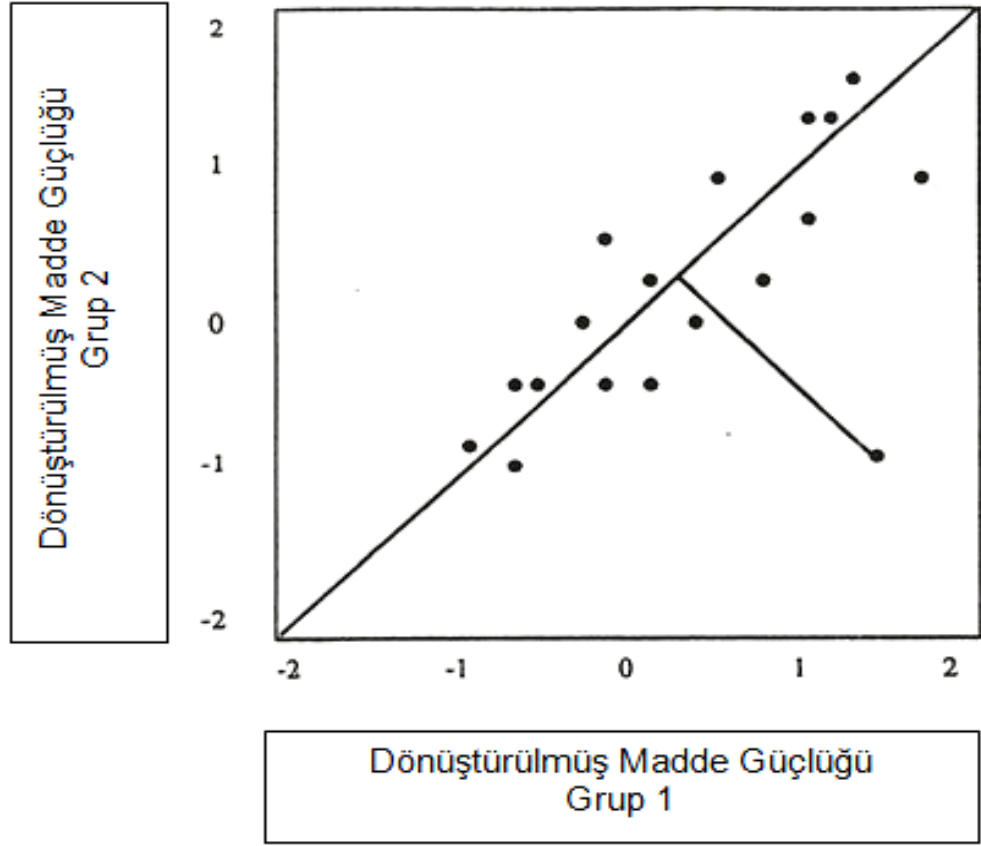
1. Dönüştürülmüş Madde Güçlük İndeksi (The Transformed Item Difficulty Index)
2. Varyans Analizi (ANOVA)
3. Korelasyona Dayanan Yöntemler
4. Mantel-Haenszel Analizi (MH)
5. Simultaneous Item Bias Test (SIBTEST)
6. Lojistik Regresyon Analizi (LR)

Araştırmada, Klasik Test Kuramı'na dayalı işlev farklılığı belirleme yöntemlerinden Mantel-Haenszel, SIBTEST ve lojistik regresyon yöntemleri, analizlerin yapılmasının ve yorumlanmasının kolaylığı, analizlerin gerçekleştirileceği yazılımların ulaşılabilir olması ve araştırmacının bu yöntemleri kullanmada yeterli olması açısından tercih edilmiştir.

### **1. Dönüştürülmüş Madde Güçlük İndeksi (The Transformed Item Difficulty Index)**

İşlev farklılaşması belirlemede kullanılan ilk yöntemlerden olan dönüştürülmüş madde güçlük indeksi, klasik test kuramına dayalı madde güçlük indeksinin ( $p$ ) karşılaştırılan gruplar için hesaplanması ile gerçekleştirilir. Angoff (1972) tarafından geliştirilen yöntemde, TID indeksini (Transformed Item Difficulty) hesaplamak için öncelikle, her bir grupta  $(1-p)$ 'ninci yüzdeye denk gelen  $z$  değeri hesaplanır. Ardından, hesaplanan  $z$  değerleri dönüştürme ile ortalaması 13, standart sapması 4 olan dağılım üzerine yerleştirilir. Fakat daha üzerinde çok çalışılması açısından elde edilen  $z$  puanı ortalaması 0, standart sapması 1 olan normal dağılıma da yerleştirilebilir. Karşılaştırılan iki grup için de yapılan bu işlem sonrasında, hesaplanan indeksler her iki grup için, Şekil 3'te görüldüğü dikdörtgen

üzerinde kesiştiği konumda nokta halinde işaretlenir (Camilli ve Shepard, 1994, sf. 26).



Şekil 3: Dönüştürülmüş Madde Güçlüğü İndeksi Yönteminde Kullanılan Çizelge

Şekil 3 incelendiğinde, iki grubu ayıran çizgiden uzaklaşan nokta (madde), sınırları içinde bulunduğu grubun aleyhine işlev farklılığı gösterir (Camilli ve Shepard, 1994). İki grubu ayıran çizgiden, diğer maddelere göre dikkate değer ölçüde uzaklaşan ve çizgi ile birleştirilen madde Grup 1 aleyhine MİF göstermektedir.

Dönüştürülmüş madde güçlüğü indeksi yöntemde gruplar ölçülen özellik anlamında eşitlenmediği (yetenek grupları oluşturulmadığı) için karşılaştırılan grupları ayırt ediciliği yüksek olan maddeler MİF gösteriyor şeklinde görülebilir. Bu nedenle, madde güçlük indeksinde dayalı olan MİF belirleme yöntemleri (dönüştürülmüş olsun ya da olmasın) yeterince geçerli sonuçlar sağlamaz (Camilli ve Shepard, 1994).

## 2. Varyans Analizi (ANOVA)

Varyans analizi (ANOVA) ile işlev farklılığı gösteren maddeleri belirleme yöntemi temelinde, madde güçlük indeksine dayalı ( $p$ ) MİF belirleme yöntemlerine benzerlik gösterir. Analiz sürecinde, bir faktörü gruplama değişkeni ve ikinci (grup içi) faktör olarak maddelerin seçildiği tekrarlı ölçümler için ANOVA kullanılır. Elde edilen sonuçlar çerçevesinde, ortalama grup farklılıklarının gruplar için temel etkide, gruplar için farklılaşan güçlüğü ise, grup-madde ortak etkisinde incelenmesi amaçlanır.

Analiz temelinde incelenen maddeye karşılaştırılan gruplarda doğru cevap verme yüzdelerine odaklandığından, dönüştürülmüş madde güçlük indeksi ile aynı problemlere sahiptir. İncelenen maddenin karşılaştırılan gruplarda güçlüğü ve ayırt ediciliği farklılaşıyorsa, söz konusu maddenin MİF gösterdiğini belirleme eğilimi gösterir. Gerçekte MİF göstermeyen maddelerin işlev farklılığı gösterdiğini belirleme özelliği nedeni ile günümüzde kullanımı önerilmemektedir (Camilli ve Shepard, 1994; Osterlind ve Everson, 2009).

## 3. Korelasyona Dayalı Yöntemler

İşlev farklılığı gösteren maddelerin belirlenmesinde 1970'li yıllarda sıklıkla, günümüzde ise nadiren kullanılan korelasyona dayalı yöntemler bulunmaktadır. Bu yöntemlerden ilkinde, karşılaştırılan iki grup için her bir maddenin güçlüğü hesaplanmakta ve en yüksek güçlükten en düşük güçlüğü göre maddeler sıralanmakta ve maddelerin "görelî güçlükleri" hesaplanmaktadır. Her iki grup için maddelerin görelî güçlük sıralamaları arasındaki korelasyona göre işlev farklılığı belirlenmektedir. Korelasyonun maksimum değeri olan 1.00'e yakın olan değerlerde, maddelerin görelî güçlüklerinin her iki grup için de benzer olduğu, dolayısı ile maddelerin her iki grupta benzer şekilde ölçme yaptığı sonucuna varılmaktadır. Dönüştürülmüş madde güçlüğü indeksi ve varyans analizine benzer şekilde madde güçlüğü dikkate alındığı ve karşılaştırılan gruplarda bulunan bireyler ölçülen özellik anlamında eşitlenmediği için MİF analizinde hangi düzeyde geçerli sonuçlar sağladığı tartışılmaktadır (Camilli ve Shepard, 1994).



İkinci korelasyona dayalı yöntemde, maddelerin ayırt edicilik indeksleri dikkate alınmakta ve ayırt edicilik indekslerinin hesaplanmasında madde ile test toplam puanı arasında nokta çift serili korelasyon katsayısı hesaplanmaktadır. Bu katsayının değerine göre, karşılaştırılan gruplarda incelenen maddelerin benzer şekilde ölçme yapıp yapmadığı belirlenmektedir. Green ve Draper (1972)'e göre, eğer bir madde, bir grup için maksimum ayırt edici olan maddeler arasındayken, diğer grup için minimum ayırt edici maddeler arasında yer alıyorsa, o maddenin yanlış ölçme yaptığından bahsedilir. Bu yöntemde kullanılan nokta çift serili korelasyon katsayısının, maddelerde ölçülen özelliğe göre gruplar büyük farklılıklar gösterdiğinde ve madde sayısının az olduğu durumlarda tutarsız sonuçlar sağladığı belirlenmiştir (Hunter, 1975). Bu anlamda yetenek düzeyleri eşitlenmeden gerçekleştirilen korelasyona dayalı yöntemlerin kullanılması önerilmemektedir (Camilli ve Shepard, 1994; Osterlind ve Everson, 2009).

#### **4. Mantel-Haenszel Yöntemi (MH)**

İki medikal araştırmacı olan Nathan Mantel ve William Haenszel tarafından 1950'li yıllarda eşleştirilmiş gruplarda uygulanacak olan bir Ki-kare prosedürü olarak geliştirilmiş ve 1985 yılında Holland tarafından güncellenerek madde işlev farklılaşmasının incelenmesi sürecinde kullanılması için önerilmiştir. Ki-kare metodunun 2×2'lik matrislerde satır ve sütunlarda yer alan değişkenler arasındaki lineer ilişkinin yardımı ile gruplar arası karşılaştırmalar yapmak amacıyla kullanılmaktadır ve günümüzde en sık kullanılan MİF belirleme yöntemlerinden biridir (Osterlind ve Everson, 2009).

Referans ve odak grupların karşılaştırılması sürecinde Ki-Kare istatistiğinin kullanıldığı yöntemde, Ki-Kare'nin örneklem büyüklüğüne üst düzeyde bağımlı olması problemini çözmek için Cochran'ın istatistiği ile düzeltme yapılır. Bu sebeple küçük boyutlu örneklerde kullanılabilir, yorumlanması kolaydır ve Madde Tepki Kuramı'na dayalı MİF belirleme yöntemlerinden elde edilen sonuçlar ile tutarlı sonuçlar sağladığı belirlenmiştir (Bertrand ve Boiteau, 2003). Mantel-Haenszel analizinin en büyük sınırlılığı yalnızca tek biçimli madde işlevi farklılaşmasını belirleyebilmesidir.

Analiz sürecinde öncelikle referans ve odak grupta yer alan cevaplayıcıların testten aldıkları toplam puanlara göre eşleştirme yapmak amacıyla yetenek grupları oluşturulur. Bu yöntemde sıklıkla 4 ya da 5 yetenek grubu oluşturulurken, heterojen puan dağılımlarında yetenek grubu sayısı artırılabilir. İncelenen her bir madde ve yetenek düzeyleri için Şekil 4'te görüldüğü gibi 2×2'lik matrisler oluşturulur.

	1	0
Referans Grubu (R)	$a_i$	$b_i$
Odak Grubu (O)	$c_i$	$d_i$
Toplam	$N_{1i}$	$N_{0i}$
	$N_{1i} = a_i + c_i$	$N_{0i} = b_i + d_i$

Şekil 4: *i* Maddesi İçin *j* Yetenek Düzeyinde Mantel-Haenszel Matrisi

$a_i$ : Referans grubunda *i* maddesine doğru cevap veren cevaplayıcı sayısı

$c_i$ : Odak grubunda *i* maddesine doğru cevap veren cevaplayıcı sayısı,

$b_i$ : Referans grubunda *i* maddesine yanlış cevap veren cevaplayıcı sayısı,

$d_i$ : Odak grubunda *i* maddesine yanlış cevap veren cevaplayıcı sayısı,

$N_{1i}$ : *i* maddesine doğru cevap veren toplam cevaplayıcı sayısı,

$N_{0i}$ : *i* maddesine yanlış cevap veren toplam cevaplayıcı sayısıdır.

$$\alpha_i = \frac{p_{ri}/q_{ri}}{p_{oi}/q_{oi}} = \frac{\frac{a_i/(a_i + b_i)}{b_i/(a_i + b_i)}}{\frac{c_i/(c_i + d_i)}{d_i/(c_i + d_i)}} = \frac{a_i/b_i}{c_i/d_i} = \frac{a_i d_i}{b_i c_i}$$

İşlev farklılaşmasının düzeyini belirlemek için farklı ölçütler bulunmakla beraber, Çizelge 1'de gösterilen  $\alpha_i$  değerine göre madde ile ilgili aşağıda yer alan yorumlar yapılabilir (Holland ve Wainer, 1993):

## Çizelge 1

### *Mantel-Haenszel Analizi İle Elde Edilen $\alpha_i$ Değerleri İçin Kabul Edilen Ölçütler*

$\alpha_i$ Değeri	MİF Düzeyi
$\alpha_i \leq 1$	A Düzeyi – Göz ardı edilebilir düzey
$1 < \alpha_i < 1.5$	B Düzeyi – Orta Düzey
$1.5 \leq \alpha_i$	C Düzeyi – Üst Düzey

Çizelge 1’de yer verilen  $\alpha_i$  değerine göre incelenen maddenin işlev farklılığı gösterip göstermediğine karar verilir.  $\alpha_i > 1$  ise, incelenen yetenek düzeyinde, referans gruptaki cevaplayıcıların maddeyi doğru cevaplama olasılığı odak gruptaki cevaplayıcılardan anlamlı düzeyde daha yüksektir.  $\alpha_i < 1$  ise, incelenen yetenek düzeyinde, odak gruptaki cevaplayıcıların maddeyi doğru cevaplama olasılığı referans gruptaki cevaplayıcılardan anlamlı düzeyde daha yüksektir (Osterlind ve Everson, 2009).

### **5. Simultaneous Item Bias Test (SIBTEST)**

SIBTEST, Shealy ve Stout (1993) tarafından geliştirilmiş, MİF belirlemek için kullanılan diğer analizlerle kıyaslandığında daha yeni olan bir yöntemdir. Yöntemin farklılığı, standartlaştırılmamış Mantel-Haenszel analizinden elde edilen değerleri lineer regresyon ile düzeltilmesidir. Bu yöntemde, referans ve odak gruplarının incelenen maddeden elde ettikleri ortalamalar, testten elde edilen toplam puanlar dikkate alınarak düzeltilir. Böylece yöntemle, ham test puanlarından hesaplama yapan Mantel-Haenszel ve benzer yöntemlere kıyasla referans ve odak gruplarını daha isabetli bir şekilde eşitlemesi söz konusu olur (Osterlind ve Everson, 2009).

SIBTEST incelenen her madde için aşağıda yer verilen  $H_0$  ve  $H_1$  hipotezlerini test eder:

$$H_0: \beta_{UNI} = 0$$

$$H_1: \beta_{UNI} \neq 0$$

Hipotez test etme sürecinde,  $\beta_{UNI}$  MİF düzeyi olarak tanımlanır ve şu şekilde hesaplanır:

$$\beta_{UNI} = \int B(\theta) f_f(\theta) d(\theta)$$

$B(\theta)$ : Sırası ile referans ve odak gruplarında yer alan cevaplayıcıların maddeye doğru cevap verme olasılıkları,

$f_f(\theta)$ : Odak grubun maddeyi doğru cevaplama olasılığı için yoğunluk fonksiyonu,

$d(\theta)$ :  $\theta$  değerinin referans ve odak gruplarda farklılaşması olarak tanımlanır (Zhou, Gierl ve Tan, 2006).

Analiz neticesinden elde edilen  $\beta_{UNI}$  değeri pozitif olduğunda maddenin referans grubu lehine, negatif ise odak grup lehine MİF gösterdiği şeklinde yorumlanır. MİF düzeyini belirlemek için ise, Roussos ve Stout (1996) tarafından Çizelge 2’de verilen şu ölçütler kullanılmaktadır:

Çizelge 2

*SIBTEST Analizi İle Elde edilen  $\beta_{UNI}$  Değerleri İçin Ölçütler*

$\beta_{UNI}$ Değeri	MİF Düzeyi
$\beta_{UNI} < 0.059$	A Düzeyi – Göz ardı edilebilir düzey
$0.059 \leq \beta_{UNI} < 0.088$	B Düzeyi – Orta Düzey
$0.088 \leq \beta_{UNI}$	C Düzeyi – Üst Düzey

Çizelge 2’de yer alan  $\beta_{UNI}$  değerleri dikkate alınarak incelenen maddenin hangi düzeyde MİF gösterdiği ve hangi grup lehine işlev farklılığı gösterdiği belirlenir. SIBTEST yöntemi, lojistik regresyon yöntemine benzer şekilde hem tek biçimli hem de tek biçimli olmayan işlev farklılığını belirleyebilmektedir.

## 6. Lojistik Regresyon Yöntemi

Günümüzde diğer MİF belirleme yöntemleri ile beraber, etkili sonuçlar sağladığı belirlenmiş ve kullanılması önerilen yöntemlerden birisi lojistik regresyon analizidir (Clauser ve Mazor, 1998; Swaminathan ve Rogers, 1990). Bu yöntemde iki kategorili bir bağımlı değişken (maddeden alınan 0 – 1 puan), bir gruplama değişkeni (1=referans grubu, 2= odak grubu), testten alınan toplam puan ve gruplama değişkeni ile toplam puanın ortak etkisinin incelendiği bir eşitlik kurulur. MİF belirlemek amacıyla kurulan lojistik regresyon eşitliği şu şekilde ifade edilir (Zumbo, 1999);

$$Y = b_0 + b_1(\text{Toplam Puan}) + b_2(\text{Gruplama Değişkeni}) \\ + b_3 (\text{Toplam Puan} * \text{Gruplama Değişkeni})$$

Bu analiz sürecinde, ilk adımda yalnızca toplam puan değişkeninin ( $R^2$  [Adım 1]), ikinci adımda toplam puan ve gruplama değişkenlerinin beraber ( $R^2$  [Adım 2]), son adımda ise toplam puan değişkeni, gruplama değişkeni ve toplam puan ile gruplama değişkeninin ortak etkisinin beraber ( $R^2$  [Adım 3]), bağımlı değişkeni hangi düzeyde yordadığı incelenir. Söz konusu üç yordama basamağının farkları dikkate alınarak etki düzeyleri hesaplanır. Etki düzeyleri 2 serbestlik derecesinde Ki-Kare istatistiği ile karşılaştırılarak incelenen madde ile ilgili MİF değerlendirilmesi yapılır.

MİF düzeyinin belirlenmesi için çeşitli ölçütler bulunmakla beraber, sık kullanılan ve araştırmada kabul edilen ölçütler, Çizelge 3'te yer alan şekilde sunulabilir (Zumbo ve Thomas, 1997):

$$\Delta R^2 = R^2[\text{Adım 3}] - R^2[\text{Adım 1}] \text{ olmak üzere,}$$

### Çizelge 3

*Lojistik Regresyon Analizi İle Elde Edilen  $\Delta R^2$  Değerleri İçin Kabul Edilen Ölçütler*

$\Delta R^2$ Değeri	MİF Düzeyi
$\Delta R^2 < 0.13$	A Düzeyi – Göz ardı edilebilir düzey
$0.13 \leq \Delta R^2 < 0.26$	B Düzeyi – Orta Düzey
$0.26 \leq \Delta R^2$	C Düzeyi – Üst Düzey

Çizelge 3'te yer verilen  $\Delta R^2$  değerlerine göre incelenen maddenin jangi düzeyde ve hangi grubun lehine işlev farklılığı gösterdiği belirlenmektedir. Lojistik regresyon analizinin diğer MİF belirleme yöntemlerine göre avantajları, sürekli "yetenek düzeyi" değişkenini kategorize etmeye gerek olmaması, tek biçimli ve tek biçimli olmayan işlev farklılaşmasını belirleyebilmesi ve iki uçlu yanıtlanmayan dereceli maddeler için de analizin gerçekleştirilebilmesidir. Bu avantajlarının yanında lojistik regresyon analizinin önemli bir sınırlılığı, tutarlı sonuçlar sağlamak için Mantel-Haenszel analizine göre daha büyük örneklemelere ihtiyaç duymasındır (Holland ve Wainer, 1993; Zumbo, 1999).

### **Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS)**

Hollanda merkezli bir kuruluş olan International Association for the Evaluation of Educational Achievement (IEA) yaklaşık 50 yıldır, öğrenci başarısının belirlenmesine yönelik uluslararası çalışmalar gerçekleştirmektedir. İlki, Birinci Uluslararası Matematik ve Fen Araştırması adıyla 1959 yılında düzenlenen bu çalışma 2003 yılından itibaren Uluslararası Matematik ve Fen Eğilimleri Araştırması adını almıştır (Şişman, Acat ve Aypay, 2011). TIMSS adı ile ilk uygulama 1995 yılında gerçekleştirilmiş ve ardından her dört yılda bir, dünyanın farklı ülkelerinde dördüncü ve sekizinci sınıfta bulunan öğrencilerin matematik ve fen bilimleri başarısını değerlendirme ve karşılaştırma sürecinde bir standart haline gelmiştir. Çalışmanın dört yıllık periyotlar halinde gerçekleştirilmesindeki temel amaç, ilk uygulamada dördüncü sınıfta eğitim gören öğrencilerin bir sonraki uygulamada sekizinci sınıfta eğitim görüyor olacak olması, dolayısı ile aynı katılımcı grubunun birden fazla sayıda değerlendirmeye tabi tutularak boylamsal bir çalışma gerçekleştirilmesine imkân sağlamasıdır (Olson, Martin ve Mullis, 2008).

Türkiye TIMSS uygulamasına ilk kez 1999 yılında sekizinci sınıf düzeyinde katılmıştır. Bu uygulama sonrasında elde edilen sonuçlara göre Türkiye, fen bilimleri testinden ülke olarak ortalama 433 puan ile düşük düzeyde başarı gösteren ülkeler arasında yer almaktadır. TIMSS uygulamasından elde edilen sonuçlar, uluslararası öğrenci başarısını değerlendirme sürecinde önemli diğer bir uygulama olan PISA sonuçları ile

beraber, eğitim sisteminin etkililiği ve bu eğitim sisteminin belli kademelerinde eğitim gören öğrencilerin başarılarının uluslararası alanda hangi düzeyde olduğunun belirlenmesi anlamında önemli görülmüştür.

İkinci kez 2007 yılında, yine sekizinci sınıflar düzeyinde TIMSS uygulamasına katılan Türkiye, fen bilimleri testinden ülke olarak ortalama 457 puan almış, TIMSS 1999 uygulaması fen bilimleri testi sonuçları ile kıyaslandığında ortalama puanını 21 puan artırmış ve fen bilimleri testinde tüm katılımcı ülkelerin ortalama puanına ( $\bar{X} = 465$ ) yaklaşmasına rağmen, TIMSS 1999 uygulaması sonuçlarına benzer şekilde düşük düzeyde başarı gösteren ülkeler arasında yer almıştır (Şişman, Acat ve Aypay, 2011). İngiltere ise TIMSS adı ile gerçekleştirilen ilk uygulama olan TIMSS 1995'den başlamak üzere, sırası ile 1999, 2003 ve 2007 uygulamalarına katılmıştır ve bu uygulamalarda fen bilimleri ortalama puanları sırası ile 533, 538, 544 ve 542 olarak hesaplanmıştır. Bu puanlar, İngiltere'yi süreç boyunca fen bilimleri alanında en başarılı beşinci ülke yapmıştır (Martin, Mullis ve Foy, 2008).

Günümüzde önemi ve sayısı hızla artan uluslararası düzeyde fen bilimleri başarısının ve okuryazarlığının karşılaştırılmasına dayalı çalışmalarda; testin tümünden ve testi oluşturan maddelerin her birinden elde edilen puanların geçerli ve karşılaştırılabilir olması büyük önem teşkil etmektedir. Söz konusu testlerden elde edilen sonuçlar ve ülkelerin başarı sıralamaları, eğitim sistemlerinde radikal kararlara ve değişimlere dayanak sağlayabilmektedir. Bu bağlamda elde edilen sonuçların, eğitim alanında son yıllarda yapılan köklü değişiklikler gerçekleştiren Türkiye ile farklı dillerde eğitim veren diğer ülkeler arasındaki eşdeğerliğinin incelenmesi önemli görülmüştür. Uluslararası uygulamalara yönelik yanlılık çalışmalarının sıklıkla matematik alanında yoğunlaşması ve TIMSS maddelerine yönelik yanlılık çalışmalarının sayıca azlığı, araştırmanın TIMSS fen bilimleri testi üzerinde gerçekleştirilmesine gerekçe olmuştur. Dile yönelik yanlılık incelemesinde Türkiye ve İngiltere örneklemi kullanılması, TIMSS uygulaması kapsamında maddelerin ilk haliyle yalnızca İngilizce geliştirildiği dikkate alındığında, maddeleri orijinal dilinde cevaplayan öğrencilerin puanları ile uyarlanmış maddeleri cevaplayan öğrencilerin puanlarının hangi ölçüde karşılaştırılabilir olduğunun belirlenmesine imkân tanımış ve farklı uluslararası yeterlik

düzeylerinde bulunan iki ülkenin karşılaştırılmasına olanak sağlamıştır. Maddeleri, geliştirildikleri orijinal dilde cevaplayan iki ülke olan İngiltere ve Amerika arasında, Amerika ile Türkiye örneklerinin karşılaştırıldığı araştırma sayısının daha fazla olması sebebi ile İngiltere tercih edilmiştir. Cinsiyete yönelik yanlılık incelemesinde ise, Türkiye’de cinsiyet gruplarının farklılaşan eğitim yaşantı ve fırsatlarına sahip oldukları, toplumsal cinsiyet rollerinin eğitim sürecinde de dikkate değer ölçüde devam ettiği düşünüldüğünden inceleme Türkiye örneğinde gerçekleştirilmiştir.

## AMAÇ

Bu araştırmada TIMSS 2007 fen bilimleri testinde maddelerin Türkiye - İngiltere örneklemi kapsamında dile göre ve Türkiye örnekleme kapsamında cinsiyete göre işlev farklılığı gösterip göstermediğinin, birden fazla yöntemle (Mantel-Haenszel, lojistik regresyon ve SIBTEST) incelenmesi amaçlanmıştır. İlgili analizler sonucunda olası yanlılık gösteren ve araştırmacıların erişimine açılmış maddelerin olası yanlılık kaynakları hakkında uzman görüşüne başvurulmuştur. Bu genel amaç doğrultusunda şu sorulara yanıt aranmıştır:

1. TIMSS 2007 uygulaması Türkiye ve İngiltere örnekleme kapsamında fen bilimleri testinde yer alan maddelerin işlevleri
  - 1.a. Mantel-Haenszel yöntemi,
  - 1.b. SIBTEST yöntemi,
  - 1.c. Lojistik regresyon yöntemiile yapılan analizlere göre, *dile* dayalı farklılık göstermekte midir?
2. TIMSS 2007 uygulaması Türkiye örnekleme kapsamında fen bilimleri testinde yer alan maddelerin işlevleri
  - 2.a. Mantel-Haenszel yöntemi,
  - 2.b. SIBTEST yöntemi,
  - 2.c. Lojistik regresyon yöntemiile yapılan analizlere göre, *cinsiyete* dayalı farklılık göstermekte midir?
3. Fen Bilimleri testinde *dile* göre işlev farklılığı gösterdiği belirlenen, araştırmacıların erişimine açılmış maddelerin olası yanlılık kaynağı konusundaki uzman görüşleri nelerdir?



4. Fen Bilimleri testinde *cinsiyete* göre işlev farklılığı gösterdiği belirlenen, araştırmacıların erişimine açılmış maddelerin olası yanlılık kaynağı konusundaki uzman görüşleri nelerdir?

## ÖNEM

Bu araştırmada, madde işlevinin farklılaşması hakkında bilgi edinmek üzere farklı özelliklere sahip olan üç yönteme (Mantel-Haenszel, SIBTEST ve lojistik regresyon) başvurulmuştur. MİF çalışmalarında birden çok yönteme dayalı inceleme yapılması önerilmektedir (Holland ve Wainer, 1993; Osterlind ve Everson, 2009). Söz konusu yöntemler kullanılarak yapılan analiz sonuçlarının tutarlılığının belirlenmesi adına alan yazında önem teşkil etmektedir.

TIMSS 2007 fen bilimleri testi Ocak 2005 – Kasım 2006 sürecini kapsayan oldukça yoğun ve ciddi bir test geliştirme sürecinin sonucunda hazırlanmıştır. Bu süreçte, maddelerin yazılması ve niteliklerinin test edilmesi için IEA bünyesinde yer alan eğitim uzmanları görev almıştır. Bu sistematik test geliştirme sürecinin, madde yanlılığına hangi düzeyde engel olabildiğinin belirlenmesi araştırmacı açısından önemli görülmüştür.

Türkiye TIMSS uygulamalarına 1999, 2007 ve 2011 yıllarında katılmıştır ve 2015 yılında gerçekleştirilecek TIMSS uygulaması için de görüşmeler sürmektedir. 1999 ve 2007 yılında yapılan TIMSS uygulamalarına yönelik Türkiye’de ve yurtdışında birçok çalışma yapılmıştır. Öğrenci başarısının uluslararası düzeyde belirlenmesi ve karşılaştırılması sürecinde TIMSS ve PISA gibi uygulamalar, eğitim politikalarına yön verme özelliği olan ve bu temel gerekçe ile dünya çapındaki eğitim uzmanları tarafından büyük önem verilen uygulamalardır.

Yanlılığın ölçme sonuçlarını etkileyen sistematik bir hata olduğu, bu hatanın ölçme sonuçlarının ve bu sonuçlara dayalı olan kararların geçerliğini tehdit ettiği göz önüne alındığında; TIMSS gibi her dört yılda bir kez, dünya çapında elliden fazla ülkede uygulanan bir sınavdan alınan sonuçların dil ve cinsiyet farklılığına göre ne derece karşılaştırılabilir olduğunun bulunması araştırmayı, bu uygulamanın sonuçlarına dayalı olarak verilecek eğitim

kararlarının isabetliliğini etkilemesi boyutunda önemli kılmaktadır. Üst uluslararası yeterlik düzeyinde yer alan İngiltere ile alt uluslar arası yeterlik düzeyinde bulunan Türkiye'nin karşılaştırılması, farklılaşan fen bilimleri başarıları gösteren iki ülkenin karşılaştırılmasına imkan sağlamış, İngiltere örneğinde TIMSS fen bilimleri orijinal maddelerini cevaplayan öğrenciler ile uyarlanan maddeleri cevaplayan Türk öğrencilerin puanlarının karşılaştırılabilir olup olmadığının belirlenmesine olanak sağlamıştır. Türkiye örneğinde ise erkek ve kız öğrencilerin eğitim fırsat ve yaşantılarının farklılaştığı, toplumsal cinsiyet rollerinin eğitim yaşantıları üzerinde etkili olduğu düşünüldüğünden cinsiyete yönelik MİF analizlerinde Türkiye örneklemini kullanılmıştır. Bu anlamda, TIMSS 2007 fen bilimleri testindeki maddelerden elde edilen puanların Türkiye örnekleminde cinsiyet gruplarına göre hangi düzeyde karşılaştırılabilir olduğunun belirlenmesi açısından önemli görülmektedir.

Bu araştırma, TIMSS ve uluslararası öğrenci başarılarını ölçmeyi ve karşılaştırmayı amaçlayan benzer çalışmaların gelecekteki uygulamalarının daha geçerli sonuçlar sağlaması amacıyla ulusal ve uluslararası düzeyde alınacak önlemlere de yol gösterici olabilecektir.

## **SAYILTILAR**

Araştırmada, madde işlevlerinin dil ve cinsiyete göre farklılaşıp farklılaşmadığının belirlenmesi için, MİF belirleme sürecinin ve bu süreçte kullanılan yöntemlerin (Mantel-Haenszel, SIBTEST ve lojistik regresyon) bir gereği olarak yetenek düzeylerinin eşitlenmesi gerekmektedir. Bu amaçla, TIMSS 2007 fen bilimleri testi sonuçlarına göre elde edilen beş olası fen bilimleri başarı puanının (Science Plausible Values) ortancasının, söz konusu beş olası puanı temsil edildiği kabul edilmiş ve fen bilimleri başarı puanı olarak dikkate alınmıştır. Türkiye ve İngiltere örneklemlerindeki öğrencilerin fen bilimleri başarı puanları eşit aralıklı dokuz farklı yetenek düzeyine ayrılmıştır.

## SINIRLILIKLAR

Bu araştırma, TIMSS 2007 fen bilimleri testinde 14 farklı kitapçıkta yer alan toplam 107 çoktan seçmeli madde ile sınırlıdır. Araştırmacının yöntemleri kullanmadaki yeterliği ve yöntemlerin analiz ve yorumlama kolaylığı göz önüne alınarak araştırma Mantel-Haenszel, SIBTEST ve lojistik regresyon yöntemleri ile sınırlandırılmıştır.

Dile göre MİF gösteren 11 farklı maddeden 5 tanesi araştırmacıların erişimine açıldığından, uzman görüşleri söz konusu beş madde ile sınırlandırılmıştır.

Dile yönelik yanlılık analizleri İngiltere ve Türkiye örneklemi ve cinsiyete yönelik yanlılık analizleri Türkiye örnekleme ile sınırlıdır.

## TANIMLAR

**Madde Etkisi (Item Impact):** Maddelerin referans ve odak gruplarda bulunan cevaplayıcılar için işlev farklılığı gösterdiği fakat yanlılık göstermediği durumlar için kullanılmıştır.

**Madde İşlevinin Farklılaşması (Differential Item Functioning):** Mantel-Haenszel, SIBTEST ve lojistik regresyon yöntemleri aracılığıyla, maddelerin referans ve odak gruplarında bulunan öğrencilerin doğru cevaplama olasılıklarının farklılaştığı durumları tanımlamaktadır.

**Yanlılık (Bias):** İşlev farklılığı gösterdiği belirlenen maddelerin uzman görüşleri sonucunda, bir grubun diğeri üzerinde avantaj veya dezavantaj sağladığını belirtmesi sonucu oluşan durumu ifade etmektedir.

## KISALTMALAR

**IEA:** International Association for the Evaluation of Educational Achievement

**MİF:** Madde İşlevinin Farklılaşması.

**NRC:** National Research Council

**SMIRC:** Science and Mathematics Item Review Committee

**TIMSS:** Trends In International Mathematics and Science Study

## BÖLÜM II

### İLGİLİ ARAŞTIRMALAR

Bu kısımda, uluslararası düzeyde öğrenci başarısının belirlenmesi ve karşılaştırılması amacı ile gerçekleştirilen TIMSS ve uluslararası düzeyde öğrencilerin farklı alanlardaki okuryazarlık düzeylerini ölçmeyi amaçlayan PISA çalışmaları için gerçekleştirilen MİF ve madde yanlılığı konulu araştırmalar, kültürler arası eşdeğerliğe dayalı çalışmalar ve dil ile fen bilimleri başarısı ilişkisini konu edinen çalışmalar kısaca özetlenmiştir. Araştırmalar, yurtdışı ve yurtiçinde gerçekleştirilmiş olmasına göre gruplanmış ve bu gruplar dahilinde araştırma yıllarına göre sıralanmıştır.

#### **Yurtdışında Gerçekleştirilen Araştırmalar**

Klieme ve Baumert (2001), 1995 yılında gerçekleştirilen TIMSS uygulaması matematik testinde Avusturya, Fransa, Almanya, İsveç, İsviçre ve Amerika Birleşik Devletleri'nden katılımcı öğrencilerin puanlarının karşılaştırılabilirliğini araştırmıştır, MIF belirleme süreçlerini işe koşarak, ülkelerin performanslarında güçlü ve zayıf konuları belirlemeyi amaçlamışlardır. Araştırmacıların çalışma öncesindeki tahminlerine paralel şekilde, farklı matematik eğitimi süreçlerine sahip söz konusu altı ülkenin ikili karşılaştırmalarında orta ve yüksek düzeyde MİF gösteren maddeler bulunmuştur. Sonuçlar, Amerika Birleşik Devletleri'nin matematiksel muhakeme düzeyinde yer alan maddelerde, İsveçli öğrencilerin uygulama ve problem çözme becerilerinde üst düzeyde başarı gösterdiği, Alman öğrencilerin bilgi düzeyindeki maddelerde daha başarısız, görsel yorumlama gereken maddelerde daha başarı olduğu, Almanca konuşan üç ülke olan İsviçre, Avusturya ve Almanya'nın maddelerde genel olarak benzer bir profil gösterdiği bulunmuştur. Araştırmacılar, ülkeler arasındaki performans farklılığının olası nedenleri olarak; farklı eğitim gelenekleri ve kültürel çerçevelerinin sınıf içindeki performansı etkilediğini, bu durumda MIF'in olası nedenlerinden olduklarını belirtmişlerdir.

Gierl ve Khaliq (2001), tercüme edilen başarı testlerinin MİF gösterme nedenlerinin inceledikleri çalışmada, 1997 yılında uygulanan "Matematik ve

Sosyal Bilimler Başarı Testi'nin verilerini kullanmışlardır. Çalışmada test geliştirme alanında uzman olan 11 kişi ve sertifikalı iki çevirmen ile toplam 13 kişilik bir komite, maddeleri inceleyerek olası yanlış ölçme yapabilecek maddeleri ve söz konusu maddelerin hangi grubun lehine ölçme yapabileceğine dair görüş bildirmişlerdir. Ardından yapılan MİF analizleri göstermiştir ki, MİF gösteren matematik maddelerin 8'inden 7'si ve sosyal bilimler testinde MİF gösteren 13 maddeden 8'i uzman ekip tarafından doğru belirlenmiştir. MİF gösteren maddeler incelendiğinde, kelime ve ifade değişikliklerinin gerçekleştirildiği veya dil/kültür ile uygun olmayan yapılar kullanıldığı belirlenmiştir.

Zhang (2002), Delaware Öğrenci Sınav Programı (Delaware Student Testing Program) çerçevesinde matematik testinde bulunan çoktan seçmeli maddelerin cinsiyet ve etnik gruplarına (siyahî, Asya kökenli, İspanyol kökenli ve beyaz) göre işlev farklılaşması gösterip göstermediğini incelemiştir. Analizlerde 3.,5., 8. ve 10. sınıflarının yanıtladığı çoktan seçmeli maddeler kullanılmıştır. Sekizinci sınıfta eğitim gördüğü öğrencilerin yanıtladığı maddeler için yapılan analizlerde, cinsiyete yönelik MİF gösteren iki maddenin bulunduğu; bu maddelerden birinin kız öğrenciler lehine, diğerinin ise erkek öğrenciler lehine işlev farklılığı gösterdiği belirlenmiştir. Asya kökenli öğrenciler arasında cinsiyete göre C düzeyinde işlev farklılığı gösteren dokuz madde bulunmuş: altı maddenin erkek öğrenciler, üç maddenin kız öğrenciler lehine MİF gösterdiği ifade edilmiştir. Siyahî öğrenciler arasında dört madde B düzeyinde MİF göstermiş, bu maddelerden üçü erkek öğrenciler lehine biri kız öğrenciler lehine ölçme yapmıştır. İspanyol kökenli öğrenci grubunda 11 MİF gösteren madde olduğu belirlenmiştir, altı maddenin kız öğrenciler lehine, beş maddenin erkek öğrenciler lehine işlev farklılığı gösterdiği belirlenmiştir. Beyaz öğrenciler arasında MİF gösteren maddeler üç olarak belirlenmiş, diğer gruplara göre daha az olduğu ifade edilmiştir. Bu maddelerden ikisinin erkekler, birinin kız öğrenciler lehine işlev farklılığı gösterdiği belirlenmiştir.

Nillas (2002), TIMSS-R sınavı verilerini kullanarak dil ile matematik arasındaki ilişkiyi incelemiştir. Araştırma çerçevesinde, 7. sınıfta iki dilde (Filipince ve İngilizce) eğitim gören 6601 Filipinli öğrencinin TIMSS-R'da elde ettikleri matematik başarıları incelenmiştir. Araştırma sonuçları göstermiştir ki,

TIMSS-R İngilizce uygulamasına katılan öğrenciler, Filipince uygulamaya katılan öğrencilerden anlamlı olarak daha yüksek matematik performansı göstermişlerdir.

Zenisky, Hambleton ve Robin (2003) gerçekleştirdikleri çalışmada, Amerika'da devlet tarafından gerçekleştirilen standart fen bilimleri sınavına katılan ilkokul, ortaokul ve lise düzeyinde yaklaşık 360.000 öğrencinin verileri dahilinde, testte yer alan maddelerin cinsiyete yönelik MİF gösterip göstermediklerini belirlemeyi amaçlamış ve MİF gösteren maddelerin hangi özellikler taşıdığını incelemiştir. Çalışmada iki basamaklı ağırlıklandırılmış madde gücü karşılaştırmasına dayalı olarak MİF incelemesi yapılmıştır ve testte hem yapılandırılmış açık uçlu sorular hem de çoktan seçmeli, toplam 222 maddenin yer aldığı belirtilmiştir. Sonuçlara göre, tüm sınıf düzeylerinde 222 maddenin 162'si (%73) MİF göstermemiş, 23'ü (%10.4) erkek öğrenciler lehine düşük düzeyde; 29'u (%13) orta ve üst düzeylerde MİF göstermiştir. Buna karşılık kız öğrenciler lehine düşük düzeyde MİF gösteren madde sayısı 8 (%3.6) olarak elde edilmiş ve kız öğrenciler lehine orta ve yüksek düzeyde MİF gösteren madde bulunmamıştır. MİF gösteren maddeler, içeriksel, görsel ve madde tipi kategorilerinde detaylı olarak incelendiğinde; çoktan seçmeli maddelerin erkeklerin lehine, açık uçlu yapılandırılmış maddelerin ise kız öğrencilerin lehine işlev farklılaşması gösterme eğiliminin olduğu ifade edilmiştir.

Bertrand ve Boiteau (2003) gerçekleştirdikleri çalışmada 1999 TIMSS uygulamasında Amerika ve Japonya örneklemelerini kullanarak madde tepki kuramına dayalı Alan metodu ve NCDIF ve klasik test kuramına dayalı lojistik regresyon ve Mantel-Haenszel (MH) yöntemlerini karşılaştırmayı amaçlamışlardır. Araştırma sonucunda madde tepki kuramı ve klasik test kuramına dayalı yöntemler birbirine oldukça yakın sonuçlar vermiş ve özellikle Mantel-Haenszel yöntemi madde tepki kuramına dayalı tekniklerle eşdeğer düzeyde tutarlı sonuçlar sağlamıştır.

Stoneberg (2004), Idaho Standart Başarı Testi'nin (Idaho Standarts Achievement Test) Okuma, Matematik ve Dil Kullanımı alt testlerinde 4., 8. ve 10. sınıf uygulamalarında yer alan maddelerin cinsiyet ve etnik yönden MİF

gösterip göstermediğini belirlemiştir. Cinsiyete yönelik MİF belirleme sürecinde odak grup olarak, kız öğrenciler, referans grubu olarak erkek öğrenciler; etnik MİF belirleme sürecinde İspanyol öğrenciler odak, beyaz öğrenciler ise referans grup olarak dikkate alınmıştır. Araştırmacı, MİF belirleme sürecinde SIBTEST ve Mantel-Haenszel yöntemlerini kullanmıştır. Araştırma sonucunda cinsiyete göre MİF gösteren maddeler, 8. sınıf uygulamasında, Okuma, Dil Kullanımı ve Matematik testlerinde yer alan tüm maddelerin sırası ile %42.9, %28.6 ve %50'sini oluşturmuştur. Araştırmacı, söz konusu maddelerin MIF göstermesinin olası nedenleri ile ilgili detaylı incelemede bulunmamıştır.

Uiterwijk ve Vallen (2005) gerçekleştirdikleri boylamsal çalışmada, işlev farklılığının oluşma sürecinde dile dayalı nedenlerin belirlenmesini amaçlamışlardır. Hollanda'da ilkokul final sınavının (Final Test of Primary Education in Netherlands) 1987, 1989 ve 1993 uygulamalarına ait verileri kullanan araştırmacılar üç basamaklı bir süreç izlemişlerdir: İlk basamağı Mantel-Haeszel ve tek parametrelili IRT yöntemleri ile MİF gösteren maddelerin belirlenmesi, ikinci basamağı, daha önce araştırmacılar tarafından geliştirilen ve göçmen öğrencilerin zorlandıkları madde tiplerini belirlemeyi amaçlayan bir ölçeğin uygulanması, üçüncü basamağı ise MİF gösteren maddelerin yanlış olup olmadığına ve olası yanlışlık kaynaklarına dair uzman grubundan görüş alınması oluşturmaktadır. Araştırmada kullanılan her iki yöntemden birisi ile MİF gösterdiği belirlenen maddelerin, işlev farklılığı gösterdiği kabul edilmiş, yöntemler arası tutarlık aranmamıştır. Referans grup olarak Hollandalı öğrencilerin, odak grup olarak ise ikinci kuşak göçmen öğrencilerin seçildiği çalışmada, üç farklı uygulamada yer alan toplam 540 sorunun 94'ünün (%17.4) en az bir yöntemle göre MİF gösterdiği belirlenmiştir. MİF gösteren tüm maddeler dikkate alındığında, 54 maddenin odak grup aleyhine, 40 tanesinin ise lehine işlev farklılığı gösterdiği belirlenmiştir. İkinci aşamada, araştırmacılar söz konusu 94 maddeyi incelemişler fakat yanlışlık kaynağı görülebilecek ortak bir madde özelliğinin bulunmadığını belirtmişler, MİF'in olası kaynağının maddelerin içerikleri veya referans ve odak gruplar arasındaki sosyoekonomik düzey farkı olabileceğini ifade etmişlerdir.

Wu ve Ercikan (2006), TIMSS 1999 uygulaması matematik testinde yer alan maddelerin kültüre göre yanlılık gösterip göstermediğini ve olası yanlılık kaynaklarını incelemiştir, yanlılık çalışmalarının sıklıkla çeviri hataları ve müfredat farklılıkları ile sınırlı kaldıklarını belirtmişlerdir. Çalışmada TIMSS uygulamalarında tutarlı olarak üst sıralarda yer alan Taiwan ve Amerika Birleşmiş Milletleri'nde yer alan öğrencilerin verileri kullanılmıştır. TIMSS 1999 uygulamasında kullanılan sekiz kitapçığın da incelendiği çalışmada 162 maddeden 34'ünün orta ve yüksek düzeyde MİF gösterdiği belirlenmiştir. Araştırmacılar, Lee (1998) tarafından ifade edilen "TIMSS yalnızca formal eğitime odaklanmakta, okul öğrenmeleri ile bağlantılı diğer öğrenmeleri ve okul dışı faktörleri göz ardı etmektedir" eleştirisine atıf yaparak, Uzakdoğu ülkelerinde eğitim bağlantılı kültüre dayalı bir özellik olan "Okul Sonrası Ek Ders Saatleri" (ELHAS) maddelerin MİF gösterme durumunun olası bir kaynağı olabileceği düşünmüşler, lojistik regresyon sürecinde, öğrencilerin okul süresi dışında kaç saat ELHAS'a katıldıklarını ifade eden ek bir ELHAS değişkeninin denklemde nasıl bir değişikliğe yol açacağını incelemişler ve bu değişkenin Taiwan lehine MIF gösteren maddeleri %30 oranında düşürdüğünü bulmuşlardır. Bu durumu göz önüne alan araştırmacılar, ELHAS'ın Taiwan ve Amerika örneklerinde maddelerin MİF gösterme nedenlerinden biri olarak göstermişlerdir.

Innabi ve Dodeen (2006) tarafından yapılan araştırmada TIMSS 1999 uygulamasında sekizinci sınıf öğrencilerinin (Jordan Bölgesi'nde yaşayan) verileri kullanılarak matematik testinde yer alan maddelerin cinsiyete göre yanlılık gösterip göstermediğinin belirlenmesi amaçlanmıştır. Matematik testinde altı ana konu alanı bulunmaktadır: kesirler ve rakam bilgisi, ölçme, veri kullanımı, analiz ve olasılık, geometri ve cebir. Araştırmada 5299 öğrencinin verileri Mantel-Haenszel (MH) metodu ile incelenmiş ve analiz edilen 124 çoktan seçmeli maddenin 37'sinde, yaklaşık olarak tüm maddelerin üçte birinde, cinsiyete göre yanlı ölçme yaptığı bulunmuştur. Yanlı ölçme yapan maddeler daha yakından incelendiğinde özellikle cebir ve analize yönelik maddelerin kız öğrencilerin lehine çalıştığı ifade edilmiştir. Buna karşılık ölçme konusundaki MİF gösteren tüm maddeler erkek öğrenciler lehine ölçme yapmıştır. Ayrıca, iki sayının toplamının en yakın değeri gibi



yargı ve yordama gerektiren maddeler baskın bir şekilde erkek öğrenciler lehine MİF göstermiştir. Araştırmada, ek olarak, maddelerin içeriğinin MİF göstermesi üzerinde dikkate değer bir etkisi olmadığı fakat “sorulara aşinalığın” MİF için belirleyici bir etken olduğu belirtilmiştir.

Kjærnsli ve Lie (2008) TIMSS 2003 veri seti ile gerçekleştirdikleri araştırmalarında her bir maddenin doğru yanıtlanma yüzdesini kullanarak ve ülkelerin performanslarını göz önüne alarak, ülkeleri benzerlik ve farklılıklarına göre gruplamayı amaçlamışlardır. Üç aşamada gerçekleştirilen araştırmanın ilk aşamasında, fen maddelerinin doğru yanıtlanma yüzdesine göre ülkeler kümeleme analizine tabi tutularak gruplanmış, ikinci basamakta, oluşturulan ülke gruplarının ortak üstünlük ve zayıflıkları incelenmiş ve son olarak üçüncü basamakta, ülkelerin gruplanma şekilleri üzerinde detaylı inceleme yapılmıştır. Sonuç olarak araştırmacılar, kullanılan dilin ülkelerin gruplanmasında oldukça önemli bir faktör olduğunu bulmuşlardır; dolayısı ile aynı dili veya benzer yapıya sahip diller kullanan ülkelerin aynı ülke grubuna girme eğilimi gösterdikleri belirlenmiştir.

Burkes (2009) gerçekleştirdiği çalışmada TIMSS 2003 uygulamasına katılan sekizinci sınıf öğrencilerinin matematik alanındaki başarılarının sosyoekonomik düzeylere göre MİF gösterip göstermediğini incelemiştir ve yanlılığın olası kaynaklarını araştırmıştır. Çalışmada Amerika’da 352 matematik sınıfında eğitim gören 2827 öğrencinin verileri ve *multilevel MİF yöntemi* (ML) kullanılmıştır. Çalışma sonucunda incelenen 71 maddenin 8 tanesinin MİF gösterdiği belirlenmiştir. MİF gösteren 8 maddenin 4’ü sayılar konusunda, 3’ü veri konusunda ve 1’i cebir konusunda bulunmaktadır. Söz konusu yanlı maddelerin hepsinin yüksek sosyoekonomik düzeyde bulunan öğrencilerin lehine ölçme yaptığı bulunmuş ve tartışmalar kısmında vurgulanmıştır. Sosyoekonomik düzeyi yüksek olan ailelerin sıklıkla çocuklarını özel okullara gönderdiğini ve bu okullarda öğrenme fırsatlarının (opportunities to learn [OTL]) daha zengin olduğunu, bu durumun daha önce ifade edilen sekiz maddenin MİF göstermesinin olası bir nedeni olabileceği belirtilmiştir.

Le (2009) PISA 2006 fen bilimleri veri setini kullandığı çalışmasında, test diline yönelik MİF gösteren maddeler ile madde tipleri arasındaki ilişkiyi ve cinsiyete yönelik işlev farklılığı gösteren madde bulunup bulunmadığını incelemiştir. Madde tepki kuramına dayalı Rasch yöntemini kullanan araştırmacı, setini kullandığı analizlerde maddelerin %5.6'sının erkek öğrencilerin ve %2.8'inin kız öğrencilerin lehine ölçme yaptığını bulmuştur.

Qian (2011) yaptığı çalışmada TIMSS 2007 uygulamasında yer alan maddelerin azınlıklara ve cinsiyet gruplarına göre yanlı ölçme yapıp yapmadığını belirlemeyi amaçlamıştır. Araştırmada Kamata (2001) tarafından geliştirilmiş çok düzeyli MİF metodu kullanılmış, performans farklılığının olası nedenleri öğrenci ve öğretmen boyutlarında incelenmiştir. Buna göre fen bilimleri testinde yer alan maddeler analiz edilmiş ve sonuçta özellikle fizik ve yer bilimleri disiplinlerinde yer alan bazı maddelerin erkek öğrencilerin lehine; buna karşılık biyoloji disiplininde yer alan bazı maddelerin ise kız öğrenciler ve azınlık gruplar lehine ölçme yapıldığı bulunmuştur. MIF gösteren maddelerin, madde güçlüğü ile ilişkili olmadığını ifade eden araştırmacı, TIMSS 2007 fen bilimleri testinde yer alan en zor maddenin (Madde 49) MİF göstermemesini bu duruma örnek olarak göstermiştir. Sonuçlara göre, kız öğrencilerin, çoktan seçmeli maddelerde daha düşük başarı gösterdiği belirlenmiş, MİF gösteren sekiz yapılandırılmış açık uçlu maddeler bulunurken, MİF gösteren çoktan seçmeli madde sayısı 26 olarak belirlenmiştir. Çoktan seçmeli maddelerin kız öğrenciler aleyhine ölçme yapma eğiliminin olası bir nedeni olarak, kız öğrencilerin sözel yeteneklerinin daha üst düzeyde olduğunu gösterilmiştir.

Babiar (2011), genel olarak fen alanı, özel olarak mühendislik alanında kadınların yeterince temsil edilmediği ve farklı standart başarı testlerinde kız ve erkek öğrencilerin başarıları arasındaki farka dikkat çekerek 2003 yılında gerçekleştirilen TIMSS uygulaması fen bilimleri testinde cinsiyete göre MİF incelemesi yapmıştır. Facets programı ile gerçekleştirilen analizler neticesinde, maddenin içeriğinin cinsiyete göre MİF üzerinde etkisi olduğu, MİF düzeyi ile madde tipi arasında ilişki olduğu fakat madde ile test edilen bilişsel düzey ile MİF düzeyi arasında bir ilişki olmadığı bulunmuştur.

## Yurtiçinde Gerçekleştirilen Araştırmalar

Yıldırım (2006) yaptığı çalışmada TIMSS ve PISA sınavlarının kültürel eşitliğini (karşılaştırılabilirliğini) sınamak ve kültürel karşılaştırılabilirlik için bozucu bir etken olan olası yanlı maddeleri Maksimum Oran Analizi (IRT-LR) ve Mantel-Haenszel (MH) ile ortaya koymayı amaçlamıştır. Çalışmada TIMSS 1999 uygulaması Türkiye ve Amerika örneklemi kullanılmıştır. Analiz sonuçları TIMSS 1999 uygulaması matematik testinde yer alan, üçü Türk öğrencilerin ve ikisi Amerikan öğrencilerin lehine olmak üzere beş maddenin MİF gösterdiğini belirlemiştir ve hem PISA hem de TIMSS uygulamalarının uyarılma süreçlerinde hatalar yapıldığı ifade edilmiştir.

Kalaycıoğlu ve Berberoğlu (2010), 2007 yılında gerçekleştirilen Üniversiteye Giriş Sınavı'nın cinsiyete göre MİF analizi konulu çalışmalarında, sınavda bulunan matematik ve fen maddelerinin kız ve erkek öğrenci gruplarına yönelik MİF gösteren maddeleri belirlemeyi, söz konusu maddelerin içeriklerini ve işlev farklılığı göstermelerinin olası nedenlerini incelemiştir. MİF analizi sürecinde Mantel-Haenszel, lojistik regresyon ve IRT-LR metotları kullanılmış ve yetenek eşitleme değişkeni olarak öğrencilerin ortaöğretim başarı puanı ortalaması dikkate alınmıştır. Analiz üç basamakta gerçekleşmiştir: 1. Maddeler MİF analizine tabi tutulmuş, 2. MİF gösterdiği belirlenen maddeler içerik analizine tabi tutulmuş ve 3. Madde kökünün yapısı incelenmiştir. Fen Bilimleri-1 testinde bulunan maddelerin %33'ünün MİF gösterdiği belirlenmiştir. Söz konusu 10 maddenin 4'ü erkek öğrenciler lehine işlev farklılığı göstermiştir. MİF gösteren maddeler daha yakından incelendiğinde, şekil veya grafiğe dayanan maddelerin erkek öğrenciler, hız veya sürat konularına dayalı maddelerin kız öğrenciler, algoritmik ve temel hesaplamalardan daha üst düzey bilişsel beceriler gerektiren maddelerin erkek öğrenciler, sözel içerikli maddeler ise kız öğrenciler lehine işlev farklılığı göstermiştir. Sonuçlar göz önüne alındığında, araştırmacılar, maddeler tarafından sınanan bilişsel becerilerin, cinsiyete dayalı MİF'in en önemli etkeni olduğunu belirtmişlerdir.

Ayan (2011), PISA 2009 fen okuryazarlığı uygulamasında cinsiyete yönelik MİF gösteren madde olup olmadığını belirlemeyi amaçlayan

çalışmasında, 12. Kitapçıkta yer alan maddeleri Mantel-Haenszel ve lojistik regresyon yöntemleri ile analiz etmiştir. Mantel-Haenszel ile yapılan analiz neticesinde dört maddenin MİF gösterdiği; söz konusu maddelerin üç tanesinin kız öğrenciler, bir tanesinin erkek öğrenciler lehine orta düzeyde işlev farklılığı gösterdiği belirlenmiştir. Lojistik regresyon yöntemi ise, bir maddenin orta düzeyde ve kızlar lehine MİF gösterdiği belirlemiştir. Buna göre iki yöntemle toplam dört maddede işlev farklılığı olduğu sonucuna varılmıştır. Yöntemlerden elde edilen sonuçların uyumluluğu açısından, her iki yöntemin ortak olarak işlev farklılığı gösterdiğini belirlediği madde bulunamamış ve MİF düzeylerinde düşük düzeyde uyum olduğu görülmüştür. Sonuç olarak her iki yöntemin MİF gösterdiği maddeler açısından uyumsuz sonuçlar ortaya koyduğu ve belirlenen MİF düzeyleri açısından uyumun düşük düzeyde olduğu belirtilmiştir.

Ulutaş (2012) PISA 2006 fen okuryazarlığı testinde yer alan maddelerin Türkiye ve Amerika Birleşik Devletleri kültürlerarası eşdeğerliğini incelemiş ve söz konusu örneklem kapsamında cinsiyete yönelik MİF gösteren maddelerin olup olmadığını belirlemeyi amaçlamıştır. İşlev farklılığı gösteren maddelerin belirlenmesi için klasik test kuramına dayanan Mantel-Haenszel, SIBTEST ve madde tepki kuramına dayanan olabilirlik oran analizini (MTK-OOA) kullanan araştırmacı, incelediği iki kitapçıkta sırası ile 16 ve 24 maddenin işlev farklılığı gösterdiğini belirlemiştir. Söz konusu toplam 40 MİF gösteren maddeden 15'inin Türk öğrenciler lehine, 25'inin Amerikalı öğrenciler lehine işlev farklılığı gösterdiği bulunmuştur. Türkiye örneğinde gerçekleştirilen cinsiyete yönelik MİF analizinde ise, üç yöntemin tutarlı olarak MİF gösterdiğini belirlediği madde bulunamamıştır. MİF gösteren maddelere yönelik uzman görüşleri sonucunda maddelerin içeriğinin ve maddelerin ölçtüğü bilişsel becerilerine göre olası yanlılık kaynağı olarak görülebileceği ifade edilmiştir.

Yurtdışı ve yurtiçinde yapılan çalışmalar bütüncül olarak değerlendirildiğinde, madde işlevinin farklılaşması ve madde yanlılığı çalışmalarının sıklıkla matematik alanında yapıldığı, fen bilimlerine yönelik çalışmaların sınırlı kaldığı görülmektedir. Çalışmaların büyük kısmında MİF analizleri gerçekleştirilmiş fakat olası yanlılık kaynaklarına yönelik uzman

görüŖü alınan araŖtırma sayısının yeterli olmadığı görülmüŖtür. AraŖtırmaların farklı örneklemler üzerinde gerekleŖtirilmesi, farklı sınavlar ve sınavların farklı alt testlerinde analizler yapılması ve MİF belirleme süreçlerinde farklılaŖan yöntemler kullanılması, araŖtırmaların ayrıŖan yönlerini oluŖturmaktadır.

## BÖLÜM 3

### YÖNTEM

Bu bölümde araştırmanın modeline, evren ve örneklemin özelliklerine, veri toplama araç ve yöntemlerine ve verilerin çözümlenmesine ilişkin bilgilere yer verilmiştir.

#### ARAŞTIRMANIN MODELİ

Bu çalışmada, TIMSS 2007 fen bilimleri testinde yer alan toplam 107 maddenin işlevlerinin dil ve cinsiyet değişkenlerine göre farklılaşıp farklılaşmadığının belirlenmesini amaçlamaktadır. Farklı kitapçıklarda yer alan maddelerin işlev farklılığı gösterip göstermediğini ve işlev farklılığı gösteren maddelerin yanlı ölçme yapıp yapmadığını ayrıntılı olarak ortaya koyması açısından betimsel bir çalışmadır.

#### EVREN ve ÖRNEKLEM

Bu çalışma, IEA tarafından gerçekleştirilen TIMSS 2007 çalışmasına Türkiye ve İngiltere’de katılan tüm öğrencilerin verileri kullanılarak gerçekleştirilmiştir. Bu açıdan, çalışmanın evrenini, 2007 yılında Türkiye ve İngiltere’de eğitim gören, en az 13.5 yaşında olan ve formal eğitim sürecinin sekizinci yılında bulunan öğrenciler oluşturmaktadır. Çizelge 4’te araştırma çerçevesinde karşılaştırılan İngiltere-Türkiye evren ve örneklemelerinde yer alan öğrenci sayıları verilmiştir.

#### Çizelge 4

*TIMSS 2007 Sekizinci Sınıf Uygulaması Türkiye ve İngiltere Evren ve Örneklemelerindeki Okul, Öğrenci Sayıları ve Cevaplayıcıların Ortalama Yaşları*

Ülke	Evren		Örneklem		Cevaplayıcıların Ortalama Yaşı
	Okul Sayısı	Öğrenci Sayısı	Okul Sayısı	Öğrenci Sayısı	
İngiltere	3,886	636,732	137	4,025	14.2
Türkiye	16,112	1,163,830	146	4,498	14.0

Çizelge 4 incelendiğinde, İngiltere ve Türkiye evrenlerinde yer alan okul ve öğrenci sayılarının oldukça farklı olduğu, Türkiye evreninde bulunan öğrenci sayısının İngiltere evreninde bulunan öğrenci sayısının yaklaşık iki katı olduğu görülmektedir. Bununla beraber, örneklemedeki okul ve öğrenci sayıları birbirine oldukça yakındır. Türkiye evreninde TIMSS uygulamasına katılan öğrencilerin ortalama yaşı 14, İngiltere evreninde 14.2 olarak belirlenmiştir. Buna göre, evrenin TIMSS 2007'nin örnekleme sürecinde, daha önce gerçekleştirilmiş TIMSS uygulamalarına benzer şekilde iki basamaklı seçkisiz oranlı örnekleme yöntemi tercih edilmiştir. Bu tercihin başlıca sebebi, ölçmeye çalışılan özellik anlamında önemli görülen çeşitli değişkenlerin evrendeki oranı ile örneklemedeki oranının eşit olmasını sağlamak, yani evreni maksimum düzeyde temsil edebilecek bir örnekleme sahip olmaktır (Olson, Martin ve Mullis, 2008).

Örnekleme sürecinin ilk adımında okullar seçkisiz olarak, örneklemedeki oranları evrendeki ile aynı kalacak şekilde seçilmiştir. İkinci aşamada, birinci aşamada belirlenen okullar içinde sınıflar yine ilk aşamada olduğu gibi seçkisiz tercihle belirlenmiştir. Belirlenen sınıflardaki tüm öğrencilere uygulama yapılmıştır. Çizelge 5'te Türkiye için bölgelere göre örnekleme süreci verilmiştir (Olson, Martin ve Mullis, 2008).

#### Çizelge 5

#### *TIMSS 2007 Türkiye Uygulaması İçin Örnekleme Değerlerinin Bölgelere Göre Dağılımı*

Coğrafi Bölge	Örnekleme Planlanan Okul Sayısı	Yeterli Sahip Olmayan Okul Sayısı	Örneklenen Okul Sayısı
Marmara	40	1	39
İç Anadolu	26	0	26
Ege	18	0	18
Akdeniz	18	0	18
Karadeniz	16	0	16
Doğu Anadolu	14	2	12
Güney Doğu Anadolu	18	1	17
Toplam	150	4	146

Çizelge 5 incelendiğinde, Türkiye'nin coğrafi bölgelerinin öğrenci sayıları göz önüne alınarak örneklenmesi planlanan ve örneklenen okul sayısı görülmektedir. Tüm coğrafi bölgelerde toplam 150 okulun örneklenmesi planlanmış, uygun niteliklere sahip 146 okul örneklenmiştir. Örneklenmesi planlanan fakat yeterliğe sahip olmayan okulların özellikle, ulaşılması zor veya öğrenci sayısının oldukça az olduğu okullar olduğu belirtilmiştir. Örneklenen okullardan bir sınıf seçkisiz olarak belirlenmiş ve bu sınıfta bulunan tüm öğrencilere uygulama yapılmıştır (Olson, Martin ve Mullis, 2008).

### **TIMSS UYGULAMASINDA KULLANILAN ÖLÇME ARAÇLARI**

TIMSS dünya çapında gerçekleştirilen ve katılımcı ülkelerin dördüncü ve sekizinci sınıf matematik ve fen bilimleri başarısını ölçmeyi amaçlayan, dört yıllık aralıklarla uygulanan uluslararası bir sınavdır. TIMSS ile farklı eğitim sistemlerine, okul organizasyon yaklaşımlarına ve eğitsel uygulamalara sahip ülkelerin başarılarını karşılaştırma amacıyla zengin bir altyapı sağlayacak şekilde veri toplanması mümkün hale gelmiştir.

TIMSS 2007 matematik ve fen bilimleri testlerinde yer alan maddelerin geliştirilmesi yaklaşık iki yıllık bir çalışmanın (Ocak 2005 - Kasım 2006) sonucudur. Bu sürece test geliştiricileri ile beraber, tüm dünyadan alanı matematik ve fen bilimleri olan birçok eğitimci katkı sağlamıştır. Maddelerin geliştirilmesi sürecinin ilk adımında TIMSS 2003 uygulamasında yer alan maddeler güncellenmiş ve revize edilmiştir. Bu maddeler ile beraber TIMSS 2007 için, Ulusal Araştırma Koordinatörlüğü (NRC) ile birlikte ve Fen Bilimleri – Matematik Madde İnceleme Komitesi (SMIRC) denetiminde yeni maddeler geliştirilmiştir. 60 katılımcı ülkeden gelen uzmanlar ile beraber Mayıs 2005'te TIMSS 2007 Madde Geliştirme Çalıştayı gerçekleştirilmiştir. Bu çalıştayda geliştirilen 472'si matematik ve 525'i fen bilimlerinde olmak üzere toplam 997 madde SMIRC tarafından incelenmiş ve maddeler ön uygulamaya tabi tutulmuştur. Madde seçiminde aşağıdaki değişkenler dikkate alınmıştır (Olson, Martin ve Mullis, 2008):

- Her madde için güçlük düzeyi
- Maddelerin düşük ve yüksek performans gösteren öğrencileri hangi düzeyde ayırt edebildiği



- Çoktan seçmeli maddelerde yer alan çeldiricilerin etkililiği
- Yapılandırılmış maddelerde puanlama güvenilirliği

TIMSS 2007 uygulamasında, daha önceki TIMSS uygulamalarına benzer şekilde çoktan seçmeli ve yapılandırılmış maddeler kullanılmıştır. Matematik ve fen bilimleri testinde yer alan tüm çoktan seçmeli maddeler doğru yanıtlandığında bir puan, yanlış yanıtlandığında puan alınmayacak şekilde yapılandırılmıştır. Yapılandırılmış her bir madde için ise dereceli puanlama anahtarı hazırlanmış, doğru cevaplanması durumunda (doğru veya kısmen doğru) alınabilecek puan, madde ile ölçülecek özelliğin bilişsel düzeyine göre bir puan ile altı puan arasında değişmektedir (Mullis, Martin ve Ruddock, 2005).

TIMSS 2007 uygulaması sekizinci sınıf düzeyinde, matematik testinde 215, fen bilimleri testinde 214 olmak üzere, toplam 429 maddeden oluşur. Bu maddeler 14 blok altında toplanmıştır, oluşturulan 14 blok kullanılarak 14 farklı kitapçık hazırlanmıştır. Bu kitapçıkların eşdeğer olması adına oldukça dikkatli bir rotasyon süreci izlenmiştir ve her bir cevaplayıcı, söz konusu 14 kitapçıktan yalnızca birini cevaplamıştır. Çizelge 6'da fen bilimleri testinde yer alan maddelerin temsil ettiği konu alanları gösterilmiştir.

#### Çizelge 6

##### *TIMSS 2007 Fen Bilimleri Sekizinci Sınıf Testinde Yer Alan Maddelerin Konu Alanları*

Fen Bilimleri Disiplini	Konu Alanı
Biyoloji	<ol style="list-style-type: none"> <li>1. Organizmaların özellikleri, sınıflaması ve yaşam süreçleri</li> <li>2. Hücreler ve işlevleri</li> <li>3. Yaşam döngüleri, üreme ve kalıtım</li> <li>4. Çeşitlilik, adaptasyon ve doğal seçim</li> <li>5. Ekosistemler</li> <li>6. İnsan sağlığı</li> </ol>
Kimya	<ol style="list-style-type: none"> <li>1. Maddenin yapısı ve sınıflaması</li> <li>2. Maddenin özellikleri</li> <li>3. Kimyasal değişim</li> </ol>

---

Fizik	<ol style="list-style-type: none"><li>1. Maddedeki fiziksel deęişim ve özellikler</li><li>2. Enerji aktarımı, ısı ve sıcaklık</li><li>3. Işık</li><li>4. Ses</li><li>5. Elektrik ve Manyetizma</li><li>6. Kuvvet ve hareket</li></ol>
Yer Bilimleri	<ol style="list-style-type: none"><li>1. Yeryüzünün yapısı ve fiziksel özellikleri</li><li>2. Yeryüzünün tarihi, süreçleri ve döngüleri</li><li>3. Yeryüzünün kaynakları, kaynakların kullanımı ve korunması</li><li>4. Güneş sisteminde yeryüzü ve evren</li></ol>

---

TIMSS 2007 uygulamasında hem matematik hem de fen bilimleri testlerinde konuların ve konuların içeriğinin açıklanması dışında, temsil edilen bilişsel alanın ayrıntılı olarak ifade edilmesi ve fen bilimleri disiplinlerine göre söz konusu bilişsel alanların temsil edilme düzeylerinin belirlenmesi önemli görülmektedir.

### **TIMSS 2007 Fen Bilimleri Testinde Ölçülen Bilişsel Beceriler**

TIMSS 2007 fen bilimleri testinde bilme, uygulama ve akıl yürütme bilişsel alanlarından oluşan üç basamaklı taksonomi kullanılmaktadır. Aşağıda söz konusu bilişsel basamakların temel yapıları açıklanmıştır.

#### **Bilme**

Bilme, öğrencinin bilimsel gerçeklere, kavramlara, araçlara ve işlemlere yönelik bilgisini ifade etmektedir. Öğrenci bu bilişsel alanda bilgiyi hatırlar, tanımlar, betimler, örneklerle ifade eder, uygun araç ve yöntemleri kullanır.

#### **Uygulama**

Uygulama alanı, sahip olunan bilginin uygun bir durumda kullanılması durumlarını kapsamaktadır. Bu bilişsel alanda öğrenciden durumlar arasındaki benzerlik ve farklılıkları belirlemesi, sınıflama yapması, uygun model veya modelleri kullanması, ilişkilendirmesi, bilgiyi yorumlaması, çözüm bulması ve açıklama yapması beklenir.

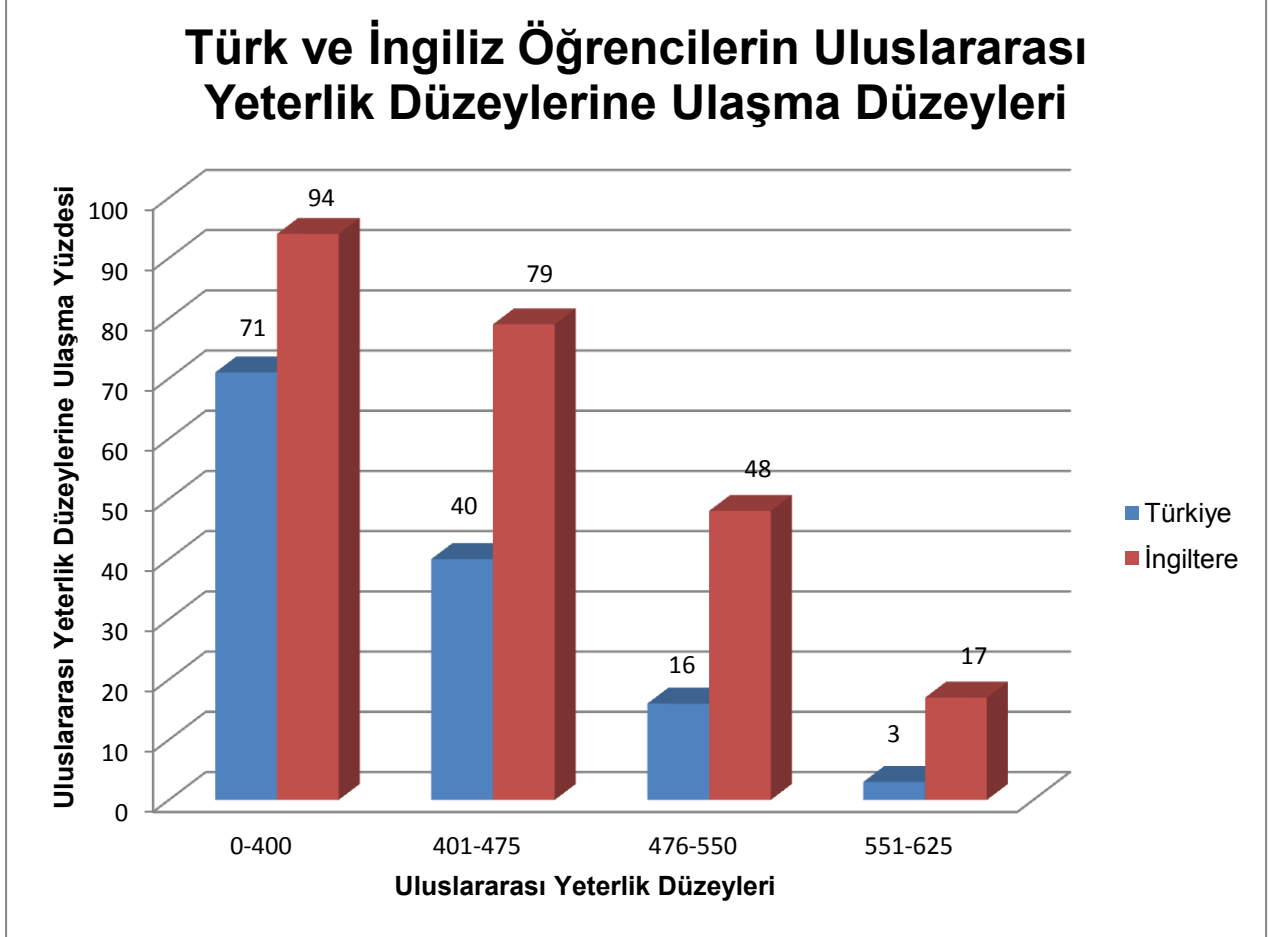
## Akıl Yürütme

Akıl yürütme alanında ise öğrenci birden çok durum veya karmaşık tek bir durum ile karşı karşıyadır. Öğrencilerden bu bilişsel alanda problemleri analiz etmesi ve çözmesi, bilgileri bütünleştirmesi ve sentez etmesi, hipotez kurması ve tahmin etmesi, plan yapması, sonuçlar çıkarması, genellemeler yapması, değerlendirmesi ve bir yargıyı doğrulaması veya yanlışlaması beklenmektedir (Mullis, Martin ve Ruddock, 2005).

TIMSS uygulamalarında kullanılan dört uluslararası yeterlik düzeyi bulunmaktadır, bu düzeyler aşağıda kısaca açıklanmıştır (Martin, Mullis ve Foy, 2008):

- *Düşük Uluslararası Yeterlik Düzeyi (400 puan):* Öğrenciler yaşama dair bilimlerde ve fizikte ancak bazı gerçeklerin farkındadır. Bu düzeyde bulunan öğrenciler, resimler ve diyagramları yorumlar, temel tabloları tamamlar, olaylarla ilgili temel sorulara kısa yanıtlar verir.
- *Orta Uluslararası Yeterlik Düzeyi (475 Puan):* Öğrenciler bir konu dizisi içeriğindeki temel bilgilerin farkındadır ve içerik ile konu arasındaki etkileşimi kurar. Bu düzeyde bulunan öğrenciler, çizelgeleri bilgileri yorumlar ve kuramsal bilgisini betimsel cevaplar vermek için işe koşabilir.
- *Üst Uluslararası Yeterlik Düzeyi (550 Puan):* Öğrenciler bilimsel sistem, dönüşüm ve prensiplere dair kavramsal bilgi ve kavrayışlarını gösterir. Bu düzeyde bulunan öğrenciler sonuçlara ulaşmak için farklı bilgi bütünlerini birleştirir, görsel bilgileri yorumlar ve bilgilerini kullanarak kısa açıklamalarda bulunurlar.
- *İleri Uluslararası Yeterlik Düzeyi (625 Puan):* Öğrenciler fizik, biyoloji, kimya ve yer bilimlerine yönelik karmaşık durumları kavrar ve görüşleri özetler. Bu düzeyde bulunan öğrenciler, bilimsel araştırmaların temellerini anlar ve bazı fiziksel kuralları nicel soruları çözmek için kullanırlar.

Yeterlik düzeylerinin açıklanmasının ardından Türk öğrencilerin sekizinci sınıf düzeyinde uluslararası yeterlik düzeylerine erişme yüzdeleri Şekil 5'te gösterilmiştir.



Şekil 5: Türk ve İngiliz Öğrencilerin Uluslararası Yeterlik Düzeylerine Ulaşma Düzeyleri

Şekil 5 incelendiğinde, Türkiye örneğinde öğrencilerin büyük kısmının düşük uluslararası yeterlik düzeyine ulaştığı, ileri uluslararası yeterlik düzeyine ulaşan öğrencilerin oranı oldukça düşüktür. Orta uluslararası yeterlik düzeylerine ulaşan öğrenciler, tüm öğrencilerin %40'sını oluşturmaktadır. Türkiye örneğindeki öğrencilerin %29'u düşük uluslararası yeterlik düzeyine ulaşamamıştır.

İngiltere örneğinde ileri uluslararası yeterlik düzeyine erişen öğrencilerin yüzdesi, Türkiye'de aynı düzeye erişen öğrencilerin yaklaşık altı katıdır. Orta düzey uluslararası düzeye erişen öğrencilerin oranı %79'dur.

Türkiye örnekleminde bu oran %49'da kalmıştır. Diğer taraftan İngiltere örnekleminde düşük uluslararası yeterlik düzeyine ulaşamayan öğrencilerin oranı %6'dır. Bu oran Türkiye örnekleminde %29'a ulaşmıştır. Tüm uluslararası yeterlik düzeyleri dikkate alındığında Türkiye ve İngiltere örneklemelerinde fen bilimleri başarısı anlamında dikkate değer bir farklılık görülmektedir.

Araştırma çerçevesinde, TIMSS 2007 uygulamasında IEA tarafından geliştirilen ve fen bilimleri testinde yer alan 107 çoktan seçmeli madde incelenmiştir. Çizelge 7'de araştırmacıların erişimine açılmış 45 çoktan seçmeli maddenin fen bilimleri disiplinlerine göre ayrımı gösterilmektedir.

#### Çizelge 7

*TIMSS 2007 Sekizinci Sınıf Uygulaması Erişime Açılan Çoktan Seçmeli Fen Bilimleri Sorularının Fen Bilimleri Disiplinlerine Göre Dağılımı*

Disiplin	Biyoloji	Fizik	Kimya	Yer Bilimleri
Madde Sayısı	16	14	8	7

TIMSS 2007 uygulaması daha önce açıklanan matematik ve fen bilimleri testi yanında, öğrenci ve genel eğitim değişkenlerini belirlemek adına dört farklı anket içermektedir. Genel olarak öğrenciler bu anketlerde yer alan maddelere, okulda ve evde sahip oldukları çevreye ve imkânlara göre yanıt vermektedirler.

### **VERİLERİN ELDE EDİLMESİ**

Araştırmada kullanılan veriler, TIMSS ve PIRLS uygulamaları için IEA resmi web sitesinde yer alan "Uluslararası Veri Bankası" sekmesi aracılığıyla ([http://timssandpirls.bc.edu/TIMSS2007/idb\\_uq.html](http://timssandpirls.bc.edu/TIMSS2007/idb_uq.html)) sağlanmıştır.

### **VERİLERİN ANALİZİ**

TIMSS 2007 verileri elde edildiği ilk haliyle, dördüncü ve sekizinci sınıflar için fen bilimleri ve matematik testlerine, öğrenci anketlerine, öğretmen anketlerine ve yönetici anketlerine verilen cevapları içeren setler halinde hazırlanmıştır. Söz konusu değişkenlere göre hazırlanan veri setleri, katılımcı

her ÷lkeye üç harfli bir kod ve üç haneli bir numara (Türkiye için ÷lke kodu: TUR, ÷lke numarası: 792) verilerek alt gruplara ayrılmıştır.

Fen bilimleri testine yönelik veri setinde, testte yer alan tüm maddelere verilen yanıtlar ile beraber her cevaplayıcının cinsiyeti, kişisel ve eğitime dair demografik özellikleri yer almaktadır. Araştırmanın amaçları doğrultusunda Türkiye ve İngiltere örneklemelerine ait veri setleri, ÷lke kodları korunarak birleştirilmiş ve alt amaçlarda gerçekleştirileceđi belirtilen analizler için uygun bir veri seti geliştirilmiştir.

Her bir alt amaç doğrultusunda gerçekleştirilecek analiz ve bu analizlerin basamaklarına geçmeden önce, uluslararası veri tabanından indirilen orijinal dosyada yer alan verilerin tüm analizler için hazır hale getirilmesi gerekmektedir. Bu amaçla yapılanlar aşamalı olarak ifade edilmiştir:

1. IEA tarafından yayımlanan orijinal veri setinde veri dosyaları ÷lkeler ve sınıflara göre isimlendirilerek ayrılmıştır. Birinci amaç için Türkiye ve İngiltere 8. sınıf örneklemeleri birleştirilerek amaca özel bir veri dosyası oluşturulmuştur.
2. Türkiye ve İngiltere 8. sınıf örneklemelerinin orijinal veri dosyalarında bulunan matematik maddeleri ve araştırmanın amacı ile ilgisiz olan diđer deđişkenler çıkarılmış, yalnızca incelenecek fen bilimleri maddeleri ve gerekli demografik deđişkenler bulunacak şekilde, veri dosyası sadeleştirilmiştir.
3. Sadeleşen dosyada her öğrenciye ait olan beş olası fen bilimleri puanının ortancasını belirleyen yeni bir deđişken tanımlanmış ve bu deđişken öğrencinin “fen bilimleri başarı puanı” olarak dikkate alınmıştır.
4. Yetenek gruplarını oluşturmak için Türkiye ve İngiltere örneklemelerine ait fen bilimleri başarı puanı dokuz eşit düzeye ayrılarak odak ve referans gruplarında bulunan öğrencilerin hangi yetenek düzeyinde yer alacakları belirlenmiştir. Bunun için, birinci basamakta oluşturulan veri setinde en yüksek ve en düşük fen bilimleri başarı puanları hesaplanmış (en düşük fen bilimleri puanı:

179.46, en yüksek fen bilimleri puanı: 780.08), söz konusu puanların ranjı hesaplanmış (Ranj=600.62) ve ranj dokuz eşit aralıklı yetenek düzeyine ayrılmıştır. Buna göre fen bilimleri puanları aralıkları ve bu aralıklara ait yetenek düzeyleri Çizelge 10'da yer almaktadır.

Çizelge 8

*Fen Bilimleri Başarı Puanlarına Göre Oluşturulmuş Eşit Aralıklı Yetenek Düzeyleri*

Fen Bilimleri Puan Aralığı	Yetenek Düzeyi
179.46 – 246.20	1
246.21 – 312.94	2
312.95 – 379.68	3
379.69 – 446.42	4
446.43 – 513.16	5
513.17 – 579.90	6
579.91 – 646.64	7
646.65 – 713.38	8
713.39 – 780.08	9

- 14 farklı kitapçık için birer analiz klasörü oluşturulmuş ve her bir kitapçıkta maddelere verilen yanıtlar iki uçlu (0 ve 1) şekilde kodlanarak veriler MİF analizlerine hazır hale getirilmiştir.
- Her bir kitapçık için Türkiye ve İngiltere örneklemi göz önüne alınarak çoktan seçmeli maddelerin işlevlerinin kullanılan dile göre farklılaşıp farklılaşmadığını belirlemek amacıyla Mantel-Haenszel, SIBTEST ve lojistik regresyon analizleri yapılmıştır.
- Bir önceki adımda, farklı yöntemlerle dile göre işlevinin farklılaştığı belirlenen maddelerin tutarlı olup olmadığını incelenmiştir.
- Her bir kitapçık için Türkiye örnekleme göz önüne alınarak çoktan seçmeli maddelerin işlevlerinin kullanılan kız ve erkek yanıtlayıcılara göre farklılaşıp farklılaşmadığını belirlemek amacıyla Mantel-Haenszel, SIBTEST ve lojistik regresyon analizleri yapılmıştır.

9. Bir önceki adımda, farklı yöntemlerle cinsiyete göre işlevinin farklılaştığı belirlenen maddelerin tutarlı olup olmadığı incelenmiştir.
10. Araştırmacıların erişimine açılan çoktan seçmeli maddelerin dil ya da cinsiyete göre işlev farklılığı gösterdiğinin belirlenmesi halinde, söz konusu maddelerin olası yanlışlık nedenlerinin belirlenmesi amacıyla “uzman grubunun” görüşleri alınmıştır. Alan uzmanları grubu, Ankara’da bulunan bir genel lise ve bir Anadolu Lisesi’nde ve görevine devam eden, en az beş yıldır görev yapan iki biyoloji, üç fizik, üç biyoloji ve iki coğrafya öğretmeninden oluşmuştur. Yabancı dil uzmanları, tercüme ve uyarlama konusunda deneyimli, bir vakıf üniversitesi ve bir genel lisede görev yapan bir İngilizce öğretmeni olarak belirlenmiştir.

Araştırmada TIMSS 2007 uygulaması fen bilimleri testinde yer alan çoktan seçmeli maddelerin dil ve cinsiyete göre yanlışlık gösterip göstermediğinin belirlenmesi amacıyla Mantel-Haenszel, SIBTEST ve lojistik regresyon analizleri yapılmıştır. Mantel-Haenszel analizi DIFAS 5.0, SIBTEST analizi SIBTEST DIFPack 1.7 ve lojistik regresyon analizi SPSS 13.0 paket programları ile gerçekleştirilmiştir.



## BÖLÜM 4

### BULGULAR VE YORUMLAR

Bu bölümde, araştırmanın amaçları çerçevesinde elde edilen bulgular ve yorumlar sırası ile özetlenmektedir.

#### 1. TIMSS 2007 uygulaması Türkiye ve İngiltere örneklemi kapsamında dile yönelik MİF gösteren maddelerin belirlenmesi

Türkiye ve İngiltere örneklemi kapsamında dile yönelik MİF gösteren maddelerin belirlenmesi sürecinde, birleştirilen veri setinde, sırası ile Mantel-Haenszel, SIBTEST ve lojistik regresyon analizleri gerçekleştirilmiş, analiz sonuçlarında elde edilen sırası ile  $\alpha_i$ ,  $\beta_{UNI}$  ve  $\Delta R$  katsayıları hesaplanmış ve bu katsayıların yönü ve nicelikleri göz önüne alınarak kitapçıklarda yer alan maddelerin hangi düzeyde MİF gösterdiği belirlenmiştir.

Üç yöntemin tutarlı olarak en az orta düzeyde (B düzeyi) MİF gösterdiğini belirlediği maddeler tek biçimli MİF, yalnızca SIBTEST ve lojistik regresyon analizlerinin birlikte en az orta düzeyde MİF gösterdiğini belirlediği maddeler ise tek biçimli olmayan MİF olarak ifade edilmiştir. Çizelge 9'da Türkiye ve İngiltere örneklemi kapsamında fen bilimleri testinde farklı kitapçıklardaki MİF gösteren maddeler, MİF türü ve söz konusu maddelerin lehine işlev farklılığı gösterdiği grup verilmiştir.

Çizelge 9

*TIMSS 2007 Uygulamasında Kullanılan Kitapçıkları Cevaplayan Öğrencilerin Dağılımı ve Her Bir Kitapçıkta Yer Alan Maddelerin Dile Göre MİF Gösterme Durumu*

Kitapçık	MİF Gösteren Madde	MH	SIBTEST	LR	MİF Türü	Lehine MİF Gösterilen Grup
1	*S032115	C	C	B	Tek Biçimli MİF	Türk Öğrenciler
	*S042028	C	C	C	Tek Biçimli MİF	Türk Öğrenciler
	*S042071	C	C	C	Tek Biçimli MİF	Türk Öğrenciler
2	*S042028	C	C	C	Tek Biçimli MİF	Türk Öğrenciler
	*S042071	C	C	B	Tek Biçimli MİF	Türk Öğrenciler
3	-	-	-	-	-	-
4	*S032385	C	C	B	Tek Biçimli MİF	Türk Öğrenciler
5	*S032385	C	C	C	Tek Biçimli MİF	Türk Öğrenciler
	S042038	C	C	B	Tek Biçimli MİF	İngiliz Öğrenciler
6	*S032257	C	C	B	Tek Biçimli MİF	İngiliz Öğrenciler
7	*S032257	C	C	B	Tek Biçimli MİF	İngiliz Öğrenciler
8	-	-	-	-	-	-
9	S042197	C	C	B	Tek Biçimli MİF	İngiliz Öğrenciler
	*S042274	C	C	B	Tek Biçimli MİF	Türk Öğrenciler
10	*S042274	C	C	B	Tek Biçimli MİF	Türk Öğrenciler
11	-	-	-	-	-	-
12	-	-	-	-	-	-
13	*S042016	C	C	C	Tek Biçimli MİF	Türk Öğrenciler
	*S042216	C	C	B	Tek Biçimli MİF	İngiliz Öğrenciler
14	*S032115	C	C	B	Tek Biçimli MİF	Türk Öğrenciler
	S042258	C	C	C	Tek Biçimli MİF	Türk Öğrenciler
	*S042016	C	C	C	Tek Biçimli MİF	Türk Öğrenciler
	*S042216	C	C	B	Tek Biçimli MİF	İngiliz Öğrenciler

*Birden fazla kitapçıkta yer alan ortak ve MİF gösteren maddeler \* ile gösterilmiştir.*

Çizelge 11 incelendiğinde, Mantel-Haenszel, SIBTEST ve lojistik regresyon analizleri tutarlı olarak, birinci kitapçıkta yer alan S032115, S042028 ve S042071 maddelerinin Türk öğrenciler lehine işlev farklılığı

gösterdiği görülmüştür. İkinci kitapçıkta yer alan maddelerden *S042028* ve *S042071* maddeleri, birinci kitapçıkta MİF gösteren maddelere benzer şekilde, Türk öğrenciler lehine işlev farklılığı göstermiştir. İlk iki kitapçığın aksine, üçüncü, sekizinci, on birinci ve on ikinci kitapçıklarda Türk veya İngiliz öğrencilerin lehine MİF gösteren madde bulunmamaktadır. Dördüncü kitapçıkta *S032385* maddesinin Türk öğrenciler lehine MİF gösterdiği belirlenmiştir. Beşinci kitapçıkta *S032385* ve *S042038* maddeleri sırası ile Türk ve İngiliz öğrenciler lehine MİF göstermiştir. Altıncı ve yedinci kitapçıklarda ortak olarak yer alan *S032257* maddesi her iki kitapçıkta da İngiliz öğrenciler lehine işlev farklılaşması göstermiştir. Dokuzuncu kitapçıkta *S042197* maddesi İngiliz öğrenciler lehine MİF gösterirken, *S042274* maddesi dokuzuncu ve onuncu kitapçıklarda Türk öğrenciler lehine işlev farklılığı göstermiştir. On üçüncü kitapçıkta *S042016* ve *S042216* maddeleri sırası ile Türk ve İngiliz öğrencilerin lehine MİF göstermiştir. On dördüncü kitapçıkta, on üçüncü kitapçığa benzer şekilde, *S042016* ve *S042216* maddeleri sırası ile Türk ve İngiliz öğrencilerin lehine, *S032115* ve *S042258* maddelerinin ise Türk öğrenciler lehine işlev farklılığı gösterdiği belirlenmiştir.

Alanyazında incelenen çalışmalar dikkate alındığında, Wu ve Ercikan (2006), TIMSS 1999 matematik maddelerinin kültüre dayalı MİF gösterdiğini bulmuştur. Yıldırım (2006) TIMSS 2009 uygulamasında Türkiye ve Amerika örneklemelerini karşılaştırarak puanların karşılaştırılabilirliğini tehdit eden, işlev farklılığı gösteren maddelerin bulunduğunu belirtmiştir. Uiterwijk ve Vallen (2005), Hollanda'da uygulanan standart testte, test diline yönelik MİF gösteren maddelerin bulunduğunu bulmuştur. Klieme ve Bauert (2001), TIMSS 1995 uygulamasında farklı eğitim geleneklerinin ve kültürel özelliklerin işlev farklılaşmasının nedenlerinden olduğunu belirtmiştir. Dilin kültürel farklılıkların en önemli unsuru olduğu ve farklılaşmanın temel kaynaklarından biri olduğu göz önüne alındığında bu çalışma, kültüre dayalı ve test diline dayalı MİF gösteren maddelerin bulunduğu söz konusu çalışmaların bulguları ile uyum göstermektedir. Ayrıca Kjaernsli ve Lie (2008), TIMSS 2003 fen bilimleri testinde yer alan maddeler üzerinde MİF gösterme durumunun, kullanılan dille ilişkili olduğunu göstermiştir. Bu durumda farklı diller kullanan ve benzer yetenekte olan öğrencilerin, maddelere doğru cevap verme olasılığının

farklılaşabileceğini göstermekte ve bu bulgu da araştırmada elde edilen bulgular ile uyum göstermektedir.

## 2. TIMSS 2007 uygulaması Türkiye örnekleme kapsamında *cinsiyete* yönelik MİF gösteren maddelerin belirlenmesi

Türkiye örnekleme kapsamında cinsiyete yönelik MİF gösteren maddelerin belirlenmesi süresinde sırası ile Mantel-Haenszel, SIBTEST ve lojistik regresyon analizleri gerçekleştirilmiş, analiz sonuçlarında elde edilen sırası ile  $\alpha_i$ ,  $\beta_{UNI}$  ve  $\Delta R$  katsayıları hesaplanmış ve bu katsayıların yönü ve nicelikleri göz önüne alınarak kitapçıklarda yer alan maddelerin hangi düzeyde MİF gösterdiği belirlenmiştir.

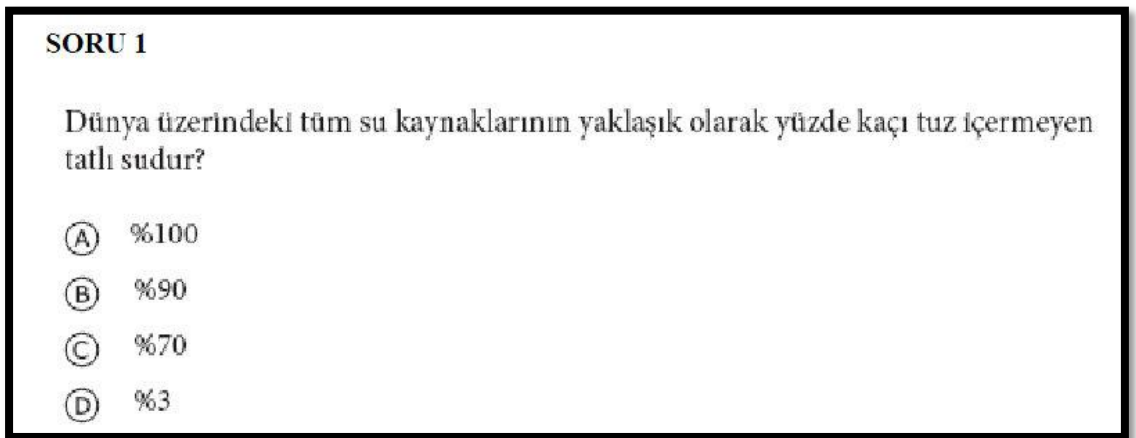
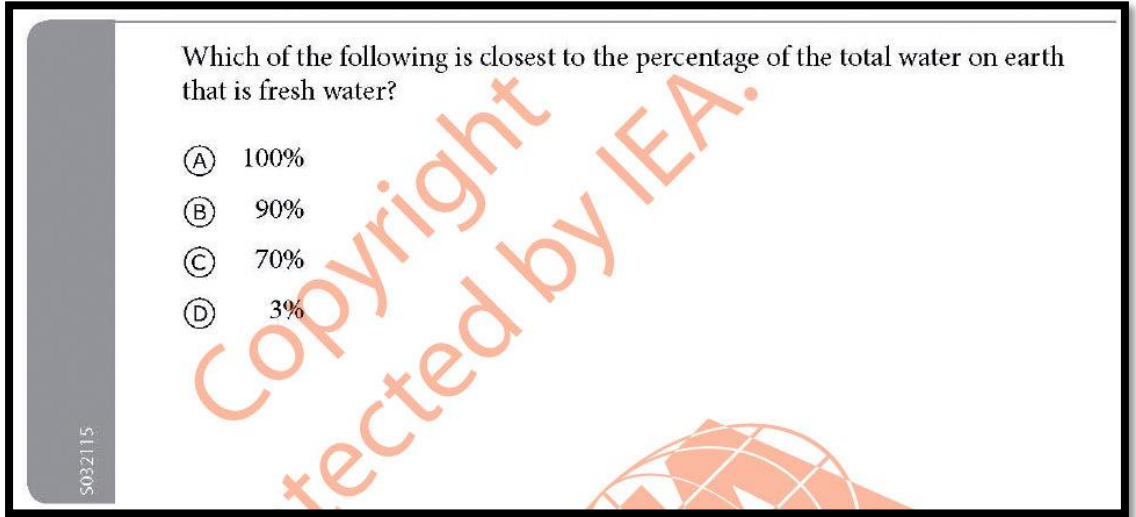
Birinci amacın bulgularının aksine, yapılan analizler neticesinde Türkiye örnekleme kapsamında, üç yöntemin tutarlı olarak tek biçimli MİF gösterdiğini belirlediği veya SIBTEST ve lojistik regresyon yöntemlerinin tutarlı olarak tek biçimli olmayan MİF gösterdiğini belirlediği bir madde bulunmamıştır.

Araştırma kapsamında elde edilen bulgunun Innabi ve Dodeen (2006), Quian (2011) ve Babier (2003) tarafından TIMSS'in farklı yıllardaki uygulama sonuçlarına yönelik elde ettikleri bulgular ile farklılaştığı, Ulutaş (2012)'in çalışmasında elde ettiği bulgular ile uyum gösterdiği belirlenmiştir. Bulgular anlamında farklılaşan çalışmalarda cinsiyete yönelik MİF gösteren maddeler bulunmuş, bu maddelerin büyük oranda erkek öğrenciler lehine işlev farklılığı gösterdiği ifade edilmiştir. Bu durumun, MİF belirleme sürecinde kullanılan yöntemlerin farklı olmasından, araştırmada tek biçimli işlev farklılaşması için üç, tek biçimli olmayan işlev farklılaşması için iki yöntemin tutarlılığının göz önüne alınması ve karşılaştırılan örneklemelerin farklılaşmasından kaynaklanabileceği düşünülmektedir.

## 3. Fen Bilimleri testinde *dile* göre işlev farklılığı gösterdiği belirlenen, araştırmacıların erişimine açılmış maddelerin olası yanlışlık kaynağı konusundaki uzman görüşleri

Mantel-Haenszel, SIBTEST ve lojistik regresyon yöntemleri ile yapılan analizler neticesinde MİF gösteren maddelerin olası yanlışlık kaynakları

konusunda uzman görüşleri alınması sürecinde, öncelikle MİF gösteren maddelerin hangilerinin arařtırmacıların erişimine açılmış olduđu belirlenmiştir. Buna göre S032115, S032257, S032385, S042028 ve S042071 olmak üzere toplam beş maddenin arařtırmacıların erişimine açılmış MİF gösteren maddeler olduđu bulunmuştur. Bu maddelerin her biri, alanında en az beş yıldır çalışan fizik, biyoloji, kimya, coğrafya ve fen-teknoloji öğretmenlerinden oluşan alan uzmanları, tercüme konusunda deneyimli yabancı dil uzmanlarından oluşan “uzman grubu”nun görüşlerine sunulmuştur. Alınan dönütler çerçevesinde bütüncül değerlendirmeler yapılmış ve söz konusu maddelere dair uzman görüşleri sırası ile sunulmuştur. Şekil 6’da S032115 maddesinin sırası ile İngilizce ve Türkçe kitapçıklardaki şekli görülmektedir.



Şekil 6: S032115 Maddesinin İngilizce ve Türkçe Kitapçıklardaki Şekli

Şekill 6’da görülen S032115 maddesinin analizler sonucunda, Türk öğrenciler lehine tek biçimli MİF gösterdiği belirlenmiştir. Coğrafya alan uzmanlarının görüşleri doğrultusunda, maddenin ve cevabın bilimsel olarak doğru olduğu fakat madde kökünde sorulan özelliğin daha iyi ifade edilebileceği belirtilmiştir. İngilizce alanında uzman ve tercüme konusunda deneyimli uzmanlar, madde kökünde sorulan özelliğin Türkçe’ye daha uyumlu bir şekilde çevrilebileceğini, orijinal kitapçıkta yer alan “fresh water” teriminin “tuz içermeyen tatlı su” yerine “içilebilir su” olarak çevrilmesinin eşdeğerlik açısından daha doğru olacağını belirtmişlerdir. Uzman grubu ile yapılan görüşmelerde, dilbilimsel eşdeğerliğin sağlanmamasının bozucu bir etki yarattığı ifade edilmiş fakat bu etkinin, Türk öğrencilerin lehine bir etki yarattığına dair açık bir kanıt olmadığını belirtmişlerdir. Bu anlamda, S032115 maddesinin yanlış ölçme yapmadığı sonucuna varılmıştır. Şekil 7’de S032257 maddesinin İngilizce ve Türkçe kitapçıklardaki şekli verilmiştir.

A sound is heard when you pluck a string on a guitar.  
What will happen to the sound if the same string is plucked harder?

(A) The volume will stay the same, and the pitch will be higher.  
(B) The pitch will stay the same, and the volume will be higher.  
(C) Both the pitch and the volume will be higher.  
(D) Both the pitch and the volume will stay the same.

Copyright ©  
Protected by IFLA

S032257

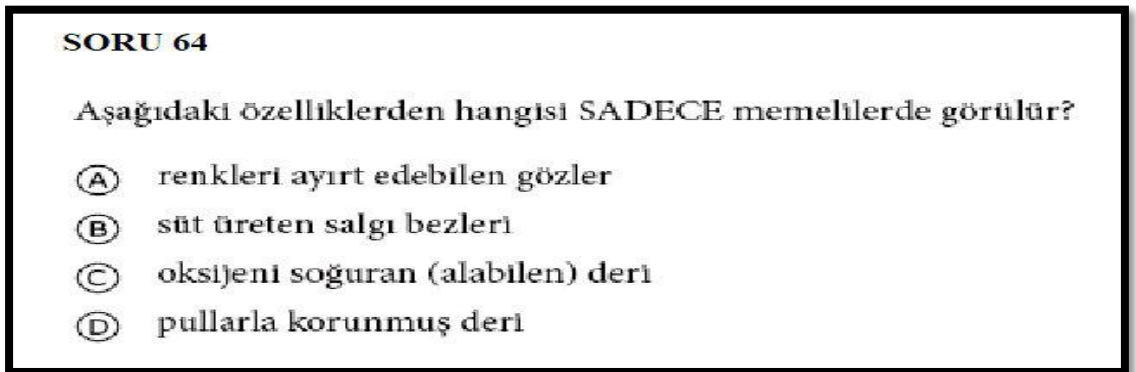
**SORU 77**

Gıtaradaki bir teli çekip bıraktığınızda bir ses duyulur. Eğer aynı tel daha sert çekilip bırakılırsa, duyulan seste nasıl bir değişiklik olur?

(A) Ses yüksekliği artar ve ses şiddeti aynı kalır.  
(B) Ses yüksekliği aynı kalır ve ses şiddeti artar.  
(C) Ses yüksekliği ve ses şiddetinin ikisi de artar.  
(D) Ses yüksekliği ve ses şiddetinin ikisi de aynı kalır.

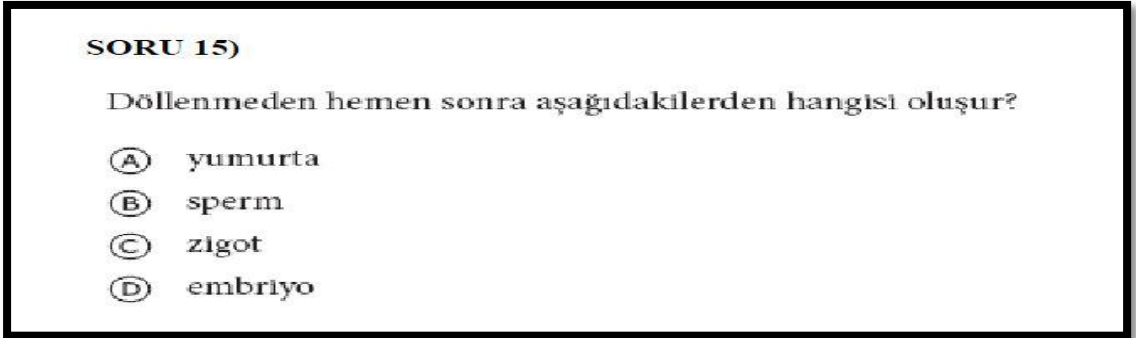
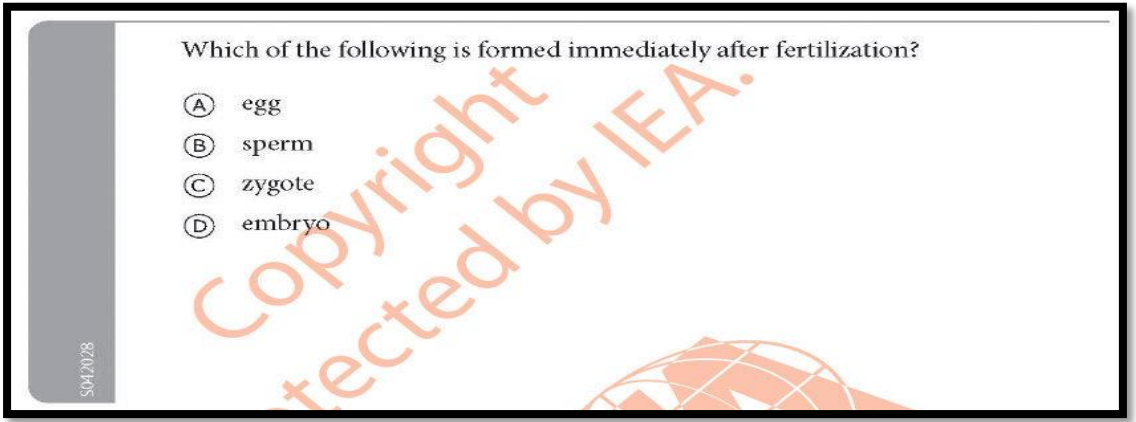
Şekil 7: S032257 Maddesinin İngilizce ve Türkçe Kitapçıklarındaki Şekli

Şekil 7’de incelenen S032257 maddesinin, gerçekleştirilen analizler neticesinde, İngiliz öğrenciler lehine tek biçimli işlev farklılaşması gösterdiği belirlenmiştir. Fizik alan uzmanları, sorunun Türkçe halinin ve cevabının bilimsel olarak doğru olduğunu belirtmişler fakat ölçülen özelliğin kavram yanılgılarına oldukça açık olduğunu belirtmişlerdir. Ek olarak, sorunun İngilizce halinde yer alan iki terime dair ifadelerin sırası, sorunun Türkçe halinde değişiklik göstermiştir ve bu konuda herhangi bir gerekçe belirtilmemiştir. Diğer taraftan gitarın bir enstrüman olarak, ölçülen özellikte kültürel bir farklılık yaratabileceği, bu enstrümana daha aşina olan öğrencilerin dahil olduğu grubun maddeyi doğru cevaplama olasılığının artacağı şüphesi ifade edilmiştir. Bu ifadeler doğrultusunda uzmanlar, A seçeneğinde yer alan ifadelerin yerinin değiştirilmesinin ve madde kökünde gitarın yer almasının İngiliz öğrencilerin lehine bir sonuç yarattığının belirgin bir kanıtı olarak görülemeyeceğini ifade etmişlerdir. Uzman görüşleri doğrultusunda, S032257 maddesinin ölçme sürecinde yanlılık göstermediği kanaatine varılmıştır. Şekil 8’de S032385 maddesinin sırası ile İngilizce ve Türkçe kitapçıklardaki hali yer almaktadır.



Şekil 8: S032385 Maddesinin İngilizce ve Türkçe Kitapçıklarındaki Şekli

Şekil 8'de verilen S032385 maddesinin Türk öğrenciler lehine tek biçimli MİF gösterdiği analizler sonucunda belirlenmiştir ve madde bilgi basamağında yer almaktadır. Biyoloji alan uzmanları, maddenin ve cevabın bilimsel olarak doğru olduğunu, madde kökünün ve seçeneklerin anlaşılır olduğunu ifade etmişler, yabancı dil uzmanları da çeviri açısından eşdeğerliğin sağlandığını belirtmişlerdir. Bu anlamda, söz konusu maddenin bilimsel doğruluk ve dilbilimsel eşdeğerlik anlamında bozucu bir özelliğe sahip olmadığı, yanlılık göstermediği sonucuna ulaşılmıştır. Şekil 9'da S042028 maddesi, İngilizce ve Türkçe kitapçıklarındaki hali ile yer almaktadır.

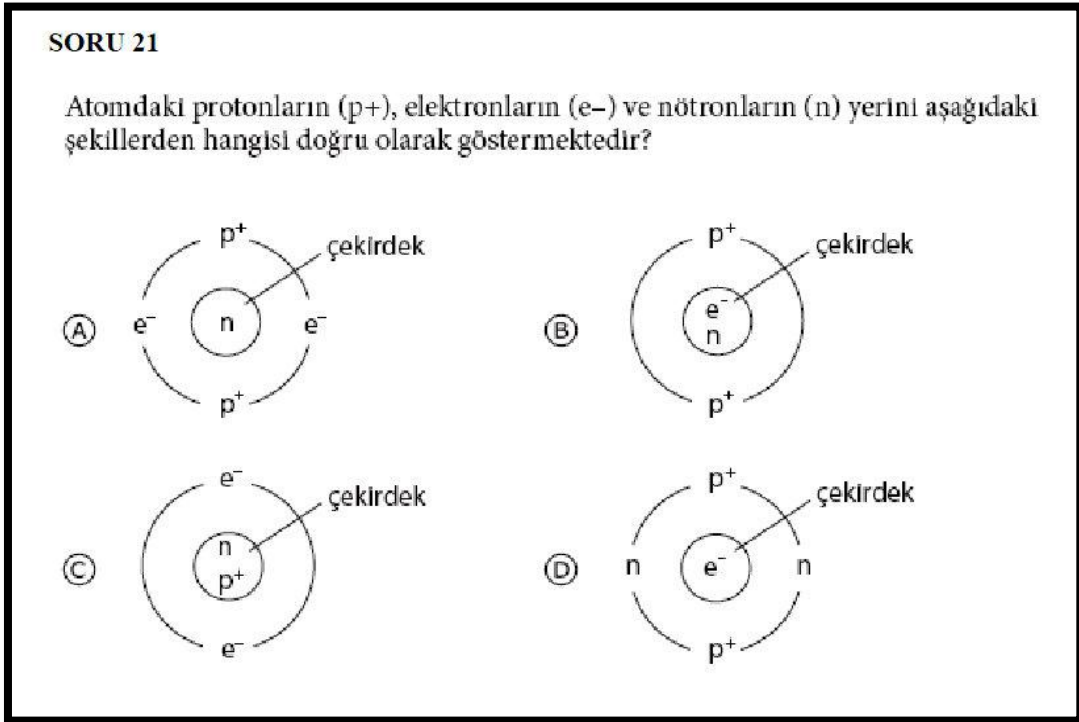
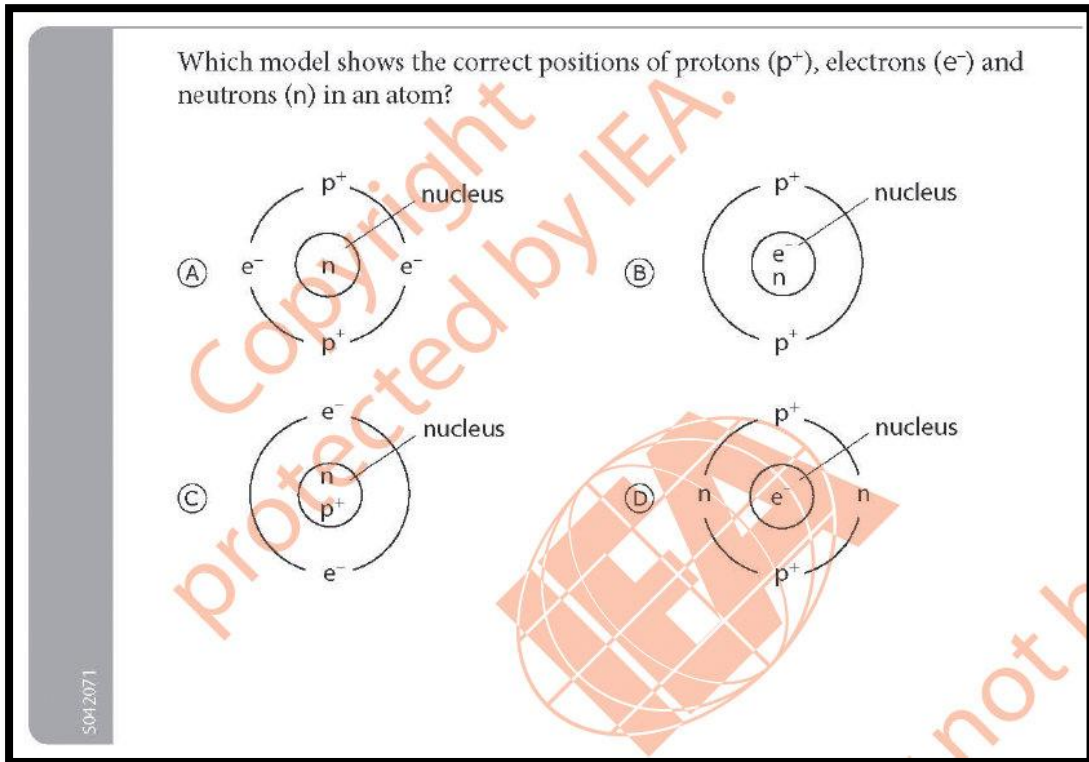


Şekil 9: S042028 Maddesinin İngilizce ve Türkçe Kitapçıklarındaki Şekli

Şekil 9'da verilen S042028 maddesinin yapılan analizler sonucunda, Türk öğrenciler lehine işlev farklılığı gösterdiği belirlenmiştir. Biyoloji alan uzmanları, maddenin ve cevabın bilimsel olarak doğru olduğunu, madde kökü ve seçeneklerin anlaşılır olduğunu ve yabancı dil alan uzmanları çeviri açısından eşdeğerliğin sağlandığını belirtmişlerdir. Bu görüşler doğrultusunda, maddenin yanlılığa sebep olabilecek özellikler göstermediği sonucuna



varılmıştır. Şekil 10'da S042071 maddesi, İngilizce ve Türkçe kitapçıklarındaki hali ile yer almaktadır.



Şekil 10: S042071 Maddesinin İngilizce ve Türkçe Kitapçıklarındaki Şekli

Şekil 10'da incelenen S042071 maddesinin Türk öğrenciler lehine tek biçimli MİF gösterdiği bulunmuştur. Söz konusu maddenin konu alanı olan "Atomun Yapısı"nın Türkiye'de hem Fizik hem de Kimya alanları müfredatında bulunduğu için, Fizik ve Kimya alan uzmanlarının görüşleri bütüncül olarak değerlendirilmiştir. Kimya ve Fizik alan uzmanları, maddenin ve cevabın bilimsel olarak doğru olduğunu, maddenin bir bütün olarak açık ve anlaşılır bir şekilde sunulduğunu belirtmişlerdir. Yabancı dil uzmanları çeviri açısından eşdeğerliğin sağlandığını belirtmişlerdir. Uzman görüşleri doğrultusunda maddenin yanlılığa neden olabilecek unsurlara sahip olmadığına karar verilmiştir.

Dile yönelik MİF gösteren maddelerin olası yanlılık nedenleri incelendiğinde, AERA' nın yanlılığın en önemli nedenlerinden olarak gösterdiği "tercüme hataları", araştırmada S032115 ve S032257 maddelerinde İngiliz ve Türk öğrencilerin puanlarının karşılaştırılabilirliği için bozucu bir etki yaratmıştır fakat bu uzmanlar bu hataların doğrudan bir yanlılık sonucu yaratmayacağını ifade etmişlerdir. Gierl ve Khaliq (2001) ve Uiterwijk ve Vallen (2005)'in çalışmalarında, dile yönelik MİF gösteren maddelerin bulunduğu ve uzman görüşleri sonrasında, bu maddelerin olası yanlılık kaynaklarının, adaptasyon sürecinde kelime ve ifade değişiklikleri yapılması olduğunu belirtmişlerdir. Araştırma çerçevesinde, özellikle S032115 ve S032257 maddelerinde bu duruma örnek teşkil eden durumlar belirlenmiştir, Bu durumda söz konusu çalışmaların bulguları ile araştırmanın bulguları ile uyum göstermektedir.

4. Fen Bilimleri testinde *cinsiyete* göre işlev farklılığı gösterdiği belirlenen, araştırmacıların erişimine açılmış maddelerin olası yanlılık kaynağı konusundaki uzman görüşleri

Mantel-Haenszel, SIBTEST ve lojistik regresyon yöntemleri ile gerçekleştirilen analizler neticesinde, Türkiye örnekleme kapsamında, üç yöntemin tutarlı olarak tek biçimli MİF gösterdiğini belirlediği veya SIBTEST ve lojistik regresyon yöntemlerinin tutarlı olarak tek biçimli olmayan MİF gösterdiğini belirlediği, araştırmacıların erişimine açılmış bir madde bulunmamıştır. Bu sonuca dayanarak, uzman görüşleri alınamamıştır.

## BÖLÜM 5

### SONUÇ VE ÖNERİLER

Bu bölümde, araştırma sorularına yönelik yapılan analizler sonucunda elde edilen sonuçlar ve bu sonuçlar doğrultusunda önerilere yer verilmiştir.

Bu araştırmada IEA tarafından geliştirilen TIMSS 2007 fen bilimleri testinde 14 kitapçıkta yer alan maddelerin dil (Türkiye-İngiltere örneklemeleri) ve cinsiyet (Türkiye örneğinde erkek-kız grupları) yönünden MİF gösterip göstermediği Mantel-Haenszel, SIBTEST ve lojistik regresyon yöntemleri ile analiz edilmiştir. Bilinmektedir ki, uluslararası alanda öğrenci başarılarının karşılaştırılmasına yönelik çalışmaların, test ile ölçülmesi istenen özellikler dışında, kültürel farklılıklardan etkilenmesi durumunda o testten ve o testi oluşturan maddelerden elde edilen puanların karşılaştırılabilirliği tartışmalı hale gelir. Bu amaçla, TIMSS 2007 fen bilimleri testindeki maddelerin dil anlamında Türkiye-İngiltere örneklemelerinde karşılaştırılabilirliği ve Türkiye örnekleme kapsamında kız ve erkek öğrenci grupları açısından karşılaştırılabilirliğine dair elde edilen sonuçlar aşağıda açıklanmıştır.

#### *Türkiye ve İngiltere Örneklemelerinde Gerçekleştirilen Dile Yönelik MİF Analizi Sonuçları*

Türkiye ve İngiltere örneklemelerinin birleştirilerek oluşturulan veri setinde yapılan analizlerde 3., 11. ve 12. Kitapçıklarda Mantel-Haenszel, SIBTEST ve lojistik regresyon analizi sonuçlarının tutarlı olarak olası yanlış ölçme yaptığını gösterdiği, diğer bir ifade ile üç yöntemin birlikte B ve C düzeyinde MİF gösterdiğini belirlediği herhangi bir madde bulunmamaktadır. Buna göre, söz konusu üç kitapçıkta bulunan maddeleri cevaplayan İngiliz ve Türk öğrencilerin maddelerden elde ettikleri puanların karşılaştırılabilir olduğu sonucuna varılmıştır. İçeriğinde bulunan dört maddenin MİF göstermesi ile 14. kitapçık, içeriğindeki en çok sayıda maddenin MİF gösterdiği kitapçık olmuş, söz konusu maddelerin üçü Türk öğrenciler lehine, biri İngiliz öğrenciler lehine MİF gösterdiği belirlenmiştir. Bu kitapçıkta 10 çoktan seçmeli madde olduğu ve bu maddelerden 4'ünün MİF gösterdiği dikkate alındığında, maddeleri açısından en az karşılaştırılabilir puanlar sunan kitapçık olarak 14. kitapçık

belirlenmiştir. Analizler, 1., 2., 4. ve 10. kitapçıklarda MİF gösterdiği belirlenen maddelerin yalnızca Türk öğrenciler lehine, 6. ve 7. kitapçıklarda MİF gösterdiği belirlenen maddelerin yalnızca İngiliz öğrenciler lehine olduğunu göstermiştir.

Araştırmada incelenen 14 farklı kitapçıkta bulunan madde setlerinin ortak maddelere sahip olabilmesi nedeni ile bir maddenin birden çok kitapçıkta bulunma durumu söz konusudur. Bulgular açıkça göstermektedir ki, birden fazla kitapçıkta MİF gösterdiği belirlenen maddeler, farklı kitapçıklarda da olsa, daima aynı grubun lehine işlev farklılığı göstermişlerdir. Aynı maddelerin farklı kitapçıklarda MİF göstermesi durumu göz önüne alındığında, 11 *farklı* maddenin tek biçimli işlev farklılığı gösterdiği ifade edilebilir.

İşlev farklılığı gösterdiği belirlenen 11'i farklı ve 8'i birden fazla kitapçıkta MİF gösteren, toplam 19 maddenin 13'ünün Türk öğrenciler lehine, 6'sının İngiliz öğrenciler lehine MİF gösterdiği saptanmıştır. Bu sonuç, Türk öğrencilerin lehine MİF gösterdiği belirlenen madde sayısının, İngiliz öğrenciler lehine MİF gösterdiği belirlenen madde sayısından daha fazla olduğunu göstermiştir. İşlev farklılığı gösteren maddelerin hepsinin tek biçimli MİF göstermesi neticesinde, tüm yetenek düzeylerinde bir grubun maddeyi doğru cevaplama olasılığının diğer gruba göre sürekli ve tutarlı olarak MİF gösterdiği belirlenen maddeler, toplam maddelerin yaklaşık %18'ini oluşturmaktadır.

### *Türkiye Örneğinde Gerçekleştirilen Cinsiyete Yönelik MİF Analizi*

#### *Sonuçları*

Türkiye örneğinde gerçekleştirilen analizler, Mantel-Haenszel, SIBTEST ve lojistik regresyon analizi sonuçlarının tutarlı olarak MİF gösterdiğini belirlediği herhangi bir maddenin bulunmadığını göstermiştir. Üç yöntem ayrı ayrı incelendiğinde, her birinin tek başına işlev farklılaşması gösterdiğini belirlediği maddeler bulunmakla beraber, her üç yöntemde tutarlı olarak B veya C düzeyinde MİF gösteren madde bulunamamıştır. Bu sonuç itibarıyla, TIMSS 2007 fen bilimleri testinden elde edilen puanların, Türkiye örneğinde kız ve erkek öğrenci grupları için aynı anlama geldiği, karşılaştırılabilir oluşu ve aynı yetenek grubunda bulunan kız ve erkek

öğrencilerin maddeleri doğru yanıtlama olasılıklarının farklılık göstermediği savunulabilir.

*Türkiye ve İngiltere Örneklemlerinde Gerçekleştirilen Dile Yönelik MİF Gösteren, Erişime Açılmış Maddelere Yönelik Uzman Görüşleri*

Analiz sonuçlarına göre MİF gösterdiği belirlenen, araştırmacıların erişimine açılmış maddelerden S032115'de çeviriye bağlı olarak anlamda belirsizlik olduğu ve terimsel eşdeğerliğin olmadığı ifade edilmiştir. Bu durum, puanların Türkiye ve İngiltere örneklemlerinde farklı anlamlar taşımaya yol açabilecek bozucu bir etki oluşturmuştur. Maddenin Türk öğrenciler lehine MİF göstermesinin olası nedenlerinden biri olarak görülmektedir. Uzmanlar, bu durumun istenmeyen bir etki yarattığını fakat açık bir yanlışlık kaynağı olarak görülemeyeceğini belirtmişlerdir.

İşlev farklılığı gösterdiği belirlenen ve erişime açılan maddelerden S032257'de ise, İngilizce ve Türkçe kitapçıklarda yer alan soruların A seçeneğinde yer alan terimlere dair ifadelerin sırası değiştirilmiştir. Seçeneklerde yer alan ifadelerdeki her türlü değişimin, cevap işlevlerinde değişime yol açabileceği ve bu değişimin gerekçelendirilmediği dikkate alınır, bu durumun yanıtlayıcıların tercihlerinde bozucu bir etki yaratabileceği ifade edilmiş fakat yine bu durumun, yanlışlığın açık bir göstergesi olmadığına dikkat çekilmiştir.

MİF gösterdiği belirlenen, erişime açılmış S032385, S042028 ve S042071 maddeleri için alan uzmanları, maddelerin bilimselliğini ve tercüme konusunda deneyimli yabancı dil uzmanları, maddelerin dilbilimsel açısından eşdeğerliğini onaylamışlardır.

Bulguların ortaya koyduğu önemli bir diğer sonuç, MİF gösterdiği belirlenen ve araştırmacıların erişimine açılmış maddelerin hepsinin bilgi düzeyinde olmasıdır. Bu anlamda MİF gösteren ve araştırmacıların erişimine açılmış maddelerden yalnızca biri İngiliz öğrenciler lehine işlev farklılık göstermiştir. Bu durumun önemli bir nedeni, Türk eğitim sisteminin diğer bilişsel düzeylerden çok bilgi düzeyindeki kazanımlara odaklanması ve eğitim sürecinde yapılan ölçmelerde daha çok bilgi düzeyinde bilişsel becerilerin göz

önüne alınması olarak görülmüştür. Süreç sonunda üst düzey bilişsel becerilerden ziyade bilgi düzeyindeki kazanımların merkezinde bulunan Türk öğrencilerin bilgi düzeyindeki maddeleri doğru cevaplama olasılıkları, İngiliz öğrencilerden tutarlı ve sürekli olarak daha yüksek bulunmuştur.

### *İşlev Farklılaşması Gösteren Maddelerin Belirlenmesi İçin Kullanılan Yöntemlerden Elde Edilen Sonuçların Tutarlılığına Yönelik Sonuçlar*

TIMSS 2007 fen bilimleri testindeki maddelerin incelenmesinde kullanılan Mantel-Haenszel ve lojistik regresyon yöntemleri büyük oranda uyumlu sonuçlar vermiştir. Dile yönelik işlev farklılığının belirlenmesi kapsamında yapılan MİF incelemesinde, Mantel-Haenszel ve lojistik regresyon yöntemleri %85.10 düzeyinde, cinsiyete yönelik MİF incelemesi kapsamında ise %97.12 düzeyinde uyum göstermiştir. SIBTEST yöntemi ise, diğer yöntemlere göre daha hassas sonuçlar sağlamış, bu yöntemle, diğer yöntemlere göre sıklıkla daha yüksek MİF düzeyleri belirlenmiştir. SIBTEST ile diğer yöntemler arasındaki uyum dile yönelik MİF analizi kapsamında %39.42, cinsiyete yönelik MİF analizi kapsamında ise %59.14 olarak hesaplanmıştır. Sonuç olarak, Mantel-Haenszel ve lojistik regresyon yöntemlerinin tutarlı sonuçlar sağladığı fakat SIBTEST yönteminin aynı düzeyde tutarlılık göstermediği ifade edilebilir.

## **ÖNERİLER**

Bu araştırma neticesinde yapılabilecek öneriler iki farklı grupta sınıflandırılabilir.

### *Araştırma Sonuçlarından Çıkan Uygulamaya Yönelik Öneriler*

1. TIMSS geliştiricisi IEA tarafından dil ve cinsiyet değişkenlerine göre gerçekleştirilecek madde yanlılığı araştırmalarının artırılması, söz konusu testten elde edilecek puanların karşılaştırılabilirliğinin güvenceye alınması anlamında bir önlem olarak görülebilir.
2. Madde yanlılığı çalışmalarında özellikle incelenen veya karşılaştıran ülkenin/ülkelerin özellikleri dikkate alınarak referans ve odak grupları oluşturulması, araştırmalardan elde edilecek sonuçların ülkelere özgü çözüm önerileri oluşturabileceği dikkate alınmalıdır.

3. Uygulamalarda kullanılan tüm kitapçıkların dâhil edildiği madde yanlılığı çalışmalarında, aynı maddelerin birden farklı kitapçıkta işlev farklılığı gösterip göstermediği ve hangi grup lehine MİF gösterdiği dikkatle incelenmeli, olası farklılıklar üzerinde uzmanların görüşüne başvurulmalıdır.
4. IEA'nın dünyanın farklı ülkelerinde TIMSS ile ilgili yapılan araştırmaları bir çatı altında toplaması, kurumda çalışan uzmanlarca değerlendirmeye tabi tutulması, güncel çalışmaların takibi için önem taşımaktadır. Bu çalışmaların ışığı doğrultusunda TIMSS 2015 ve sonraki uygulamaların daha geçerli, güvenilir ve karşılaştırılabilir sonuçlar sağlaması sağlanabilir.

#### *Araştırmacılara Yönelik Öneriler*

1. Araştırmada TIMSS 2007 fen bilimleri testinde yer alan çoktan seçmeli maddeler üzerinde inceleme yapılmıştır. Gerçekleştirilecek çalışmalarda TIMSS 2011 uygulamasında matematik ve fen bilimleri testinde yer alan maddeler yanlılık açısından incelenebilir.
2. Araştırma fen bilimleri testinde yer alan çoktan seçmeli sorular ile sınırlandırılmıştır. Gerçekleştirilecek çalışmalarda sınırlandırılmış açık uçlu maddeler de MİF analizlerine dâhil edilebilir.
3. Araştırmada Klasik Test Kuramı'na dayalı Mantel-Haenszel, SIBTEST ve lojistik regresyon analizi yöntemleri MİF belirleme süreçlerinde kullanılmıştır. Gerçekleştirilecek güncel çalışmalarda, örneklem büyüklüğü ve araştırmacı yeterliğinin uygun olduğu durumlarda, Madde Tepki Kuramı'na dayalı MİF belirleme yöntemleri kullanılabilir.
4. TIMSS madde geliştirilme sürecinde, orijinal madde ve madde setlerinin bulunduğu kitapçıkların hepsi İngilizce dilinde geliştirilip ardından katılımcı ülkelerin dillerine çevrilmesi nedeniyle, farklı dilde sınava giren ülkeler için dil değişkeni için madde yanlılığı analizleri gerçekleştirilebilir.

## KAYNAKÇA

- Ayan C. (2011). *PISA 2009 Fen Okuryazarlığı Alt Testinin Değişen Madde Fonksiyonu Açısından İncelenmesi*. Yüksek Lisans Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Babiar T. C. (2011). Exploring Differential Item Functioning (DIF) With the Rasch Model: A Comparison of Gender Differences on Eighth Grade Science Items in The United States and Spain. *Journal of Applied Measurement, 12:2, 144-164*.
- Bertrand R.. N. Boiteau (2003). *Comparing the Stability of IRT-based and non IRT-based DIF methods in Different Cultural Contexts Using TIMSS Data*. (<http://www.eric.ed.gov/PDFS/ED476924.pdf>)
- Burkes. L.L. (2009). *Identifying Differential Item Functioning Related to Socioeconomic Status and Investigating Sources Related to Classroom Opportunities to Learn*. (<http://proquest.umi.com/pqdlink?Ver=1&Exp=02-24-2017&FMT=7&DID=1818417291&RQT=309&attempt=1&cfc=1>).
- Camilli G.. Shepard L. A. (1994). *Methods for Identifying Biased Test Items. Volume 4*. California: SAGE Publications. Inc.
- Clauser B. E.. Mazor K. M. (1998). Using Statistical Procedures to Identify Differentially Functioning Test Items. *Educational Measurement: Issues and Practice* (<http://ncme.org/linkservid/80E850A0-1320-5CAE-6E0F967D1CCD586B/showMeta/0/>)
- Crocker L.. Algina J. (1986). *Introduction to Classical and Modern Test Theory*. Orlando. Florida: Holt. Rinehart and Winston Inc.
- Erden M. (2011). *Eğitim Bilimine Giriş*. (5. Baskı). Arkadaş Yayınevi: İstanbul
- Foy P.. Olson J. F. (2009). *TIMSS 2007 User Guide for the International Database*. TIMSS & PIRLS International Study Center.



- Gierl M. J., Khaliq S. N. (2001). Identifying Sources of Differential Item and Bundle Functioning on Translated Achievement Tests: A Confirmatory Analysis. *Journal of Educational Measurement*, 38:2, 164-187.
- Holland P.W., Wainer H. (1993). *Differential Item Functioning*. Hillside. New Jersey: Lawrence Erlbaum Associates. Publishers.
- Innabi. H., Dodeen H. (2006). Content Analysis of Gender-related Differential Item Functioning TIMSS Items in Mathematics in Jordan. *School Science and Mathematics*, 106:8, 328-337.
- Kalaycıoğlu D. B., Berberoğlu G. (2011). Differential Item Functioning Analysis of the Science and Mathematics Items in the University Entrance Examinations in Turkey. *Journal of Psychoeducational Assessment*, 29:467.
- Karami H., Nodoushan M. A. S. (2011). Differential Item Functioning (DIF): Current Problems and Future Directions. *International Journal of Language Studies*, 2011; 5-4: 133-142.
- Kjærnsli M., Lie S. (2008): Country profiles of scientific competence in TIMSS 2003, *Educational Research and Evaluation: An International Journal on Theory and Practice*, 14:1, 73-85.
- Klieme E., Baumert J. (2001). Identifying National Cultures of Mathematics Education: Analysis of Cognitive Demands and Differential Item Functioning in TIMSS. *European Journal of Psychology of Education*, 16:3, 385-402.
- Le, L.T. (2009). *Investigating Gender Differential Item Functioning Across Countries and Test Languages for PISA Science Items*. International Journal of Testing. (<http://www.intestcom.org/Downloads/ITC2006Brussels/Session%203.2.6%20Le.pdf>)
- Lee, J. (1998). *Missing links in international educational studies: Comparing the US with Eastern Asian countries in TIMSS*. (<http://iejll.synergiesprairies.ca/iejll/index.php/iejll/article/view/462/124>)

- Martin M. O., Mullis I., Foy P. (2008). *TIMSS 2007 International Science Report*. TIMSS & PIRLS International Study Center.
- McMillan J. H. (2007). *Classroom Assessment. Principles and Practice for Effective Standards-Based Instruction (4. Edition)*. Chapter 3: High-Quality Classroom Assessment - Test Fairness. Boston: Pearson Education. Inc.
- Moses T., Miao J., Dorans N. (2010). *A Comparison of Methods for Estimating Conditional Item Score Differences in Differential Item Functioning (DIF) Assessments*. ETS: Princeton, New Jersey
- Mullis I. V. S., Martin M. O., Ruddock G. J., O'Sullivan C. Y., Arora A., Erberber E. (2007). *TIMSS 2007 Assessment Frameworks*. TIMSS & PIRLS International Study Center.
- Nillas L. (2002). Does Language Make a Difference: A TIMSS-R Analysis. *The Mathematics Educator*, 6:2, 95-112.
- Olson J. F., Martin M. O., Mullis I. V. S. (2008). *TIMSS 2007 Technical Report*. TIMSS & PIRLS International Study Center.
- Osterlind S. J., Everson H. T. (2009). *Differential Item Functioning: Second Edition*. California: SAGE Publications. Inc.
- Perrone. M. (2006). *Differential Item Functioning and Item Bias: Critical Considerations in Test Fairness*. Teachers College. Columbia University Working Papers in TESOL & Applied Linguistics.
- Qian. X. (2011). *A Multi-Level Differential Item Functioning Analysis of Trends In International Mathematics And Science Study: Potential Sources Of Gender And Minority Difference Among U.S. Eighth Graders' Science Achievement*. (<http://gradworks.umi.com/34/65/3465815.html>).
- Reynolds. C. R., Livingston R. B., Willson V. (2006). *Measurement and Assessment in Education*. Chapter 16: The Problem of Bias in Educational Assessment. Boston: Pearson Education. Inc.
- Roever. C. (2005). *"That's not fair!". Fairness, bias and differential item functioning in language testing*. SLS Brownbag.

- Roussos. L.. Stout W. (1996). Simulation Studies of the Effects of Small Sample Size and Studied Item Parameters on SIBTEST and Mantel-Haenszel Type I Error Performance. *Journal of Educational Measurement*, 33:2, 215-230.
- Shealy R.. Stout W. F. (1993). A Model Based Standardization Approach that Separates True Bias/DIF From Group Ability Differences And Detects Test Bias/DTF As Well As Item Bias/DIF. *Psychometrika*. 159-194.
- Stoneberg B. D. (2004). *A Study of Gender Based and Ethnic Based Differential Item Functioning (DIF) in the Spring 2003 Idaho Standards Achievement Tests Applying the Simultaneous Bias Test (SIBTEST) and the Mantel-Haenszel Chi-Square Test.* (<http://www.eric.ed.gov/PDFS/ED489949.pdf>)
- Swaminathan H., Rogers H. J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement* 27; 361-370.
- Şişman M.. Acat M. B.. Aypay A.. Karadağ E. (2011). *TIMSS 2007 Ulusal Matematik ve Fen Raporu. 8. Sınıflar.* Ankara: Vaktaş Okul Donatım. Basın-Yayım.
- Uiterwijk H., Vallen T. (2005). Linguistic Sources of Item Bias For Second Generation Immigrants in Dutch Tests. *Language Testing* 22:211
- Ulutaş, S. (2012). *PISA 2006 Fen Okuryazarlığı Testindeki Maddelerin Yanlılık Bakımından Araştırılması.* Yüksek Lisans Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü.
- Van der Linden. W. J. (1998). A Discussion of Some Methodological Issues in International Assessments. *International Journal of Educational Research* 29. Pages: 569-577.
- Walker C. M. (2011). What's the DIF? Why Differential Item Functioning Analyses Are an Important Part of Instrument Development and Validation. *Journal of Psychoeducational Assessment* 2011 29: 364.

- Wu A. D., Ercikan K. (2009). Using Multiple-Variable Matching to Identify Cultural Sources of Differential Item Functioning. *International Journal of Testing*, 6:3, 287-300.
- Yıldırım H. H. (2006). *The Differential Item Functioning (DIF) Analysis Of Mathematics Items In The International Assessment Programs*. Yayınlanmamış Doktora Tezi, Ankara: Ortadoğu Teknik Üniversitesi.
- Zenisky A. L., Hambleton R. K., Robin F. (2003). *DIF Detection and Interpretation in Large Scale Science Assessments: Informing Item Writing Practices*. Center for Educational Assessment MCAS Validity Report No. 1. (CEA-429).
- Zhang Y. (2002). *DIF in a Large Scale Mathematics Assessment: The Interaction of Gender and Ethnicity*. American Educational Research Association Conference, New Orleans.
- Zhou J., Gierl M. J. Tan X. (2006). *Evaluating the Performance of MULTISIB Using Different Multidimensional Matching Criteria*. ([http://www2.education.ualberta.ca/educ/psych/crame/files/ncme06\\_jz.pdf](http://www2.education.ualberta.ca/educ/psych/crame/files/ncme06_jz.pdf))
- 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*. (<http://www.educ.ubc.ca/faculty/zumbo/DIF/handbook.pdf>)
- 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. ([http://educ.ubc.ca/faculty/zumbo/papers/Zumbo\\_LAQ\\_reprint.pdf](http://educ.ubc.ca/faculty/zumbo/papers/Zumbo_LAQ_reprint.pdf))

## EK-1

### 1. Kitapçık

Madde Numarası	Madde	$\chi^2$	$\alpha_i$	MİF Düzeyi
1	S032115	70.188	-1.996	<b>C</b>
2	S032403	7.734	0.545	<b>A</b>
3	S032273	25.983	1.018	<b>B</b>
4	S032620	3.711	0.381	<b>A</b>
5	S042009	27.365	-1.149	<b>B</b>
6	S042059	16.529	0.761	<b>A</b>
7	S042028	132.214	-2.840	<b>C</b>
8	S042001	30.138	1.151	<b>B</b>
9	S042276	1.938	0.269	<b>A</b>
10	S042279	0.774	-0.178	<b>A</b>
11	S042071	106.778	-2.030	<b>C</b>
12	S042405	0.780	-0.179	<b>A</b>

### 2. Kitapçık

Madde numarası	Madde	$\beta$	Standart Hata	p	MİF Düzeyi
1	S042009	-0.117	0.038	.002	<b>C</b>
2	S042059	0.266	0.041	.000	<b>C</b>
3	S042028	-0.434	0.033	.000	<b>C</b>
4	S042001	0.271	0.038	.000	<b>C</b>
5	S042276	0.169	0.043	.000	<b>C</b>
6	S042279	0.026	0.042	.534	-
7	S042071	-0.372	0.036	.000	<b>C</b>
8	S042405	0.039	0.040	.332	-
9	S022183	0.173	0.040	.000	<b>C</b>
10	S022276	-0.052	0.039	.177	-
11	S022115	-0.101	0.038	.008	<b>C</b>
12	S022019	0.111	0.037	.003	<b>C</b>
13	S022002	0.100	0.041	.015	<b>C</b>
14	S022294	-0.094	0.042	.025	<b>C</b>
15	S022150	-0.136	0.041	.001	<b>C</b>
16	S022042	0.284	0.040	.000	<b>C</b>

### 3. Kitapçık

Madde Numarası	Madde	Aşamalı Lojistik Regresyonun Basamaklarındaki R <sup>2</sup> Değerleri			MİF Ki-Kare Testi	MİF R <sup>2</sup> Farkı ( $\Delta R^2$ )	MİF Düzeyi
		Adım1	Adım2	Adım3			
		1	S022183	0.271			
2	S022276	0.254	0.258	0.260	3.401 p=.182	0.006	-
3	S022215	0.242	0.287	0.296	29.681 P=.000	0.054	<b>A</b>
4	S022019	0.286	0.286	0.303	9.253 p=.010	0.017	<b>A</b>
5	S022002	0.335	0.345	0.347	7.522 p=.023	0.012	<b>A</b>
6	S022294	0.132	0.141	0.142	5.181 p=.075	0.010	-
7	S022150	0.147	0.170	0.171	12.302 p=.002	0.024	<b>A</b>
8	S022042	0.325	0.346	0.350	15.163 p=.000	0.025	<b>A</b>
9	S042013	0.323	0.335	0.336	6.569 p=.062	0.013	-
10	S042006	0.075	0.090	0.097	10.669 p=.005	0.022	<b>A</b>
11	S042054	0.100	0.121	0.123	11.162 p=.004	0.023	<b>A</b>
12	S042061	0.070	0.092	0.097	12.763 p=.002	0.027	<b>A</b>
13	S042109	0.380	0.386	0.399	12.289 p=.002	0.019	<b>A</b>
14	S042232C	0.005	0.016	0.017	5.574 p=.062	0.012	-
15	S042294	0.090	0.181	0.183	45.965 p=.000	0.093	<b>A</b>
16	S042150	0.094	0.106	0.125	14.901 p=.000	0.031	<b>A</b>

#### 4. Kitapçık

Madde Numarası	Madde	$\chi^2$	$\alpha_i$	MİF Düzeyi
1	S042013	2.522	0.458	<b>A</b>
2	S042006	15.855	-0.832	<b>A</b>
3	S042054	20.624	0.876	<b>A</b>
4	S042061	18.114	0.815	<b>A</b>
5	S042109	3.368	0.436	<b>A</b>
6	S042232C	0.000	0.014	<b>A</b>
7	S042294	59.631	-1.834	<b>C</b>
8	S042150	4.957	0.439	<b>A</b>
9	S022290	0.388	0.156	<b>A</b>
10	S022054	2.710	-0.348	<b>A</b>
11	S022181	19.704	-0.993	<b>A</b>
12	S022208	39.167	-1.366	<b>B</b>
13	S022126	11.413	0.700	<b>A</b>
14	S032385	103.958	-2.316	<b>C</b>
15	S032035	9.543	-0.737	<b>A</b>
16	S032683	0.169	-0.106	<b>A</b>
17	S032258	16.403	0.810	<b>A</b>

#### 5. Kitapçık

Aşamalı Lojistik							
Madde Numarası	Madde	Regresyonun Basamaklarındaki R <sup>2</sup> Değerleri			MİF Ki-Kare Testi	MİF R <sup>2</sup> Farkı ( $\Delta R^2$ )	MİF Düzeyi
		Adım1	Adım2	Adım3			
		1	S022290	0.277			
2	S022054	0.126	0.135	0.157	15.605 p=.000	0.031	<b>A</b>
3	S022181	0.194	0.252	0.254	32.682 p=.000	0.060	<b>A</b>
4	S022208	0.12	0.122	0.123	1.014 p=.602	0.003	-
5	S022126	0.024	0.043	0.086	25.458 p=.000	0.062	<b>A</b>
6	S032385	0.023	0.289	0.292	131.043 p=.000	0.269	<b>C</b>
7	S032035	0.179	0.252	0.254	37.276 p=.000	0.075	<b>A</b>
8	S032683	0.243	0.246	0.246	2.123 p=.346	0.003	-

9	S032258	0.324	0.342	0.349	15.163 p=.000	0.025	<b>A</b>
10	S042304	0.206	0.211	0.211	2.846 p=.241	0.005	-
11	S042038	0.002	0.150	0.154	49.195 p=.000	0.152	<b>B</b>
12	S042306	0.151	0.152	0.165	7.362 p=.025	0.014	<b>A</b>
13	S042272	0.040	0.049	0.052	4.240 p=.120	0.012	-
14	S042238A	0.119	0.120	0.121	0.937 p=.626	0.002	-
15	S042141	0.359	0.374	0.374	8.870 p=.012	0.015	<b>A</b>
16	S042215	0.039	0.111	0.117	35.125 p=.000	0.078	<b>A</b>

## 6. Kitapçık

Madde Numarası	Madde	$\chi^2$	$\alpha_i$	MİF Düzeyi
1	S042304	0.713	-0.203	<b>A</b>
2	S042038	10.172	0.719	<b>A</b>
3	S042306	0.131	0.090	<b>A</b>
4	S042272	0.519	0.166	<b>A</b>
5	S042238A	1.742	0.266	<b>A</b>
6	S042141	23.406	1.060	<b>B</b>
7	S042215	17.849	-0.895	<b>A</b>
8	S032606	7.461	-0.702	<b>A</b>
9	S032672	13.792	0.758	<b>A</b>
10	S032392	6.217	-0.649	<b>A</b>
11	S032425	42.532	-1.373	<b>B</b>
12	S032257	84.945	1.940	<b>C</b>
13	S032663	0.009	0.000	<b>A</b>
14	S032660	12.434	0.711	<b>A</b>

## 7. Kitapçık

Madde numarası	Madde	$\beta$	Standart Hata	p	MİF Düzeyi
1	S032606	-0.187	0.033	.000	<b>C</b>
2	S032672	0.151	0.045	.001	<b>C</b>
3	S032392	-0.204	0.041	.000	<b>C</b>
4	S032425	-0.246	0.045	.000	<b>C</b>



5	S032257	0.290	0.036	.000	<b>C</b>
6	S032663	-0.118	0.042	.005	<b>C</b>
7	S032660	0.217	0.049	.000	<b>C</b>
8	S042053	-0.022	0.041	.590	-
9	S042015	-0.187	0.042	.000	<b>C</b>
10	S042309	0.018	0.044	.677	-
11	S042182	-0.065	0.042	.127	-
12	S042126	0.263	0.039	.000	<b>C</b>
13	S042210	-0.117	0.043	.007	<b>C</b>
14	S042257	-0.060	0.045	.184	-

## 8. Kitapçık

Aşamalı Lojistik							
Madde Numarası	Madde	Regresyonun Basamaklarındaki R <sup>2</sup> Değerleri			MİF Ki-Kare Testi	MİF R <sup>2</sup> Farkı (ΔR <sup>2</sup> )	MİF Düzeyi
		Adım1	Adım2	Adım3			
		1	S042053	0.454			
2	S042015	0.299	0.303	0.326	15.630 p=.000	0.027	<b>A</b>
3	S042309	0.064	0.084	0.087	11.155 p=.004	0.023	<b>A</b>
4	S042182	0.258	0.258	0.271	7.137 p=.028	0.013	<b>A</b>
5	S042126	0.119	0.201	0.213	45.238 p=.000	0.094	<b>A</b>
6	S042210	0.069	0.075	0.085	7.290 p=.026	0.016	<b>A</b>
7	S042257	0.120	0.123	0.124	1.713 p=.425	0.004	-
8	S032542	0.137	0.154	0.169	16.012 p=.000	0.032	<b>A</b>
9	S032645	0.224	0.248	0.253	15.274 p=.000	0.029	<b>A</b>
10	S032502	0.273	0.273	0.274	0.087 p=.957	0.001	-
11	S032184	0.092	0.094	0.100	4.099 p=.129	0.008	-
12	S032394	0.200	0.203	0.203	1.486 p=.476	0.003	-
13	S032151	0.175	0.176	0.240	34.135 p=.000	0.065	<b>A</b>

## 9. Kitapçık

Madde Numarası	Madde	$\chi^2$	$\alpha_i$	MİF Düzeyi
1	S032542	0.143	-0.090	<b>A</b>
2	S032645	10.086	-0.071	<b>A</b>
3	S032502	2.396	-0.354	<b>A</b>
4	S032184	0.320	-0.121	<b>A</b>
5	S032394	0.210	0.107	<b>A</b>
6	S032151	4.753	-0.459	<b>A</b>
7	S042073	0.545	0.236	<b>A</b>
8	S042017	2.830	0.352	<b>A</b>
9	S042007	6.554	-0.593	<b>A</b>
10	S042024	12.931	-0.735	<b>A</b>
11	S042095	0.865	-0.237	<b>A</b>
12	S042063	13.743	1.194	<b>B</b>
13	S042197	75.841	1.648	<b>C</b>
14	S042112	11.543	0.643	<b>A</b>
15	S042274	68.499	-1.810	<b>C</b>

## 10. Kitapçık

Madde numarası	Madde	$\beta$	Standart Hata	p	MİF Düzeyi
1	S042073	-0.023	0.031	.451	-
2	S042017	0.093	0.041	.023	<b>C</b>
3	S042007	-0.176	0.037	.000	<b>C</b>
4	S042024	-0.079	0.043	.071	-
5	S042095	0.017	0.036	.637	-
6	S042063	0.142	0.027	.000	<b>C</b>
7	S042197	0.321	0.039	.000	<b>C</b>
8	S042112	-0.010	0.044	.817	-
9	S042274	-0.318	0.038	.000	<b>C</b>
10	S032465	0.243	0.040	.000	<b>C</b>
11	S032315	0.139	0.041	.001	<b>C</b>
12	S032579	-0.150	0.040	.000	<b>C</b>
13	S032024	0.079	0.042	.059	-
14	S032141	-0.120	0.038	.001	<b>C</b>
15	S032463	-0.244	0.037	.000	<b>C</b>
16	S032514	-0.136	0.044	.002	<b>C</b>

## 11. Kitapçık

Aşamalı Lojistik							
Madde Numarası	Madde	Regresyonun Basamaklarındaki R <sup>2</sup> Değerleri			MİF Ki-Kare Testi	MİF R <sup>2</sup> Farkı (ΔR <sup>2</sup> )	MİF Düzeyi
		Adım1	Adım2	Adım3			
		1	S032465	0.269			
2	S032315	0.258	0.294	0.316	34.515 p=.000	0.058	<b>A</b>
3	S032579	0.052	0.076	0.077	11.838 p=.003	0.025	<b>A</b>
4	S032024	0.208	0.235	0.251	24.204 p=.000	0.043	<b>A</b>
5	S032141	0.138	0.138	0.161	11.819 p=.003	0.023	<b>A</b>
6	S032463	0.323	0.376	0.376	31.708 p=.000	0.053	<b>A</b>
7	S032514	0.108	0.113	0.122	7.038 p=.030	0.014	<b>A</b>
8	S042042	0.262	0.286	0.295	16.568 p=.000	0.033	<b>A</b>
9	S042003	0.038	0.046	0.056	8.447 p=.015	0.018	<b>A</b>
10	S042110	0.164	0.174	0.191	13.692 p=.001	0.027	<b>A</b>
11	S042222C	0.257	0.276	0.297	22.857 p=.000	0.040	<b>A</b>
12	S042065	0.317	0.319	0.33	6.506 p=.039	0.013	<b>A</b>
13	S042280	0.339	0.388	0.391	33.241 p=.000	0.052	<b>A</b>
14	S042218	0.277	0.278	0.281	2.261 p=.323	0.004	-
15	S042312	0.049	0.05	0.052	1.269 p=.530	0.003	-
16	S042217	0.247	0.26	0.265	10.375 p=.006	0.018	<b>A</b>

## 12. Kitapçık

Madde Numarası	Madde	$\chi^2$	$\alpha_i$	MİF Düzeyi
1	S042042	14.255	-1.040	<b>B</b>
2	S042003	8.083	-0.585	<b>A</b>
3	S042110	6.517	0.535	<b>A</b>
4	S04222C	3.050	0.388	<b>A</b>
5	S042065	0.428	0.217	<b>A</b>
6	S042280	4.642	-0.577	<b>A</b>
7	S042218	6.329	-0.534	<b>A</b>
8	S042312	3.064	-0.388	<b>A</b>
9	S042217	4.408	-0.456	<b>A</b>
10	S032611	41.765	-1.406	<b>B</b>
11	S032156	1.325	0.248	<b>A</b>
12	S032087	20.704	-0.933	<b>A</b>
13	S032279	7.732	-0.568	<b>A</b>
14	S032238	11.982	-0.754	<b>A</b>
15	S032160	0.000	0.018	<b>A</b>
16	S032654	24.733	-0.964	<b>A</b>
17	S032510	2.571	-0.384	<b>A</b>
18	S032158	0.038	-0.056	<b>A</b>

## 13. Kitapçık

Madde numarası	Madde	$\beta$	Standart Hata	p	MİF Düzeyi
1	S032611	-0.096	0.038	.012	<b>C</b>
2	S032156	0.212	0.039	.000	<b>C</b>
3	S032087	-0.103	0.040	.010	<b>C</b>
4	S032279	-0.045	0.040	.260	-
5	S032238	-0.086	0.039	.027	<b>B</b>
6	S032160	0.093	0.037	.012	<b>C</b>
7	S032654	-0.176	0.041	.000	<b>C</b>
8	S032510	-0.041	0.037	.272	-
9	S032158	0.062	0.040	.120	-
10	S042258	-0.243	0.038	.000	<b>C</b>
11	S042016	-0.485	0.033	.000	<b>C</b>
12	S042068	-0.123	0.037	.001	<b>C</b>
13	S042216	0.487	0.036	.000	<b>C</b>
14	S042249	0.273	0.037	.000	<b>C</b>
15	S042164	0.077	0.037	.038	<b>B</b>

## 14. Kitapçık

Madde Numarası	Madde	Aşamalı Lojistik Regresyonun Basamaklarındaki R <sup>2</sup> Değerleri			MİF Ki-Kare Testi	MİF R <sup>2</sup> Farkı (ΔR <sup>2</sup> )	MİF Düzeyi
		Adım1	Adım2	Adım3			
		1	S032115	0.098			
2	S032403	0.151	0.156	0.157	2.891 p=.236	0.006	-
3	S032273	0.131	0.204	0.215	43.223 p=.000	0.084	<b>A</b>
4	S032620	0.106	0.116	0.131	11.867 p=.003	0.025	<b>A</b>
5	S042258	0.058	0.276	0.276	112.045 p=.000	0.218	<b>C</b>
6	S042016	0.023	0.300	0.306	141.946 p=.000	0.283	<b>C</b>
7	S042068	0.197	0.204	0.216	9.874 p=.007	0.019	<b>A</b>
8	S042216	0.271	0.416	0.433	99.895 p=.000	0.162	<b>B</b>
9	S042249	0.216	0.226	0.232	8.657 p=.013	0.016	<b>A</b>
10	S042164	0.279	0.289	0.300	11.471 p=.003	0.021	<b>A</b>

## EK-2

### 1. Kitapçık

Madde Numarası	Madde	$\chi^2$	$\alpha_i$	MİF Düzeyi
1	S032115	0.581	0.226	A
2	S032403	2.031	-0.382	A
3	S032273	0.745	0.293	A
4	S032620	0.295	0.170	A
5	S042009	1.226	-0.336	A
6	S042059	0.804	0.254	A
7	S042028	0.057	-0.092	A
8	S042001	0.763	0.374	A
9	S042276	0.002	0.039	A
10	S042279	0.005	-0.011	A
11	S042071	1.128	0.311	A
12	S042405	0.702	-0.229	A

### 2. Kitapçık

Madde numarası	Madde	$\beta$	Standart Hata	p	MİF Düzeyi
1	S042009	0.037	0.048	.444	-
2	S042059	0.053	0.052	.305	-
3	S042028	-0.146	0.052	.005	C
4	S042001	0.014	0.036	.696	-
5	S042276	0.022	0.056	.694	-
6	S042279	-0.009	0.057	.873	-
7	S042071	-0.136	0.050	.006	C
8	S042405	-0.109	0.054	.043	C
9	S022183	-0.001	0.048	.983	-
10	S022276	-0.186	0.050	.000	C
11	S022115	-0.014	0.050	.772	-
12	S022019	-0.004	0.054	.945	-
13	S022002	0.019	0.051	.711	-
14	S022294	0.162	0.051	.001	C
15	S022150	0.056	0.052	.286	-
16	S022042	0.084	0.053	.110	-

### 3. Kitapçık

Aşamalı Lojistik							
Madde Numarası	Madde	Regresyonun Basamaklarındaki R <sup>2</sup> Değerleri			MİF Ki-Kare Testi	MİF R <sup>2</sup> Farkı ( $\Delta R^2$ )	MİF Düzeyi
		Adım1	Adım2	Adım3			
		1	S022183	0.152			
2	S022276	0.196	0.211	0.215	5.338 p=.069	0.019	-
3	S022115	0.200	0.221	0.233	8.872 p=.012	0.033	<b>A</b>
4	S022019	0.361	0.361	0.364	1.051 p=.591	0.003	-
5	S022002	0.206	0.214	0.222	4.189 p=.123	0.016	-
6	S022294	0.108	0.124	0.124	4.242 p=.120	0.016	-
7	S022150	0.182	0.188	0.191	12.284 p=.002	0.009	<b>A</b>
8	S022042	0.168	0.181	0.182	3.650 p=.161	0.014	-
9	S042013	0.251	0.251	0.260	2.252 p=.324	0.009	-
10	S042006	0.045	0.055	0.057	2.954 p=.228	0.012	-
11	S042054	0.035	0.045	0.048	3.182 p=.204	0.013	-
12	S042061	0.008	0.011	0.026	4.050 p=.132	0.018	-
13	S042109	0.198	0.218	0.232	9.379 p=.009	0.034	<b>A</b>
14	S042232C	0.002	0.003	0.003	0.199 p=.905	0.001	-
15	S042294	0.195	0.199	0.200	1.300 p=.522	0.005	-
16	S042150	0.007	0.007	0.010	0.628 p=.730	0.003	-

#### 4. Kitapçık

Madde Numarası	Madde	$\chi^2$	$\alpha_i$	MİF Düzeyi
1	S042013	0.294	0.190	<b>A</b>
2	S042006	1.209	0.289	<b>A</b>
3	S042054	1.670	-0.343	<b>A</b>
4	S042061	0.430	0.197	<b>A</b>
5	S042104	3.802	0.568	<b>A</b>
6	S042232C	3.431	0.481	<b>A</b>
7	S042294	0.744	0.097	<b>A</b>
8	S042150	3.220	-0.468	<b>A</b>
9	S022290	7.721	0.721	<b>A</b>
10	S022054	0.007	-0.048	<b>A</b>
11	S022181	0.012	-0.003	<b>A</b>
12	S022208	0.176	-0.141	<b>A</b>
13	S022126	1.453	-0.328	<b>A</b>
14	S032385	1.751	-0.453	<b>A</b>
15	S032035	0.066	-0.108	<b>A</b>
16	S032683	0.082	-0.111	<b>A</b>
17	S032258	3.093	0.450	<b>A</b>

#### 5. Kitapçık

Madde numarası	Madde	$\beta$	Standart Hata	p	MİF Düzeyi
1	S022290	0.254	0.051	.000	<b>C</b>
2	S022054	-0.021	0.057	.712	-
3	S022181	-0.040	0.055	.468	-
4	S022208	-0.178	0.052	.001	<b>C</b>
5	S022126	-0.040	0.054	.461	-
6	S032385	-0.064	0.042	.128	<b>B</b>
7	S032035	0.033	0.049	.502	-
8	S032683	-0.128	0.053	.016	<b>C</b>
9	S032258	0.001	0.055	.980	-
10	S042304	-0.052	0.055	.349	-
11	S042038	0.133	0.041	.001	<b>C</b>
12	S042306	-0.046	0.056	.416	-
13	S042272	0.070	0.055	.200	-
14	S042238A	-0.001	0.041	.984	-
15	S042141	-0.046	0.056	.410	-
16	S042215	0.000	0.052	.997	-



## 6. Kitapçık

Aşamalı Lojistik							
Madde Numarası	Madde	Regresyonun Basamaklarındaki R <sup>2</sup> Değerleri			MİF Ki-Kare Testi	MİF R <sup>2</sup> Farkı (ΔR <sup>2</sup> )	MİF Düzeyi
		Adım1	Adım2	Adım3			
		1	S042304	0.177			
2	S042038	0.045	0.079	0.086	8.432 p=.015	0.041	<b>A</b>
3	S042306	0.158	0.171	0.180	6.035 p=.049	0.022	<b>A</b>
4	S042272	0.101	0.101	0.102	0.322 p=.851	0.001	-
5	S042238A	0.081	0.093	0.094	3.331 p=.189	0.013	-
6	S042141	0.198	0.199	0.199	0.258 p=.879	0.001	-
7	S042215	0.026	0.027	0.032	1.395 p=.498	0.006	-
8	S032606	0.188	0.188	0.198	2.761 p=.251	0.010	-
9	S032672	0.102	0.111	0.112	2.532 p=.282	0.010	-
10	S032392	0.044	0.044	0.056	2.145 p=.342	0.012	-
11	S032425	0.236	0.238	0.240	1.165 p=.559	0.004	-
12	S032257	0.008	0.012	0.012	0.669 p=.716	0.004	-
13	S032663	0.085	0.116	0.116	8.189 p=.017	0.031	<b>A</b>
14	S032660	0.029	0.030	0.031	0.606 p=.738	0.002	-

## 7. Kitapçık

Madde Numarası	Madde	$\chi^2$	$\alpha_i$	MİF Düzeyi
1	S032606	2.123	-0.487	<b>A</b>
2	S032672	0.092	-0.096	<b>A</b>
3	S032392	0.227	0.248	<b>A</b>
4	S032425	0.547	-0.211	<b>A</b>

5	S032257	0.001	0.061	<b>A</b>
6	S032663	0.741	0.243	<b>A</b>
7	S032660	1.318	-0.325	<b>A</b>
8	S042053	2.304	-0.437	<b>A</b>
9	S042015	0.473	0.223	<b>A</b>
10	S042309	1.829	0.337	<b>A</b>
11	S042182	0.363	0.184	<b>A</b>
12	S042126	4.401	0.518	<b>A</b>
13	S042210	0.490	-0.191	<b>A</b>
14	S042257	0.676	0.226	<b>A</b>

### 8. Kitapçık

Madde numarası	Madde	$\beta$	Standart Hata	p	MİF Düzeyi
1	S042053	-0.056	0.050	.258	-
2	S042015	-0.026	0.050	.611	-
3	S042309	-0.086	0.057	.129	-
4	S042182	0.001	0.055	.980	-
5	S042126	0.074	0.058	.199	-
6	S042210	0.070	0.055	.205	-
7	S042257	-0.041	0.054	.444	-
8	S032542	0.084	0.057	.140	-
9	S032645	-0.039	0.053	.464	-
10	S032502	0.018	0.048	.711	-
11	S032184	-0.172	0.054	.001	<b>C</b>
12	S032394	-0.021	0.053	.690	-
13	S032151	0.051	0.054	.342	-

### 9. Kitapçık

Aşamalı Lojistik							
Madde Numarası	Madde	Regresyonun Basamaklarındaki R <sup>2</sup> Değerleri			MİF Ki-Kare Testi	MİF R <sup>2</sup> Farkı ( $\Delta R^2$ )	MİF Düzeyi
		Adım1	Adım2	Adım3			
		1	S032542	0.200			
2	S032645	0.219	0.219	0.219	0.037 p=.982	0.000	-
3	S032502	0.343	0.344	0.356	3.668 p=.160	0.013	-
4	S032184	0.099	0.101	0.101	0.548	0.002	-

					p=.760		
5	S032394	0.174	0.174	0.174	0.122	0.000	-
					p=.941		
6	S032151	0.177	0.205	0.221	12.244	0.044	<b>A</b>
					p=.002		
7	S042073	0.161	0.161	0.161	0.152	0.000	-
					p=.927		
8	S042017	0.211	0.215	0.220	2.582	0.009	-
					p=.274		
9	S042007	0.274	0.279	0.284	2.916	0.010	-
					p=.233		
10	S042024	0.239	0.250	0.251	3.580	0.012	-
					p=.167		
11	S042095	0.192	0.209	0.211	5.253	0.019	-
					p=.072		
12	S042063	0.260	0.312	0.313	14.912	0.053	<b>A</b>
					p=.001		
13	S042197	0.005	0.005	0.009	0.801	0.004	-
					p=.670		
14	S042112	0.002	0.004	0.007	1.248	0.005	-
					p=.536		
15	S042274	0.044	0.046	0.046	0.453	0.002	-
					p=.797		

### 10. Kitapçık

Madde Numarası	Madde	$\chi^2$	$\alpha_i$	MİF Düzeyi
1	S042073	0.006	0.019	<b>A</b>
2	S042017	5.080	-0.695	<b>A</b>
3	S042007	0.057	-0.110	<b>A</b>
4	S042024	8.896	-0.793	<b>A</b>
5	S042095	0.062	-0.093	<b>A</b>
6	S042063	11.098	1.023	<b>B</b>
7	S042197	2.613	0.536	<b>A</b>
8	S042112	0.232	-0.141	<b>A</b>
9	S042274	2.540	0.389	<b>A</b>
10	S032465	0.007	-0.008	<b>A</b>
11	S032315	0.942	0.297	<b>A</b>
12	S032579	9.270	-0.818	<b>A</b>
13	S032024	0.366	0.189	<b>A</b>
14	S032141	1.145	0.338	<b>A</b>
15	S032463	0.040	-0.093	<b>A</b>
16	S032514	0.066	-0.088	<b>A</b>

## 11. Kitapçık

Madde numarası	Madde	$\beta$	Standart Hata	p	MİF Düzeyi
1	S032465	-0.098	0.054	.069	<b>C</b>
2	S032315	0.005	0.052	.919	-
3	S032579	0.011	0.054	.833	-
4	S032024	0.064	0.053	.230	-
5	S032141	0.065	0.051	.200	-
6	S032463	-0.093	0.052	.071	<b>C</b>
7	S032514	-0.079	0.053	.138	-
8	S042042	0.009	0.047	.845	-
9	S042003	0.013	0.052	.797	-
10	S042110	-0.149	0.056	.007	<b>C</b>
11	S042222C	-0.118	0.056	.036	<b>C</b>
12	S042065	0.107	0.050	.033	<b>C</b>
13	S042280	-0.056	0.053	.283	-
14	S042218	0.056	0.052	.284	-
15	S042312	0.092	0.056	.100	<b>C</b>
16	S042217	0.097	0.054	.071	<b>C</b>

## 12. Kitapçık

Aşamalı Lojistik							
Madde Numarası	Madde	Regresyonun Basamaklarındaki R <sup>2</sup> Değerleri			MİF Ki-Kare Testi	MİF R <sup>2</sup> Farkı ( $\Delta R^2$ )	MİF Düzeyi
		Adım1	Adım2	Adım3			
		1	S042042	0.283			
2	S042003	0.087	0.090	0.100	3.219 p=.200	0.013	-
3	S042110	0.056	0.056	0.057	0.381 p=.826	0.001	-
4	S042222C	0.121	0.129	0.132	3.019 p=.221	0.011	-
5	S042065	0.219	0.278	0.278	15.133 p=.000	0.059	<b>A</b>
6	S042280	0.399	0.399	0.399	0.076 p=.963	0.000	-
7	S042218	0.222	0.227	0.232	3.041 p=.219	0.010	-
8	S042312	0.103	0.103	0.112	2.273 p=.321	0.009	-

9	S042217	0.310	0.317	0.320	2.924	0.010	-
					p=.232		
10	S032611	0.408	0.409	0.409	0.278	0.001	-
					p=.870		
11	S032156	0.182	0.191	0.198	4.404	0.016	-
					p=.110		
12	S032087	0.078	0.078	0.078	0.126	0.000	-
					p=.939		
13	S032279	0.134	0.147	0.151	4.551	0.017	-
					p=.103		
14	S032238	0.354	0.354	0.354	0.103	0.000	-
					p=.950		
15	S032160	0.140	0.141	0.148	2.102	0.008	-
					p=.350		
16	S032654	0.180	0.189	0.190	2.747	0.010	-
					p=.253		
17	S032510	0.242	0.243	0.248	1.758	0.006	-
					p=.415		
18	S032158	0.106	0.108	0.109	0.824	0.003	-
					p=.662		

### 13. Kitapçık

Madde Numarası	Madde	$\chi^2$	$\alpha_i$	MİF Düzeyi
1	S032611	0.000	-0.030	<b>A</b>
2	S032156	0.184	0.143	<b>A</b>
3	S032087	0.180	0.126	<b>A</b>
4	S032279	0.046	-0.078	<b>A</b>
5	S032238	0.003	-0.045	<b>A</b>
6	S032160	4.651	-0.552	<b>C</b>
7	S032654	1.454	0.314	<b>A</b>
8	S032510	1.111	0.288	<b>A</b>
9	S032158	0.006	-0.010	<b>A</b>
10	S042258	0.449	0.190	<b>A</b>
11	S042016	3.180	-0.481	<b>A</b>
12	S042068	0.030	0.073	<b>A</b>
13	S042216	0.001	-0.022	<b>A</b>
14	S042249	1.481	-0.327	<b>A</b>
15	S042164	0.167	0.153	<b>A</b>

#### 14. Kitapçık

Madde numarası	Madde	$\beta$	Standart Hata	p	MİF Düzeyi
1	S032115	0.030	0.058	.606	-
2	S032403	0.042	0.059	.476	-
3	S032273	0.021	0.050	.669	-
4	S032620	0.039	0.052	.451	-
5	S042258	0.050	0.054	.357	-
6	S042016	-0.202	0.054	.000	<b>C</b>
7	S042068	0.058	0.052	.270	-
8	S042216	-0.077	0.057	.173	-
9	S042249	-0.017	0.057	.771	-
10	S042164	0.046	0.050	.357	-

### Ek-3

Sayın uzman,

“TIMSS 2007 Fen Bilimleri Testindeki Maddelerin Dil ve Cinsiyet Yanlılığı Açısından İncelenmesi” adlı tezimde içerdikte bulunan maddelerin dile göre, İngiliz öğrenciler ile karşılaştırıldığında, Türk öğrenciler lehine ölçme yaptığı bulunmuştur.

**Sizden, alan uzmanı olarak, içerikte bulunan soru veya soruların ve yanıtlarının bilimsel açıdan doğru olup olmadığının belirtmeniz beklenmektedir. Eğer soruda bilimsel anlamda bir yanlışlık/belirsizlik varsa, o halde bu yanlışlığı/belirsizliği belirtiniz.**

Katkılarınız için teşekkür ederim.

**Uzmanın,**

Adı Soyadı:

Görev Yaptığı Okul:

Uzmanlık Alanı:

Arş. Gör. H. Eren SUNA

Başkent Üniversitesi

Eğitim Bilimleri Bölümü

Ölçme ve Değerlendirme Alanı

**SORU 64**

Aşağıdaki özelliklerden hangisi SADECE memelilerde görülür?

- Ⓐ renkleri ayırt edebilen gözler
- Ⓑ süt üreten salgı bezleri
- Ⓒ oksijeni soğuran (alabilen) deri
- Ⓓ pullarla korunmuş deri

ÖĞRENME ALANI	Biyoloji
BİLİŞSEL ALAN	Bilgi
DOĞRU YANIT	B

**Yorumunuz:**



## EK-4

Sayın uzman,

“TIMSS 2007 Fen Bilimleri Testindeki Maddelerin Dil ve Cinsiyet Yanlılığı Açısından İncelenmesi” adlı tezimde içerdikte bulunan maddelerin dile göre, İngiliz öğrenciler ile karşılaştırıldığında, Türk öğrenciler lehine ölçme yaptığı bulunmuştur.

**Sizden, yabancı dil uzmanı olarak, içerikte maddenin İngilizce ve Türkçe versiyonlarının madde köklerinin ve seçeneklerinin dilbilimsel açıdan eşdeğer doğru olup olmadığının ve varsa, tercüme sürecindeki belirsizlik ve yanlışlıkları belirtmeniz beklenmektedir.**

Katkılarınız için teşekkür ederim.

**Uzmanın,**

Adı Soyadı:

Görev Yaptığı Okul:

Uzmanlık Alanı:

Arş. Gör. H. Eren SUNA

Başkent Üniversitesi

Eğitim Bilimleri Bölümü

Ölçme ve Değerlendirme Alanı

### Maddenin İngilizce Kitapçığındaki Versiyonu

Which of the following is formed immediately after fertilization?

Ⓐ egg  
Ⓑ sperm  
Ⓒ zygote  
Ⓓ embryo

Copyright  
Protected by IEA.

S042028

### Maddenin Türkçe Kitapçığındaki Versiyonu

**SORU 15)**

Döllenmeden hemen sonra aşağıdakilerden hangisi oluşur?

Ⓐ yumurta  
Ⓑ sperm  
Ⓒ zigot  
Ⓓ embriyo

Yorumunuz: