

PISA Fen Başarı Testinin Madde Yanlılığının Kültür ve Dil Açısından İncelenmesi*

An Investigation of Item Bias in PISA Science Test in Terms of The Language and Culture

Nezaket Bilge UZUN

Mersin Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Mersin, Türkiye

Selahattin GELBAL

Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitimde Ölçme Ve Değerlendirme Anabilim Dalı, Ankara, Türkiye

Makale Geliş Tarihi: 22.03.2016

Yayına Kabul Tarihi: 27.03.2017

Özet

Bu çalışmada, PISA 2006 testi maddelerinin farklı alt gruplarda DMF analizleri yürütülmüştür. DMF analizleri Mantel Haenszel (MH), lojistik regresyon (LR), Alan indeksleri yöntemleriyle gerçekleştirilmiştir. Türkiye, Avusturalya, İngiltere ve Kanada başarı testine ilişkin veriler kullanılmıştır. Uygulanan kitapçıklardan farklı kültür farklı dil gruplaması için birinci ve beşinci kitapçık; aynı kültür farklı dil ile farklı kültür aynı dil gruplaması için beşinci kitapçık seçilmiştir. DMF'nin olası nedenlerini belirlemek için alan uzmanlarının görüşlerine başvurulmuştur. Dilsel ve kültürel farklılıklar arttıkça DMF'li madde sayısının da arttığı gözlenmiştir. Kullanılan tekniğe göre DMF'li madde sayısı değişmektedir. DMF belirleme yöntemlerine göre DMF'li ve DMF'siz maddeler için bir tutarlılık bulunmamaktadır. Benzer yanlılık araştırmalarında olduğu gibi bu çalışmada da Türkçe ve İngilizce formdaki açıklanan maddelerdeki DMF'nin olası temel nedenleri; çeviri, program, kültür ve dile bağlı farklılıklar olarak belirlenmiştir.

***Anahtar Kelimeler:** Değişen madde fonksiyonu, Mantel Haenszel, Lojistik Regresyon, Alan İndeksleri Yöntemleri, PISA*

Abstract

In this study, differential item functioning (DIF) analyses of Science items of PISA 2006 tests were carried out between different samplings. Mantel Haenszel (MH), logistic regression(LR) and signed - unsigned area indexes methods were used. The research group of this study consists of the samples of Australia, Canada; England, Turkey. In order to investigate the sources of DIF field specialist opinions were consulted for released multiple choice items of science test. It is observed that as the linguistic and cultural differences increased between countries, the number of DIF items increased. The number of DIF items varied significantly according to the procedure used. There was not consistency according to DIF detecting method in DIF or non-

DIF items. Generally; like other results of bias researchs this research indicated that the main possible reasons for DIF is due to differences in translation, curriculum,cultural relevance, linguistic differences across Turkish and English versions of the tests of released items.

Keywords: *Differential Item Functioning, Mantel-Haenszel, Logistic Regression, Signed-Unsigned Area Indexes, PISA*

1. Giriş

Ulusal ve uluslararası düzeylerde farklı amaçlarla birçok sınav düzenlenmektedir. Bu sınavlara katılan bireyler çeşitli açılardan farklılaşmaktadır. Sınav dilinin değişmesi, farklı demografik özelliklere ait olma, farklı sosyo ekonomik düzeye sahip olma, cinsiyet, kültürel özellikler...vs gibi farklılıklar sınavlara katılan bireylerin performanslarını etkilemektedir (Asil ve Gelbal, 2012; Ercikan, 1998) . Bu nedenle ölçme aracından elde edilen puanların yorumlanmasında, bireylerin performanslarını etkileyebilecek değişkenler ile alt grupların özellikleri dikkate alınmalıdır.

Bir ölçme aracı geliştirildiğinde, ölçme aracından alınan puanların farklı alt gruplara bağımlı olmadığı kabul edilir. Diğer bir ifadeyle, ölçme aracını cevaplayan grup üyeleri için değişmez olması beklenir (Millsap ve Meredith, 1992; Millsap, 2005). Eğer, ölçme aracının ölçtüğü özellikler farklı gruplar için değişmez değilse, yanlılık içeriyorsa, bu ölçme aracından elde edilen verilerle yapılan karşılaştırmaların, yorumların, alınan kararların geçerliği sorgulanabilir. Bu nedenle özellikle son yıllarda “değişen madde fonksiyonu (DMF) ve madde yanlılığı” konularının araştırılmasına yoğunlaşılmaktadır. Madde yanlılığı; aynı yetenek düzeyinde olan fakat farklı gruplardan gelen iki kişinin bir maddeyi doğru cevaplama olasılıklarının aynı olmaması durumu olarak tanımlanmaktadır (Osterlind, 1983; Hambleton ve Rogers,1996; Zumbo,1999; Raju ve Ellis, 2002). Diğer bir ifadeyle; bir ölçme aracı ile ölçülen özellik bakımından aynı yetenek düzeyinde olan fakat cinsiyet, sosyoekonomik düzey, kültür...vs gibi farklı alt gruplarda yer alan bireylerin ölçme aracında yer alan sorulara verdikleri cevapların benzer olması beklenir. Madde yanlılığının araştırılması süreci hem istatistiksel yönden, hem de bu maddelerin farklılığının kaynağı açısından ne olduğunun araştırılmasını içerir. Farklılığın sadece istatistiksel olarak gösterildiği duruma “değişen madde fonksiyonu (DMF)” adı verilir. DMF’nin ortaya çıkmasının iki nedeni vardır. Bunlar; alt gruplar arasındaki gerçek farklılık ve madde yanlılığıdır (Camilli ve Shepard,1994). Her zaman bir testte bir grubun puanının başka bir grubun puanından farklı olması testin ya da incelenen maddelerin yanlı olduğu anlamına gelmez. Örneğin ergenlik döneminde cinsiyet grupları arasındaki gelişimsel farklılıklar, cinsiyet bağlamına ilişkin büyüme eğrileri arasındaki farklılıklar gerçek farklılıklardır. Yanlılık genel olarak DMF sonuçlarına dayalı olarak değerlendirilmektedir. Zumbo (1999, s:6) DMF’yi yanlılık analizlerindeki yeni standart olarak belirtmiştir.

PISA, TIMSS, PIRLS gibi kültürler-arası çalışmalar, ülkelerdeki eğitim sistemini değerlendirmek ve geliştirmek için geniş bakış açıları sağlamaktadır. Büyük çaplı sı-

nav verileri kullanılarak yapılan yanlılık çalışmalarının, ülkeler arası karşılaştırmaların anlamlılığı bakımından önemli bir ihtiyaç olduğu düşünülmektedir. Bilindiği üzere bu sınavların her uygulama döneminde farklı bir odak alanı bulunmaktadır. PISA 2006 yılının temel odağı fen bilimleridir. Bu çalışmanın amacı, PISA 2006 fen başarı testinde yer alan maddelerin aynı kültür farklı dil, farklı kültür aynı dil ve farklı kültür farklı dil çerçevesinde değişen madde fonksiyonunun farklı teknikler ile analiz edilmesi ve karşılaştırılmasıdır. Genel olarak madde yanlılığı belirleme teknikleri, klasik test kuramına (KTK) dayalı teknikler ve madde tepki kuramına (MTK) dayalı teknikler olarak sınıflandırılabilir. Çalışmanın temel amacı; klasik test kuramı ve madde tepki kuramı dahilinde DMF belirlemek için kullanılacak tekniklerden elde edilen sonuçların benzerlik ve farklılıklarının ortaya çıkarılması ve teknikler arasındaki uyumun belirlenmesidir. Diğer bir amaç ise değişen madde fonksiyonu tespit edilen maddelerin olası DMF sebeplerinin araştırılmasıdır. Özellikle uluslararası nitelikte olan uygulamalarda kullanılan ölçme araçlarının farklı alt gruplarda aynı özelliği ölçmesi yapılan karşılaştırmaların anlamlı olmasına olanak verir. Bu nedenle yapılan karşılaştırmalı çalışmaların anlamlılığı açısından da özellikle geniş ölçekli ölçme araçlarında DMF ve olası DMF nedenlerinin araştırılması önemlidir.

2. Yöntem

Bu araştırma; PISA bilişsel alanda yer alan maddelerin kültüre ve dile göre yanlılığının farklı tekniklerle kültürler arası karşılaştırmalı incelenmesine yönelik tasarlanmış betimsel bir çalışmadır.

Çalışma grubunu 2006 yılında yapılan PISA araştırma taramasına katılan Türkiye, Avustralya, İngiltere, Kanada öğrenci örneklemini oluşturmaktadır. Çalışma grubu oluşturulurken dil ve kültüre göre yapılabilecek gruplamalar temel alınmıştır. Farklı kültürlerle sahip olup PISA kitapçığını aynı dilde yanıtlayan grubu Avustralya ve Kanada; aynı kültürün bir parçası olup; soru kitapçığını farklı dillerde yanıtlanabildiği grubu Kanada; hem kültürün hem de kitapçık dilinin değiştiği grubu ise Türkiye ve İngiltere oluşturmaktadır. Aynı kültür farklı dil gruplamasında Kanada beşinci kitapçık verileri; farklı kültür aynı dil gruplaması için Avustralya ve İngiltere beşinci kitapçık verileri; yanlılık çalışmalarının da yürütüldüğü farklı kültür aynı dil için Türkiye ve İngiltere birinci ve beşinci kitapçık verileri kullanılmıştır. Çalışmada kullanılan veriler Tablo 1’de gösterildiği gibi oluşturulmuştur.

Tablo 1: PISA’ya Katılan ve Beşinci Kitapçığı Cevaplayan Öğrenci Sayılarına Ait İstatistikler

Ülke	Sınavı Giren Öğrenci Sayısı	5. kitapçığı alan öğrenci sayısı	1. kitapçığı alan öğrenci sayısı
FARKLI KÜLTÜR AYNI DİL			
Avustralya	14.170	1124	
Kanada	22.646	1744	

Ülke	Sınava Giren Öğrenci Sayısı	5. kitapçığı alan öğrenci sayısı	1. kitapçığı alan öğrenci sayısı
AYNI KÜLTÜR FARKLI DİL			
Kanada (İngilizce form)	17.555	1352	
Kanada (Fransızca form)	5.091	392	
FARKLI KÜLTÜR FARKLI DİL			
İngiltere	13.152	1008	1430
Türkiye	4.942	377	380

Tablo 1'deki istatistikler incelendiğinde, araştırmada karşılaştırılan gruplar için kullanılan örneklemelerin büyüklüğü 377 ile 1744 arasında değişmektedir. 13 kitapçık türünden birinci ve beşinci kitapçık tüm çoktan seçmeli soruların fen okur-yazarlığı ile ilgili olması ve en çok çoktan seçmeli soru sayısını barındırması bakımından çalışılacak kitapçıklar olarak belirlenmiştir. Birinci kitapçıkta yer alan 19 çoktan seçmeli soru ve beşinci kitapçıkta yer alan 21 çoktan seçmeli soru ile çalışma yürütülmüştür. PISA 2006 raporunda 13 kitapçıkta yer alan çoktan seçmeli sorulardan sadece 10'u açıklanmıştır. Bu nedenle farklı yöntemlerle yapılan DMF analizleri iki kitapçıkta yer alan toplam 37 çoktan seçmeli soru üzerinden, DMF'nin olası nedenlerinin araştırılması ise açıklanan 10 soru üzerinden yürütülmüştür. Çalışmada Türkçe ve İngilizce form için açıklanan maddelerdeki olası DMF nedenleri uzman görüşlerine başvurulmuş ve değerlendirilmiştir. Uzmanlara Türkçe ve İngilizce formda yer alan maddelerin gruplardan (Türkiye ve İngiltere) herhangi birinin lehine çalışıp çalışmadığı; çalıştığını düşünmesi halinde bunun olası nedenini belirtmesi istenmiştir. Çalışmada 15 uzmana ulaşılarak; açıklanan maddelere ilişkin görüşleri alınmış ve uzmanlarca bildirilen olası yanlışlık nedenleri bulgular bölümünde özetlenmiştir. Bu formu yanıtlayan uzmanların fen bilimleri alanında mezun olmaları (fen bilimleri alanında öğretmenlik yapmaları) ve iyi derecede İngilizce bilmeleri dikkate alınmıştır.

Veri dağılımı ve varsayımların sağlanıp sağlanmamasına yönelik çalışmaların ardından DMF analizlerine geçilmiştir. Bu çalışmada KTK teknikleri içerisinde yer alan Mantel-Haenszel (MH) tekniği, Lojistik regresyon (LR) tekniği ve MTK kapsamında yer alan madde karakteristik eğrileri arasındaki alanlar hesaplanarak elde edilen alan indeksleri tekniğine bağlı değişen madde fonksiyonu belirleme teknikleri kullanılmıştır.

KTK kapsamında DMF belirlemede kullanılan Mantel-Haenszel (MH) yaklaşımı bir ki-kare tekniğidir. Yöntemde odak ve referans gruplarının her ikisi de kendi içinde yanıtlayıcıların yeteneklerini eşitlemek için iki ya da beş alt gruba bölünür. Denkleştirilmiş iki gruptaki performans olasılık oran değerleri dikkate alınarak karşılaştırılır. Olasılık oranının daha kolay yorumlanabilmesi için bu oran değeri "delta ölçeği" adı verilen bir ölçek değerine dönüştürülür. Hesaplanan değer, 0'a eşitse maddede DMF yoktur. Hesaplanan değer 0'dan büyükse madde odak grup lehine, 0'dan küçükse referans grup lehine çalışmış olduğuna karar verilir. Lojistik regresyon analizleri odak ve referans gruptaki yanıtlayıcıların belli bir ölçüm kriterleri çerçevesinde bir mad-

deye doğru cevap verme olasılığını hesaplamaya dayanır. Madde bağımlı değişken olarak düşünülür. Bağımlı değişken test maddesinin 0-1 şeklinde kodladığı değerden oluşur. Odak ve referans grubun tanımlanması bağımsız değişkendir. Bağımsız değişken ise ikili veya sürekli veri niteliğindedir. Puanlar(madde puanları) üzerinde grup etkisinin anlamlı olup olmadığına bakılır. Eğer madde grup etkileşimi anlamlı ise bu maddenin yanlı olabileceğine karar verilir. MTK yöntemlerinde aynı yetenek düzeyine sahip odak ve referans grubun bir maddeye doğru cevap verme olasılıkları arasındaki farklar incelenir. Başka bir ifade ile MTK ile yapılan alan indeksleri yönteminde; alt gruplardan (odak ve referans grup) elde edilen madde parametreleri veya madde karakteristik eğrileri karşılaştırılır. Madde karakteristik eğrilerinin aynı düzlem üzerindeki görünümünün incelenerek yanlılık hakkında kanıt aranır. Bu yöntemle madde yanlılığı belirleme yöntemine Raju'nun alan ölçümü denilmektedir. Aynı düzlem üzerine konulan madde karakteristik eğrileri eğer çakışır veya madde karakteristik eğrileri arasındaki alanın ölçümü 0 olursa, yanlılığın olmadığına karar verilir. Madde karakteristik eğrileri arasındaki alanın büyüklüğü sıfırdan uzaklaştıkça, maddenin yanlılığı da artmaktadır (aktaran: Devenci, 2008; Lord, 1980; Raju, 1988). MH ve LR analizi için EZDIFF programı; alan indeksleri tekniği için BILOG OLD ve Frans Oort tarafından geliştirilen CALCBIAS programı kullanılmıştır. Verilerin analizlere uygunluğunu değerlendirebilmek için öncelikle kullanılan PISA-2006 verilerine ait betimsel istatistikler hesaplanmıştır. Elde edilen sonuçlar tablo 2'de verilmiştir.

Tablo 2: Çalışma Grubuna İlişkin Betimsel İstatistikler

	Ülke adı	N	Ortalama	Basıklık	Çarpıklık	S.S	KR-20
5.Kitapçık	AVUSTRALYA	1067	13,98	-,457	-,495	4,13	0,78
	KANADA	1647	13,78	-,664	-,421	4,25	0,79
	TÜRKİYE	333	10,38	-,568	,326	4,08	0,74
	İNGİLTERE	968	12,84	-,927	-,188	4,45	0,81
	TOPLAM	3829	13,35	-,781	-,328	4,36	0,80
1.Kitapçık	TÜRKİYE	365	9,99	-,728	,288	3,84	0,74
	İNGİLTERE	790	12,83	-,563	-,421	3,82	0,79
	TOPLAM	1158	11,93	-,871	-,195	4,05	0,80

Tablo 2'deki dağılımların biçimi, çarpıklık ve basıklık değerleri sıfır değerine yakın olduğundan; dağılımların gruplar için simetrik ve normale yakın olduğu kabul edilebilir. Testin hesaplanan güvenilirlik katsayısının alt gruplar için 0,74 ile 0,81 arasında değiştiği, kullanılan verilerin güvenilir olduğu söylenebilir.

MTK'nın en önemli sayıtlarından biri testlerin tek boyutlu bir niteliği ölçmesidir (Hambleton ve Swaminathan, 1985). Bu sayıtlı madde setine uygulanan temel bileşenler analizi ile özdeğerlerin incelenmesi yoluyla test edilebilir. Bu nedenle tek boyutluluk sayıtlısını kontrol etmek için, PISA'ya ait veriler faktör analiziyle incelenmiştir. Yapılan analiz sonucunda tüm gruptan elde edilen özdeğerler dikkate alınmıştır. Tek boyutluluk aynı zamanda testteki maddelerin yerel bağımsızlığı için kanıt

olarak kullanılmıştır. Birinci ve beşinci kitapçıkta yer alan 19 ve 21 soruluk testin tek boyutluluk ve yerel bağımsızlık değerlendirmeleri için özdeğerlere ilişkin faktörler ile açıkladığı varyans oranları Tablo 3'te gösterilmiştir.

Tablo 3: PISA-2006 Beşinci Kitapçıkta Yeralan 21 ve Birinci Kitapçıkta Yeralan 19 Soruluk Teste Ait Faktör Analizi Sonuçları

		Özdeğer	Açıklanan varyans	
5. kitapçık	TÜM GRUP	1. Faktör	4,37	20,811
		2. Faktör	1,023	4,872
		3. Faktör	1	4,761
	AVUSTRALYA	1. Faktör	4,152	19,773
		2. Faktör	1,11	5,285
		3. Faktör	1,101	5,243
		4. Faktör	1,043	4,98
	KANADA	1. Faktör	4,301	20,481
		2. Faktör	1,065	5,072
		3. Faktör	1,013	4,825
		4. Faktör	1,01	4,81
	TÜRKİYE	1. Faktör	3,524	16,782
		2. Faktör	1,4	6,665
		3. Faktör	1,265	6,023
		4. Faktör	1,153	5,491
	İNGİLTERE	1. Faktör	4,519	21,517
2. Faktör		1,151	5,481	
3. Faktör		1,094	5,21	
4. Faktör		1,044	4,969	
1. kitapçık	TÜM GRUP	1. Faktör	4,199	22,102
		2. Faktör	1,098	5,778
		3. Faktör	1,036	5,452
	İNGİLTERE	1. Faktör	4,153	21,855
		2. Faktör	1,177	6,196
		3. Faktör	1,046	5,507
	TÜRKİYE	1. Faktör	3,535	18,605
		2. Faktör	1,279	6,732
		3. Faktör	1,186	6,242
		4. Faktör	1,136	5,979

Tablo 3 incelendiğinde beşinci kitapçığı yanıtlayan Avustralya, Kanada, Türkiye ve İngiltere veri grubun oluşturduğu 21 soruluk teste ait belirlenen ilk dört faktöre ait özdeğerler ve açıkladıkları toplam varyans tüm grup ve alt gruplar için yapılan faktör analizi sonuçlarına göre tüm alt gruplar için ilk boyuttaki özdeğerlerin ikinci boyuttan büyük olduğu, ikinci ve üçüncü boyutlardaki özdeğerlerin büyük ölçüde benzer büyüklükte olduğu bulunmuştur. Benzer bulgu birinci kitapçığı alan alt gruplar için de geçerlidir. Yapılan faktör analizi sonucu incelendiğinde birinci ve ikinci faktör arasın-

daki keskin düşüşler tek boyutluluk için kanıt olarak kabul edilmiştir. Tek boyutluluk aynı zamanda testteki maddelerin yerel bağımsızlığı için de kanıt olarak kullanılmıştır (Lord ve Novic 1968; Hambleton ve Swaminathan 1985; Baker 2001).

Her model için yetenek kestirimlerine yönelik elde edilen $-2 \log$ (olabilirlik) değerlerinin farkını almak suretiyle bir uyum iyiliği testi (" χ^2 " Kay Kare) yapılarak her iki kitapçık verileri, parametrelerinin kestirimlerinde üç parametrelili lojistik modelin uyumlu olduğu sonucuna ulaşılmıştır. MTK'ya bağlı alan indeksleri tekniği ile yapılan çalışmalar incelendiğinde DMF içeren maddeye karar verilirken iç ölçütlerden faydalandığı belirlenmiştir. Bu ölçütler, işaretli ve işaretsiz alan ölçümlerinin ortalaması, medyan değerleri, bu değerlerin bir standart sapma veya çeyrek sapma üstü ile alan ölçümlerinin z değerlerinin, anlamlılığı biçiminde listelenebilir (Gondal; 2001; Öğretmen ve Doğan, 2004; Deveci, 2008). Bu çalışmada iç ölçüt olarak maddelerin alan indeksleri değerlerine ait olan medyan değerlerinin 1 çeyrek sapma üstü kullanılmıştır. Hesaplanan iç ölçütler çalışma grubunda belirtilen her bir grupta için tablo 4'te yer almaktadır. Alan indeksleri tekniğiyle maddelerin DMF içerip içermediği medyan+ 1 çeyrek sapma değeri temel alınarak değerlendirilmiştir.

Tablo 4: Alan İndeksleri Tekniğinin Yorumlanmasında Kullanılan İç Ölçülere İlişkin İstatistikler

	İŞARETLİ ALAN İNDEKSİ			İŞARETSİZ ALAN İNDEKSİ		
	Medyan	Çeyrek Sapma	Ç.Sapma +Medyan	Medyan	Çeyrek Sapma	Ç.Sapma +Medyan
Aynı Kültür- Farklı Dil	0,283	0,256	0,539	0,594	0,658	1,251
Farklı Kültür Aynı Dil	-0,305	-0,397	-0,702	0,501	0,486	0,987
Farklı Kültür Farklı Dil						
1.KİTAPÇIK	1,020	1,311	2,331	2,823	2,832	5,655
5.KİTAPÇIK	0,709	0,606	1,315	1,158	1,508	2,667

3. Bulgular

Bu bölümde üç farklı alt grupta için (aynı kültür farklı dil: Kanada İngilizce form – Kanada Fransızca form örneği; farklı kültür aynı dil Avustralya - İngiltere örneği; farklı kültür farklı dil Türkiye- İngiltere örneği) MH, LR ve alan indeksleri tekniği ile DMF içeren maddeler belirlenmiştir. Tablo5 'de üç grup için yapılan MH sonuçları bulunmaktadır.

Tablo 5: PISA 2006 Birinci ve Beşinci Kitapçık Alt Testinde Mantel – Haenszel Tekniği ile Değişen Madde Fonksiyonu Bulunan Maddelere İlişkin Analiz Sonuçları

	Kitapçık-Madde	A	χ^2	p değeri	Δ MH	SE Δ MH	DMF düzeyi
Aynı Kültür Farklı Dil	5-3	1.483	7.421	0.006	-0.927	0.332	A
	5-6	1.332	4.537	0.033	-0.674	0.310	A
	5-11	1.830	16.179	0.000	-1.420	0.353	B
	5-17	1.375	5.850	0.016	-0.749	0.302	A
	5-18	0.597	9.581	0.002	1.212	0.385	B
	5-19	0.570	11.175	0.001	1.320	0.380	B
Farklı Kültür Aynı Dil	5-1	0.733	5.962	0.015	0.729	0.292	A
	5-15	1.331	6.474	0.011	-0.672	0.259	A
	5-16	1.308	7.545	0.006	-0.630	0.226	A
	5-18	0.491	30.629	0.000	1.671	0.302	C
	5-19	0.645	13.414	0.000	1.032	0.278	B
	5-20	1.467	15.068	0.000	-0.901	0.229	A
Farklı Kültür- Farklı Dil (Türkiye-İngiltere)	1-1	1.504	6.378	0.012	-0.959	0.371	A
	1-3	1.384	4.406	0.036	-0.702	0.330	A
	1-4	0.578	11.819	0.001	1.289	0.371	B
	1-5	3.118	51.650	0.000	-2.673	0.380	C
	1-8	2.984	48.883	0.000	-2.569	0.380	C
	1-10	1.440	6.151	0.013	-0.857	0.366	A
	1-13	1.643	9.110	0.003	-1.167	0.375	A
	1-14	3.289	57.521	0.000	-2.798	0.381	C
	1-15	0.679	6.349	0.012	0.911	0.350	A
	1-18	1.756	15.712	0.562	-1.323	0.327	B
	5-1	1.393	4.274	0.039	-0.778	0.364	A
	5-2	3.589	52.415	0.000	-3.003	0.420	C
	5-3	1.459	6.482	0.011	-0.888	0.341	A
	5-4	1.863	16.191	0.000	-1.463	0.362	B
	5-6	0.637	8.530	0.003	1.062	0.353	B
	5-9	0.397	35.051	0.000	2.169	0.365	C
	5-12	0.490	25.284	0.000	1.675	0.337	C
	5-16	0.520	20.342	0.000	1.537	0.340	B
	5-17	0.491	24.751	0.000	1.671	0.333	C
5-18	2.487	35.744	0.000	-2.141	0.362	C	
5-19	1.537	8.025	0.005	-1.011	0.351	B	

Tablo 5’de belirtilen MH sonuçlarına göre; aynı kültür farklı dil gruplaması için beşinci kitapçıkta yer alan 21 maddenin altısı DMF içermektedir. Bu maddelerin üçü önemsenmeyecek düzeyde (A) diğer üçü de orta düzeydedir (B). 3, 6, 11 ve 17. maddeler referans grubun (dili İngilizce olanlar) lehine, 18. ve 19. maddelerin ise odak grubun (dili Fransızca olanlar) lehine DMF içermektedir. Farklı kültür aynı dil gruplaması için 6 madde DMF içermektedir. Bu maddelerin dördünün A düzeyinde, biri B, diğeri de C düzeyindedir. MH analiz sonuçları incelendiğinde; 1., 18. ve 19. maddelerin İngiltere’nin lehine DMF içerdiği, 15., 16. ve 20. maddelerin ise referans grubun (Avustralya) lehine DMF içerdiği görülmektedir. Farklı kültür farklı dil gruplaması için birinci kitapçıkta 19 maddeden 10’u DMF içermektedir. Maddelerin beşi A düzeyinde, ikisi B, üçü ise C düzeyindedir. MH sonuçlarına göre; 4. ve 15. Maddeler odak grubun (Türkiye) lehine, DMF gösteren diğer maddelerin ise referans grubun (İngiltere) lehinedir. Beşinci kitapçıkta yer alan 21 maddeden 11 maddenin farklı kültür ve dil için DMF içerdiği bulunmuştur. Bu maddelerin ikisi A, dördü B, beşi de C düzeyinde DMF içermektedir. MH analiz sonuçları incelendiğinde; 1., 2., 3., 4., 18. ve 19. maddelerin referans grubun (İngiltere) lehine; 6., 9., 12., 16. ve 17. maddelerin ise odak grubun (Türkiye) lehinedir. Zieky (2003, s:12) ön deneme uygulaması mümkün olmayan başarı ölçümlerinde “B” düzeyinde bulunan DMF’nin ihmal edilebileceğini belirtmiştir. Bu çerçeveden bakıldığında DMF içeren ve yanlılığın araştırılabileceği madde sayısı farklı kültür aynı dil için bir iken; farklı kültür farklı dil gruplamasında birinci kitapçıkta üç; beşinci kitapçıkta ise sekizdir. Elde edilen bulgulardan da görüldüğü üzere gruplar arası farklılıklar arttıkça DMF gösteren madde sayısı da artmaktadır. Üç farklı gruplama için yapılan LR analizi sonuçları tablo 6’da verilmiştir.

Tablo 6: PISA 2006 Birinci Ve Beşinci Kitapçıkta Logistik Regresyon Tekniği İle Değişen Madde Fonksiyonu Bulunan Maddelere İlişkin Analiz Sonuçlar

Gruplar	Kitapçık-Madde	Grup etkisi (p)	Grup*Toplam puan etkisi (p)	ΔR^2	DMF Türü	DMF düzeyi
Aynı Kültür Farklı Dil	5-4	0.042	0.014	0.001	ÇBDMF	A
	5-19	0.215	0.021	0.009	ÇBDMF	A
	5-20	0.028	0.025	0.000	ÇBDMF	A
Farklı Kültür Aynı Dil	5-1	0.018	0.141	0.005	TBDMF	A
	5-3	0.008	0.033	0.001	ÇBDMF	A
	5-16	0.004	0.022	0.004	ÇBDMF	A
	5-18	0.003	0.282	0.024	TBDMF	A
Farklı Kültür Farklı Dil	1-3	0.005	0.0003	0.000	ÇBDMF	A
	1-4	0.0000	0.00008	0.032	ÇBDMF	A
	1-5	0.861	0.031	0.030	ÇBDMF	A
	1-8	0.812	0.006	0.041	ÇBDMF	B
	1-15	0.101	0.006	0.022	ÇBDMF	A
	1-17	0.003	0.0077	0.007	ÇBDMF	A

Gruplar	Kitapçık-Madde	Grup etkisi (p)	Grup*Toplam puan etkisi (p)	ΔR^2	DMF Türü	DMF düzeyi
Farklı Kültür Farklı Dil	5-2	0.004	0.767	0.069	TBDMF	B
	5-4	0.210	0.011	0.014	ÇBDMF	A
	5-11	0.091	0.039	0.001	ÇBDMF	B
	5-12	0.000005	0.008	0.029	ÇBDMF	A
	5-14	0.002	0.008	0.001	ÇBDMF	A
	5-16	0.001	0.169	0.022	TBDMF	A

Kanada alt gruplarının beşinci kitapçıkta yer alan 21 fen maddesinden lojistik regresyon analizi sonuçlarına göre üç maddede DMF olduğu görülmektedir. Bu maddelerin tümünde grup madde etkileşimi manidar bulunmuştur. Bu; maddelerin tümünde çok biçimli DMF bulunduğu bir göstergesidir. Bir diğer ifadeyle; yetenek düzeyi ile grup kategorileri arasında bir kesişim söz konusudur. Gruplar arası başarı olasılığı tüm yetenek düzeylerinde sabit kalmamaktadır. DMF gösteren maddelerin DMF düzeyleri ise Gierl ve arkadaşlarının (2000) belirlediği ölçütlere göre A düzeyinde bulunmuştur. Farklı kültür aynı dil gruplaması için 21 maddeden 4 maddenin DMF içerdiği sonucuna ulaşılmıştır. LR analizi sonuçlarına göre 1. ve 18. madde tek biçimli değişen madde fonksiyonu (TBDMF) içermektedir. TBDMF’de gruplar arası başarı olasılığı tüm yetenek düzeylerinde sabit kalmaktadır ; .3. ve 16. madde ise çok biçimli değişen madde fonksiyonu (ÇBDMF) içermektedir. DMF gösteren dört maddenin de DMF düzeyleri Gierl ve arkadaşlarının belirlediği ölçütler çerçevesinde önemli bulunmamıştır. Farklı kültür farklı dil gruplaması için birinci kitapçıkta altı madde ÇBDMF içermektedir. Bu altı maddenin düzeyleri Beşinci kitapçıkta ise altı madde DMF içermektedir. Bunlardan 2. ve 16. maddeler TBDMF gösterirken diğer dördü ÇBDMF göstermektedir. Altı maddenin dördünün önemsenmeyecek düzeyde, ikisinin ise orta düzeyde DMF içerdiği saptanmıştır. Alan indeksleri için Tablo 4’e göre DMF kararı verilirken kullanılacak iç ölçütler temel alınarak yapılan DMF analiz sonuçları tablo 7’deki gibidir.

Tablo 7: PISA 2006 Birinci ve Beşinci Kitapçık Testinde Alan İndeksleri Tekniği ile Değişen Madde Fonksiyonu Bulunan Maddelere İlişkin Analiz Sonuçlar

	Kitapçık-Madde No	İşaretili alan indeksleri	İşaretsiz alan indeksleri	DMF Biçimi
Aynı Kültür Farklı Dil	5-2	3,3	3,69	ÇBDMF
	5-3	0,63	0,63	TBDMF
	5-4	0,45	1,68	ÇBDMF
	5-6	0,28	2,12	ÇBDMF
	5-8	0,54	0,60	ÇBDMF
	5-16	0,55	0,56	TBDMF
	5-17	1,05	1,67	ÇBDMF
	5-20	0,2	2,92	ÇBDMF

	Kitapçık-Madde No	İşaretli alan indeksleri	İşaretsiz alan indeksleri	DMF Biçimi
Farklı Kültür Aynı Dil	5-2	-4,24	5,64	ÇBDMF
	5-16	-0,37	3,36	ÇBDMF
	5-20	-0,8	0,80	TBDMF
Farklı Kültür Farklı Dil	1-1	-24	30,95	ÇBDMF
	1-2	34,84	39,89	ÇBDMF
	1-3	2,68	3,00	ÇBDMF
	1-17	34,69	39,51	ÇBDMF
	1-18	35,92	40,59	ÇBDMF
	1-19	35,97	40,98	ÇBDMF
	5-2	2,82	2,82	TBDMF
Farklı Kültür Farklı Dil	5-4	2,02	2,74	ÇBDMF
	5-12	0,1	3,57	ÇBDMF
	5-14	3,05	8,68	ÇBDMF
	5-16	0,22	3,17	ÇBDMF
	5-18	1,46	1,52	ÇBDMF

Tablo 7 incelendiğinde aynı kültür farklı dil işaretli alan indeksleri için, 2, 3, 8, 16 ve 17 nolu maddelerin; işaretsiz alan indeksleri için 2, 4, 6, 17 ve 20 nolu maddelerin DMF gösterdiği anlaşılmaktadır. İşaretli alan indekslerinde DMF belirlemek için kullanılan ölçüte göre DMF gösterdiği belirlenen tüm maddelerin işareti pozitifdir. Pozitif işaretli olan maddeler referans grubun (dili İngilizce olanlar) lehine DMF içermektedir. Raju (1990) ve Gondal (2001) çalışmalarında işaretsiz alan indeksinin, işaretli alan indeksinden oldukça büyük olduğu durumlarda, DMF ortaya çıktığını ve işaretsiz alan indekslerinin maddelerde DMF'nin hangi grup lehine olduğu hakkında bilgi vermediklerini ve Çok Biçimli Değişen Madde Fonksiyonunu (ÇBDMF) belirlemede, işaretsiz alan indekslerinin etkisinin daha büyük olduğunu belirtmişlerdir. DMF'nin araştırıldığı grupların madde karakteristik eğrileri farklılık gösterirse, madde DMF içerir. Bu maddelerden Gondol ve Raju'nun açıklamaları doğrultusunda alan indekslerinin birbirlerine göre durumları incelendiğinde 3. ve 16. madde dışındaki tüm maddelerde işaretli alan indeksi değerinin işaretsiz alan indeksi değerinden küçük olduğu; dolayısıyla bu maddelerin ÇBDMF'ye sahip olduğu sonucuna ulaşılmıştır. Farklı kültür aynı dil örneklemin için tablo 7'de yer alan işaretli alan indeksleri yöntemleri sonuçlarına göre; beşinci kitapçıkta yeralan 21 maddeden iki madde DMF içermektedir. Bu maddelerin her ikisinin de işareti negatiftir. Negatif işaretli olan bu maddeler odak grubun (İngiltere) lehine DMF içermektedir. İşaretsiz alan indeksleri için hesaplanan 0,987 değeri ölçüt olarak alındığında, 2 ve 16. maddede DMF olduğu sonucuna ulaşılmaktadır. Farklı kültür farklı dil için yapılan alan indeksleri sonuçları incelendiğinde; birinci kitapçıkta yeralan 19 maddeden altı madde DMF içermektedir. İşaretli alan indeksleri için DMF gösteren tüm maddelerin işareti pozitifdir. İşareti pozitif olan bu maddeler referans grubun (İngiltere) lehine DMF içermektedir. Birinci kitapçık için işaretsiz alan indekslerine göre 5 maddenin DMF içerdiği sonucuna ulaşılmaktadır. Bu maddelerin hepsi işaretli alan indekslerine göre de yanlı bulunan maddelerdir. İşaretsiz alan indeksleri büyüklükleri 3'ün üstünde ise sıra dışı değerler

olarak kabul edilmekte ve bu değerler yanlılığın en güçlü kanıtı olarak kabul edilmektedir (Raju, 1990). Bu açıklama doğrultusunda 3. madde dışındaki tüm maddelerin bu tanımlamaya uyduğu ve DMF gösteren tüm maddelerin ÇBDMF gösterdiği belirlenmiştir. Beşinci kitapçığa ait olan işaretli alan indeksleri değerleri incelendiğinde 21 maddeden dördünde DMF olduğu sonucuna ulaşılmıştır, bu dört maddenin hepsi İngiltere lehine DMF göstermektedir. İşaretsiz alan indeksleri değerlerine göre beş maddede DMF bulunmaktadır. İşaretsiz alan indeksleri ve işaretli alan indeksleri değerlerinin birbirlerine göre durumu incelendiğinde beşinci kitapçıkta yeralan ve DMF gösteren maddelerden 2. madde dışındaki tüm maddelerin ÇBDMF gösterdiği sonucuna ulaşılmıştır. Elde edilen bulgular incelendiğinde; karşılaştırılan alt gruplar arasındaki farklılıklar arttıkça DMF gösteren madde sayıları ve düzeylerinin arttığı sonucuna ulaşılmıştır. Bu sonuç, Atalay (2010)'ın ve Asil (2010)'in yapmış oldukları çalışmalarda dilsel ve kültürel farklılıklar artıkça DMF gösteren maddelerin arttığı sonucu ile tutarlılık göstermektedir.

Yapılan DMF analizleri gibi elde edilen değerler arasındaki uyum 3 grup için ayrı ayrı incelenmiştir. Bu amaçla LR için etki büyüklüğü değerleri; MH için hesaplanan ki-kare değerleri, işaretli alan indeksleri ve işaretsiz alan indeksleri ile elde edilen değerler arasındaki sıra farkları korelasyonu incelenmiştir. Elde edilen bulgular tablo 8'de yer almaktadır.

Tablo 8: Mantel – Haenszel, Logistik Regresyon ve Alan İndeksleri Tekniği İle Elde Edilen Değişen Madde Fonksiyonu Değerleri Arasındaki Sıra Farkları Korelasyonları

		MH	LR	İAİ	İSAİ
Aynı Kültür Farklı Dil	MH	1	0,681**	0,11	0,19
	LR		1	-0,1	0,02
	İAİ			1	0,642**
Farklı Kültür Aynı Dil	MH	1	0,538*	-0,02	0,39
	LR		1	0,35	-0,11
	İAİ			1	-0,65
Farklı Kültür Farklı Dil (1. Kitapçık)	MH	1	0,45	-0,33	0,01
	LR		1	-0,21	0,05
	İAİ			1	0,491*
Farklı Kültür Farklı Dil (5.Kitapçık)	MH	1	0,891**	-0,05	0,27
	LR		1	-0,07	0,11
	İAİ			1	0,43

Tablo 8'de yeralan bulgulara göre: MH; LR ve Alan indeksleri ile elde edilen DMF değerleri arasındaki sıra farkları korelasyonları; aynı kültür farklı dil için MH ile LR ve İşaretsiz alan indeksleri ile İşaretsiz alan indeksleri değerleri arasındaki sıra farkları korelasyon katsayıları 0,01 düzeyinde; farklı kültür aynı dil için MH ile LR arasındaki sıra farkları korelasyon katsayısı 0,05 düzeyinde farklı kültür farklı dil (1. kitapçık) için işaretli alan indeksleri ile işaretsiz alan indeksleri arasındaki sıra farkları

korelasyon katsayısı 0,05 düzeyinde ve son olarak farklı kültür farklı dil (5. kitapçık) için MH ile LR arasındaki sıra farkları korelasyon katsayısı 0,01 düzeyinde anlamlı bulunmuştur. Bu bulgulardan yola çıkarak, madde yanlılığı tespitinde kullanılan farklı metotların sonuçları arasında uyumlu sonuçlar bulunamamıştır. Ancak KTK'ya ait tekniklerle MTK'ya bağlı tekniklerin kendi aralarında uyumlu olduğu sonucuna ulaşılmıştır.

DMF'nin olası nedenleri araştırılırken şu iki sınırlık göz ardı edilmemelidir. Birincisi belirlenen yanlılık sebepleri belirgin olmamasına rağmen gerçek yanlılık sebebi olarak addedilmemelidir. İkincisi ise bu nedenler o maddenin bir grubun lehine olma durumunu açıkça göstermez. Uzman görüşleri madde içeriği ya da kapsam alanı ile ilgili DMF'ye bağlı sebepler hakkında genel bir bakış sağlayabilir ancak söz konusu olan bu bakış açısının farklı alt gruplarda “niçin” veya “nasıl” farklı olduğunu açıklamak için yeterli olmayacaktır. Ayrıca giriş bölümünde de değinildiği gibi yanlılık çalışmalarında DMF gerek şart olarak gösterilir. Ancak DMF belirlenen her madde “yanlıdır” denemez. Bu maddenin karşılaştırılan gruplardan birindeki gerçek farklılık olarak ta değerlendirilebilir.

Türkçe ve İngilizce test formlarında açıklanan PISA maddelerini inceleyerek görüşlerini belirten 15 uzman yanlı olduğunu düşündüğü maddeler için; olası yanlılık sebepleri olarak; çeviri, müfredat farklılıkları ve uyarlama yanlışlıkları sonucu metnin ağır bir dile dönüşmesi, sosyo ekonomik düzeyden kaynaklanan farklılıklar olduğunu belirtmişlerdir. Açıklanan çoktan seçmeli maddelerin karmaşık ve uzun metinlerle ilişkilendirilmesinin özellikle Türk öğrencilerin alışkın olmadığı bir formata dönüşmesinin bazı maddelerin yanıtlanmasını zorlaştırdığı fikrinin hakim olduğu sonucuna ulaşılmıştır. Açıklanan maddelerden birinci ve beşinci kitapçıkta yer alan ve yapılan DMF analizleri sonucu DMF içerdiği belirlenen maddelere ait uzman görüşlerine aşağıda yer verilmiştir.

Uzmanlar; 1. ve 5. kitapçıkta yer alan 1 nolu maddenin; laboratuvar koşullarının Türkiye’de yeterince karşılanmadığı, bir çok okulda yeterince deney malzemesine ulaşılamadığı...vs gibi sosyo ekonomik düzeyle bağlantılı olan sebepler, teknolojik farklılıklar, bilim ve teknolojiadaki gelişmeleri takip eden birey sayısının azlığını olası yanlılık nedeni olarak belirtilmiştir. Bu maddenin İngiltere lehine yanlı olduğu konusunda görüş bildiren uzmanlardan bazıları maddeye ait olan okuma parçasının Türk öğrenciler için karmaşık bir anlatıma sahip olduğunu, paragrafta kullanılan dilin ilköğretim 8. Sınıf öğrencileri için ağır olduğunu belirtmiştir. Uzmanların bir kısmı kullanılan soru tipinin Türk öğrencilerin alışık olmadığı bir stil olduğu vurgulamıştır. PISA sınavında yer alan farklı soru tipleri, eğitimleri boyunca genellikle tek soru tipine alışmış öğrencilerde sıkıntı yaratabilmektedir. Aslında çoktan seçmeli soru tipine alışık olan Türk öğrenciler sorunun farklı bir biçimde sorgulanmasından ötürü dezavantaj sahibi olabilir. Yani sorunun tipinin yanı sıra nasıl sorgulandığının da yanlılık nedeni olabileceği konusunda bir görüş birliği bulunmaktadır. Yapılan analizler incelendiğinde bu maddenin MH tekniği ile yapılan analiz sonucunda her iki kitapçıkta da

A düzeyinde İngiltere lehine DMF içerdiği belirlenmiştir. LR tekniğine göre madde DMF içermemektedir. AİT tekniğine göre her iki kitapçıkta da maddenin İSAİ sonuçlarına göre DMF içerdiği belirlenmiştir

1. kitapçıkta yer alan 8. madde için (Grand Kanyon okuma parçasına ait madde); uzmanlar bu maddenin Türkiye fen ve teknoloji dersi müfredatı kapsamında olmadığını, okuma parçasının kısmen coğrafya dersi müfredatında Jeoloji konusu içerisinde yer alabileceğini belirtmişlerdir. Ayrıca İngilizce formdaki “shale” kelimesinin karşılığı Türkçe formda “kilin sıkışması ile oluşan tortul kayaç” biçimindedir. Uzmanlar ilköğretim 8. Sınıf öğrencilerinin bir kelime ile belirtilebilecek bir kavramın uzun bir kelime grubuyla açıklanmasının karışıklığa ve anlaşmazlığa neden olabileceği konusunda hemfikirlerdir. Grisay (2003)’e göre, kelime ve cümlelerin ortalama uzunlukları diller arasında farklılık gösteren bir özellik olduğunu ve ölçme araçlarının uyarlanması aşamasında çevirmenler tarafından tamamen kontrol edilemediğini belirtmiştir. Benzer biçimde soruya ait okuma parçasının başlığının “Grand Kanyon” gibi özel bir isim olması nedeniyle Türk öğrencilerin aleyhine işleyebileceği belirtilmiştir. “Grand Kanyon” un daha bilindik bir bölge olması sebebiyle İngilizce formu alan öğrenciler lehine çalışmasının beklendiği ve bu maddenin yanlış olarak kabul edilebileceği konusunda uzmanlar ortak görüş bildirmişlerdir. Uzmanlar PISA fen maddelerinin çoğunluklu olarak İngiltere lehine işleyeceği görüşündedirler. Bu madde için yapılan DMF analizleri incelendiğinde maddenin MH tekniğine göre C düzeyinde, LR tekniği analiz sonuçlarına göre ise madde B düzeyinde çok biçimli DMF göstermektedir. Alan indeksleri tekniği sonuçlarına göre ise madde DMF göstermemektedir.

Uzmanlar açıklanan maddelerden 5. kitapçıkta yer alan 12. maddenin İngilizce formu alan öğrenciler lehine yanlış olmasını beklediklerini belirtmişlerdir. Yanlılığın olası nedeni olarak; koşullar ve malzeme temininde yaşanan sıkıntıları, müfredattaki farklılıkların ve mevcut metindeki çeviri yetersizliğini belirtmiştir. Uzmanların bir kısmı bu maddenin ülkemize has sınıfsal farklılıklardan dolayı Türkiye’ nin farklı bölgelerinde bile farklı işleyebileceği hususunda görüş bildirmiştir. Benzer olarak görüş bildiren uzmanlardan biri bu sorunun yanıtlanmasında öğrencilerin “kontrollü deney” hakkında bilgi sahibi olması gerektiğini vurgulamış ve ülkemizde yürütülen Fen bilimleri derslerinin ağırlık olarak çok azının laboratuvar koşullarında yürütüldüğünü bildirmiştir. DMF analizleri incelendiğinde bu maddenin MH tekniğine göre yapılan analizlerde C düzeyinde, LR tekniğine göre yapılan analizlerde madde A düzeyinde DMF içermektedir. AİT tekniği sonuçlarına göre madde hesaplanan İAİ’ye göre Türkiye lehine yanlılık gösterirken, İSAİ’ ye göre DMF göstermemektedir.

5. kitapçıkta yer alan 17. madde için uzmanlar; okuma metni ve soru arasında ilişkinin kurulamayışı, kullanılan sözcük ve kavram nedeniyle maddenin yanlış işlendiğinin düşünüldüğü belirtilmiştir. Bunun ise çeviriden kaynaklandığını ve çevirinin bu anlamsal farklılıklara neden olduğu görüşündedirler. Belirgin örneğini şu biçimde vermişlerdir. Uzmanlar okuma parçasında virüslerden bahsedildiğini fakat soruda bakterilerle ilgili soru yöneltildiğini belirtmiştir. Benzer biçimde bu madde için diğer

olası yanlılık nedeni olarak; Türkiye’de verilen eğitimin daha çok teorik olması ve teorik eğitimin pratik eğitime yansımaması, uygun laboratuvar koşullarının sağlanamaması sıralanmıştır. Yapılan DMF analizleri incelendiğinde bu maddenin MH tekniğine göre C düzeyinde, LR ve alan indeksleri tekniğine göre ise DMF göstermediği belirlenmiştir.

Benzer DMF analizi çıktısına sahip 5. kitapçık 18. madde için de uzmanlar olası yanlılık sebebi olarak; metin içerisindeki karışık ifadelerin ve genel kültür düzeyini etkileyen faktörlerin bu maddeyi yanıtlamada yanlılığa sebep olduğunu bildirmişlerdir.

Açıklanan maddeler için yapılan yanlılık araştırmasında Türkçe ve İngilizce formda yer alan açıklanan maddeleri inceleyen uzmanların bu maddeler için belirttikleri olası yanlılık nedenleri dil ve kültürle ilgili yapılan yanlılık çalışmalarında ortaya konan sebeplerle uyusmaktadır (Gierl ve Khaliq, 2001; Ercikan, 1998; Grisay, 2003; Yıldırım ve Berberoglu, 2009; Asil, 2011).

4. Sonuç ve Tartışma

Bu çalışmanın temel amacı, PISA 2006 fen başarı testinde yer alan maddelerin aynı kültür farklı dil, farklı kültür aynı dil ve farklı kültür farklı dil çerçevesinde değişen madde fonksiyonunun farklı teknikler ile analiz edilmesi ve karşılaştırılması idi. Bu temel amaç çerçevesinde PISA 2006 başarı testinde yer alan çoktan seçmeli fen maddeleri, üç farklı DMF belirleme yöntemine göre, testteki madde kapsamı bakımından yanıtlayıcıların dili ve kültürü bakımından incelenmiştir. Yapılan analiz sonucunda yanıtlayıcıların dilinin ve kültürünün DMF’ye neden olduğu sonucuna ulaşılmıştır. Elde edilen bulgular, PISA 2006 Fen başarı testinin çoktan seçmeli sorularının madde yanlılığı bakımından çeşitli açılardan tartışılır olduğunu göstermektedir. Elde edilen bulgular incelendiğinde; karşılaştırılan alt gruplar arasındaki farklılıklar arttıkça DMF gösteren madde sayıları ve düzeylerinin arttığı sonucuna ulaşılmıştır.

DMF içeren maddelerin sayısının yüksek olması, bu testten alınan puanların karşılaştırılması durumunda bizleri doğru yorumlardan uzaklaştıracaktır. Dolayısı ile PISA testinin, dil ve kültür bakımından çalışılan gruplar içerisindeki bazı gruplara avantaj sağlıyor görünmesi bu test ile alınan kararların, geçerliliğini ve güvenilirliğini de tartışılır kılmaktadır. Ancak unutulmamalıdır ki DMF’nin doğrudan madde yanlılığının bir göstergesi değildir. Bu DMF’nin araştırıldığı alt gruplarda maddenin gerçek etkisi de olabilir. Bu nedenle öncelikli olarak maddenin gruplarda nasıl çalıştığı, grubun başarısı gibi etmenlerin de araştırılması uygun olacaktır. Aynı zamanda PISA sınavına katılan tüm ülkeler için benzer çalışmalar yapıldığında karşılaştırmaların doğruluğuna ilişkin genellemeler yapılması mümkün olabilecektir. Bu nedenle en azından PISA gibi uluslararası testlerde düzenli olarak DMF analizleri yapılması önerilmektedir. Yapılan DMF analizleri sonrasında DMF gösteren maddelerin neden DMF gösterdiği nitel araştırmalarla desteklenmeli ve söz konusu maddelerde yanlılığa neden olan

özellik ortadan kaldırılamıyorsa o madde ya da maddeler testlerden çıkarılmalıdır.

Elde edilen bulgulardan da anlaşılacağı üzere teknikler birbirlerinden çoğu kez farklılaşmakta ve elde edilen değerler birbiriyle uyumlu gözükmemektedir. Farklı DMF belirleme teknikleri ile yapılan diğer bir çok çalışmada aynı bulgu mevcuttur (Camilli ve Shepard, 1994; Bakan Kalaycıoğlu ve Kelecioğlu, 2011; Deveci Ateşok, 2008) .Bu sonuç DMF analizlerinde karşılaşılan en temel problemlerden biri olarak karşımıza çıkmaktadır (Karami ve Nodoushan; 2011). Bu araştırmada Klasik Test Kuramına (KTK) bağlı tekniklerin kendi arasında; Madde Tepki Kuramına (MTK) bağlı tekniklerin de kendi aralarında uyumlu olduğu yönündedir. Bu tekniklerin farklılaştıkları noktaların bu farklılığa neden olduğu düşünülmektedir. Bir bakıma KTK yöntemlerinden LR ve MH teknikleri benzer bileşenlere sahiptir, her ikisinde de toplam test puanları üzerinden işlem yapılmasına rağmen kategorilendirmelerde kullanılan ölçütlerin bu uyumu bozduğu düşünülmektedir. Bu çalışmada DMF’li maddelere karar vermede daha etkili bir yöntem olduğu savunulan (Camilli, 2006) etki büyüklüğü üzerinden sonuçlara ulaşılmıştır. Yöntemler kısmında MTK’ya bağlı ölçütler belirlenirken vurgulandığı gibi KTK’ya bağlı yöntemlerde de etki büyüklüğü ölçütü değişebilmektedir. Nitekim Lopez-Pina (2004) yaptıkları çalışma sonucunda, ΔR^2 istatistiğinin yorumlanması için yeni ölçütlerin belirlenmesi gerektiğini vurgulamışlardır. Kalaycıoğlu ve Kelecioğlu (2011) yapmış oldukları çalışmada ΔMH için 0,010 ve 0,020 değerlerini kullanmışlardır. Öte yandan MTK’na dayalı yöntemlerin kullanılmasında sağlanması gereken varsayımların çeşitliliği ve bu varsayımlardan hareketle elde edilen madde parametre kestirimlerinin; MTK ile KTK’na göre DMF belirleme yöntemleri arasındaki uyumsuzluğun veya farklılığın nedeni olabileceği düşünülmektedir.

Abedi 2004 ve 2008 yıllarında ELLs (English Language Learners) sınavında aynı dil formunu alan farklı alt yapıya sahip öğrenciler üzerinde yaptığı çalışmada karmaşık - zor cümle yapılarında farklı alt gruplardan gelen bireylerin performanslarının karşılaştırılmasında sıkıntılar olabileceğini vurgulamıştır. Bu sonuç; kültürlerin farklılaştığı ancak alınan form dilinin değişmediği durumlarda bu çalışmada farklı tekniklerle belirlenen DMF’li maddeler için bir dayanak olabilir. Alt örneklemelerin farklılaşmasının yanı sıra; ortaya çıkan farklılıkların en temel sebeplerinden birinin; yöntemlerin DMF belirlemede izledikleri istatistiksel yolun ve aşamaların farklılaşması olduğu düşünülmektedir. Yöntemlerle elde edilen sayılar DMF kararı için kullanılmakta; ancak elde edilen bulgulara farklı matematiksel yollarla ulaşılmaktadır.

Bu çalışmada üzerinde durulan dil ve kültürden kaynaklanan yanlılığı, hem kültür farklılıklarının hem de dil farklılıklarının farklı biçimde etkileyeceği ve etkilediği düşünülmektedir. Kültürün aynı dilin farklı olduğu Kanada örneğinde olduğu gibi alan indeksleri tekniği ile bulunan DMF’li madde sayısı kültürün farklı dilin aynı olduğu Avustralya-İngiltere örneğine göre göreceli olarak fazla bulunması çalışılan bu örneklemelerde DMF de dilin daha belirleyici bir etken olduğunu düşündürülebilir. Ancak dil, hem bütün insan nüfusları için ortak bir yetenek, hem de kültür topluluklarını

birbirinden ayıran farklılaşmış bir uygulamadır. Kültür ise, insanın toplumsal çevreye genel uyarlanma yöntemi olduğu kadar, aynı zamanda topluluktan topluluğa farklılaşma gösteren bir yaşam ve düşünce tarzıdır (Beals R.L.ve Hoijer H.; 1972). Nitekim aynı alt grup için MH tekniğiyle bulunan DMF’li madde sayısı eşitken; LR tekniği ile DMF olduğu belirlenen madde sayısı Avustralya-İngiltere grubunda alan indeksleri tekniğinde belirlenen madde sayısının üstünlüğü kadar olmasa da fazla bulunmuştur. Bu tanımlardan ve açıklamalardan dilin ve kültürün birbiri içerisinde kavramlar olduğu, ayrı ayrı düşünülmesinin yanlış yorumlamalara neden olabileceği anlaşılmaktadır. Bu çalışma içinde kültürün, dilin hem de dil ve kültürün ortak etkileşiminin DMF’li madde sayısı üzerinde etkili olduğu, birinin diğerine baskın olmayacağı düşünülmektedir.

Maddelerin kültürler ve diller arasında farklı fonksiyon göstermesinin olası nedenlerinin neler olduğu konusunda uzman görüşleri incelendiğinde; birden fazla dile çevrilen ölçme araçlarında çeviriden kaynaklanan farklılıkların ve kültürel farklılıkların yanlılığa neden olabileceği sonucuna ulaşılmıştır. Maddelerin kültürler ve diller arasında farklı fonksiyon göstermesinin olası nedenlerinin neler olduğu konusunda uzman görüşlerinden ortaya çıkan ortak sonuç genel biçimiyle şu biçimde özetlenebilir. Uzmanlar PISA maddelerinin yanlılık göstermesinin temel nedeninin çeviri, müfredat farklılıkları ve uyarlama yanlılıkları sonucu metnin ağır bir dile dönüşmesi, sosyo ekonomik düzeyden kaynaklanan farklılıklar olduğunu belirtmişlerdir. Maddelerin karmaşık ve uzun metinlerle ilişkilendirilmesinin özellikle Türk öğrencilerin alışkın olmadığı bir formata dönüşmesinin bazı maddelerin yanıtlanmasını zorlaştırdığı fikrinin hakim olduğu sonucuna ulaşılmıştır. Uzmanlar açıklanan PISA çoktan seçmeli fen maddelerinin çoğunluklu olarak İngiltere lehine işleyeceği görüşündedirler. Bu sonuçtan yola çıkarak uzmanların inceledikleri sorularla ilgili temel önerileri şu biçimdedir: Kısa, basit ve aktif cümleler kullanılması; deyim atasözü gibi kelime grupları ile; kültüre özgü sözcük ve terimlerin kullanılmasından kaçınılması gerekmektedir. Maddeler yazılırken kültürler arası farklılıklar dikkate alınmalıdır. Toplumlar mevcut kavramları ile görüş geliştireceğinden özellikle madde yazımında kavramların yüzeysel anlamları ve derin anlamları üzerine çalışmalar yapmak yararlı olacaktır.

Değişen madde fonksiyonu (DMF) çalışmalarında “Differential Item Functioning” ifadesini daha uygun bir çeviri ile karşılayan ortak dil birliği sağlanması önerilebilir. Her ne kadar DMF alanyazınımızda yaygınlaşmış olsa da “maddenin ayrımcılık yapacak biçimde çalışması” ifadesinin Türkçe çevirisi ortada değişen diferansiyel bir fonksiyon varmış gibi algılanmaktadır.

5. Kaynakça

Abedalaziz N.,2011, Detecting DIF using Item Characteristic Curve Approaches.The International Journal of Educational and Psychological Assessment , Vol. 8(2)

- Abedi J. (2008). Classification System for English Language Learners: Issues and Recommendations Educational Measurement: Issues and Practice; 27(3), 17-31.
- Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher*, 33, 4-14.
- Asil, M. (2010). Uluslararası Öğrenci Değerlendirme Programı (PISA) 2006 öğrenci anketinin kültürler arası eşdeğerliğinin incelenmesi. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi, Ankara.
- Asil, M. Ve Gelbal, S. (2012). PISA Öğrenci Anketinin Kültürler Arası Eşdeğerliği, Eğitim ve Bilim, 37, 166, 236-249.
- Atalay K. (2010) *PISA 2006 Öğrenci Anketinde Yer Alan Tutum Maddelerinin Değişen Madde Fonksiyonu Açısından İncelenmesi*. Yüksek Lisans Tezi Eğitim Bilimleri Bölümü, Hacettepe Üniversitesi, Ankara.
- Bakan Kalaycıoğlu, D. ve Kelecioğlu, H. (2011). Öğrenci Seçme Sınavı'nın Madde Yanlılığı Açısından incelenmesi. *Eğitim ve Bilim*, 36, 3-13.
- Baker, Frank. (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221-256). Westport: American Council on Education ve Praeger Publishers.
- Camilli, G., Shepard, L.A. (1994). Methods for identifying biased test items. Thousand Oaks: Sage.
- Deveci Ateşok, N. (2008) Üniversitelerarası Kurul Yabancı Dil Sınavının Madde Yanlılığı Bakımından İncelenmesi; Doktora Tezi, Ölçme ve Değerlendirme Anabilim Dalı, Ankara Üniversitesi
- Deveci, N. (2008). Üniversitelerarası Kurul Yabancı Dil Sınavı'nın Madde Yanlılığı bakımından incelenmesi. Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Eğitimde Psikolojik Hizmetler Anabilim Dalı. Yayınlanmamış Doktora Tezi, Ankara.
- Ercikan, K., (1998), Translation Effects In International Assessments, International Journal of Educational Research, 29, 543-553.
- Gierl, M., Jodoin, G. M. Ackerman, T. A. (2000). Performance of Mantel-Haenszel, Simultaneous Item Bias Test, and Logistic Regression when the proportion of DIF items is Large. Paper Presented at the Annual Meeting of the American educational Research Association (AERA). New Orleans, Louisiana, USA.
- Giril 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; Vol. 38, No. 2.; pp. 164-187*
- Gondal, M., B,(2001). *Pakistan'daki 4. Sınıf Ve Urdu (Ulusal Dil) Başarı Test Maddelerinin Madde Yanlılığı Açısından İncelenmesi*. Yayınlanmamış Doktora Tezi, Orta Doğu Teknik Üniversitesi, Eğitim Bilim Bölümü, Ankara.
- Grisay, A. (2003). Translation Procedures in OECD/PISA 2000 International Assessment. *Language Testing*, 20(2), 225-240
- Hambleton, R. K., Swaminathan, H. (1985) Item Response Theory: Principles and Applications. Boston: Kluwer-Nijhoff Publishing.
- Hambleton, R.K. Rogers, H., J.(1996). Developing an Item Bias Review Form. Which is Available Through ERIC/AE.

- Lord, M.F., Novic, R. M. (1968). Statistical theories of mental test scores. New York: Addison-Wesley Publishing Company.
- Millsap, R.E. (2005). Four unresolved problems in studies of factorial invariance. In A. Maydeu-Olivares & J.J. McArdle (Eds.) Contemporary Psychometrics. (pp. 153-171) Mahwah, NJ: Lawrence Erlbaum.
- Millsap, R.E. Meredith, W. (1992). Component analysis in multivariate aging research. Experimental Aging Research, 18, 203-212
- Osterlind, S.J. (1983). Test item bias. Beverly Hills: Sage
- Öğretmen, T. (2006). Uluslararası Okuma Becerilerinde Gelişim Projesi (PIRLS) Testinin psikometrik Özelliklerinin İncelenmesi: Türkiye- Amerika Birleşik 2001 Devletleri Örneği. Doktora Tezi. Eğitim Bilimleri Bölümü, Hacettepe Üniversitesi.
- Öğretmen, T., Doğan, N. (2004). OKÖSYS Matematik alt testine ait maddelerin yanlılık analizi. İnönü Üniversitesi Eğitim Fakültesi Dergisi. 8, 61-76.
- PISA 2006 Uluslararası Öğrenci Başarılarını Değerlendirme Programı Ulusal Ön Rapor (2007) <http://earged.meb.gov.tr/dosyalar> adresinden indirilmiştir.
- Raju, N. S. (1990). The area between two item characteristic curves. Psychometrika, 53, 495–502.
- Raju, N. S., Ellis, B. B. (2002). Differential item and test functioning. In F. Drasgow & N. Schmitt (Eds.), Measuring and analyzing behavior in organizations (p. 156-188). San Francisco, CA: Jossey-Bass.
- Yıldırım H, Berberoglu G. ;Judgmental and statistical DIF analyses of the PISA-2003 mathematics literacy items. International Journal of Testing, 2009; 9(2):108-121. (PsycInfo)
- Yıldırım, H.(2006). The Differential Item Functioning (DIF) Analysis Of Mathematics İtems in The International Assessment Programs. Yayımlanmamış Doktora Tezi. Orta Doğu Teknik Üniversitesi, Matematik Eğitim Bölümü. Ankara.
- Zieky, M. (2003). A DIF primer. Princeton, NJ: Educational Testing Service.
- Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores. Ottawa, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense

Extended Abstract

In comparability investigations, the presence of differential item functioning (DIF) is considered to be an indication of possible bias. Differential Item Functioning (DIF) analysis has become the new standard in psychometric bias analysis. DIF statistical techniques are based on the principle that if different groups of test-takers (e.g., males and females) have roughly the same level of something (e.g., knowledge), then they should perform similarly on individual test items regardless of group membership (Zumbo, 1999). In this study, differential item functioning (DIF) analyses of Science items of PISA 2006 tests were carried out between different samplings in respect to language and culture. DIF methods allow one to judge whether items are functioning in the same manner in various groups of examinees. Mantel Haenszel (MH), logistic regression(LR) and signed - unsigned area indexes methods were used for DIF detection analyses.

The research group of this study consists of the Australia sample comprising 1124 students, the Canada sample comprising 1744 students; the England sample comprising 1008 students, the Turkey sample comprising 377 students; took the fifth booklets and the England sample comprising 1430 students, the Turkey sample comprising 380 students; took the first booklets. These countries were selected due to the differences in cultural relevance and linguistic are the possible main reasons for differential item functioning (DIF). In order to investigate the sources of DIF (possible bias sources) field specialist opinions were consulted. Because, DIF is required, but not sufficient, for item bias. It is important to know that DIF and bias are not same. DIF occurs when examinees from different groups show differing probabilities of success on (or endorsing) the item after matching on the underlying ability that the item is intended to measure. Bias occurs when examinees of one group are less likely to answer an item correctly than examinees of another group because of some characteristic of the test item or testing situation that is not relevant to the test purpose (Zumbo, 1999).

In the study, in Canadian sampling, DIF was found in three items at B level and three items at negligible level according to the MH technique and in three item at negligible level according to LR technique, in five items according to each fo signed - unsigned area indexes methods. In Australia- England sampling DIF was found in one item at B level and C level, four items at negligible level according to the MH technique and in four items at negligible level according to LR technique, in two items according to each fo signed - unsigned area indexes methods. In England- Turkey sampling for the first booklet; ten items included DIF according to MH results; five of them were at A level, two of them were at B level and three of them were at C level according to the MH technique those of the items eight of them favored English form, where two of them favored Turkish form. DIF was found in five items at negligible level and one item at B level according to LR technique, in six items according to each for signed-unsigned area indexes methods. In England-Turkey sampling for the fifth booklet; in two items at A level and in four items at B level and in five items at C level according to the MH technique and infour item at negligible level and two items at B level according to LR technique, in six items according to each fo signed-unsigned area indexes methods had DIF.

It is observed that as the linguistic and cultural differences increased between countries, the number of DIF items increased. The number of DIF items varied significantly according to the procedure used. The correlation coefficients for the same culture-different language between LR and MH were significant, Non-signed area indexes and Signed area indexes were significant at $\alpha = 0,01$. The correlation coefficients for the different culture-same language between LR and MH were significant at $\alpha = 0,05$, For the different culture and language LR and MH were significant at $\alpha = 0,01$ for the first booklet and Non-signed area indexes and Signed area indexes were significant at $\alpha = 0,01$ for the fifth booklet.

As mentioned above; DIF is a necessary, but not sufficient, condition for item bias. Thus, if DIF is not apparent for an item, then no item bias is present. However, if DIF is apparent, then its presence is not sufficient to declare item bias. Because of this field specialist opinions were consulted. When the experts' opinions were examined; the main reasons for DIF were changes in word difficulty, changes in item format, differences in cultural relevance, translation, curriculum, linguistic differences and changes in content across Turkish and English versions of the tests of released items.