

PISA 2009 OKUMA TESTİ MADDELERİNİN YANLILIĞI ÜZERİNE BİR ÇALIŞMA¹

Dr. Deniz Tuğçe Özmen

Ankara Üniversitesi

Özet

Bu araştırmanın amacı, PISA 2009 okuma becerileri testinde kullanılan maddelerin Türkiye-Amerika Birleşik Devletleri (ABD), Türkiye-Birleşik Krallık (BK), ABD-BK grupları arasında DMF (değişen madde fonksiyonu) gösterip göstermediğinin incelenmesi, DMF içeren maddelerin olası yanlılık nedenlerinin belirlenmesidir. Bu amaçla, Mantel Haenszel (MH), SIBTEST, Madde Tepki Kuramı Olabilirlik Oran (MTK-OO) ve b parametrelerinin karşılaştırılması yöntemleri kullanılmıştır. Araştırma grubunu, PISA 2009 okuma becerileri testinde dört ve altı numaralı kitapçıkları yanıtlayan Türk öğrenci grubundaki 767, ABD öğrenci grubundaki 796, BK öğrenci grubundaki 765 öğrenci oluşturmaktadır. Çeşitli yöntemlere göre DMF gösteren maddelerin madde formatı, okurun metne yaklaşımı ve avantaj sağladığı gruba göre dağılımları incelenmiştir. Elde edilen sonuçlara göre, Türk ve ABD’li, Türk ve BK’daki öğrenci grupları arasında en az altı numaralı kitapçıkta (%31 ve %33) en fazla ise dört numaralı kitapçıkta (%36 ve %49) DMF içeren madde bulunduğu, bu maddelerin odak ve referans gruplar arasında dağılımının yaklaşık olarak eşit olduğu belirlenmiştir. Madde formatı açısından incelendiğinde, referans grubun lehine DMF gösteren maddelerin çoğunlukla açık uçlu olduğu saptanmıştır. Ayrıca, metne yaklaşım açısından incelendiğinde, “bilgiyi bir araya getirme ve yorumlama” yaklaşımına göre geliştirilmiş DMF içeren maddelerin çoğunlukla referans grubun lehine olduğu sonucuna ulaşılmıştır. DMF gösteren maddelerin, sözcüklerin güç anlaşılması, madde formatına aşına olma durumu, çeviri hataları ve kültürel farklılıklardan kaynaklandığı yönünde uzmanlar tarafından ortak görüş bildirilmiştir.

Anahtar Sözcükler

Değişen madde fonksiyonu, PISA 2009, Mantel Haenszel, SIBTEST, Madde Tepki Kuramı-Olabilirlik Oran, b parametreleri.

© 2014, *Eğitim Bilimleri ve Uygulama*, 13 (26), 147-165

1 Yazarın, Prof. Dr. Nizamettin Koç danışmanlığında hazırlanmış olduğu doktora tezinden yararlanılarak hazırlanmıştır.

A STUDY ON PISA 2009 READING TEST ITEMS IN TERMS OF BIAS

Dr. Deniz Tugce Ozmen

Ankara University

Abstract

The aim of this research is to determine the items indicating differential item functioning (DIF) and its possible sources of bias among Turkey, United States of America (USA) and the United Kingdom (UK) in PISA 2009 reading test. In this context, the methods of Mantel Haenszel (MH), Simultaneous Item Bias Test (SIBTEST), Item Response Theory Likelihood Ratio (IRTLR) and comparison of b parametres have been used. The research group involves 767 pupils from Turkey, 796 pupils from USA and 765 pupils from the UK who answered the questions in Booklet 4 and Booklet 6 of PISA 2009 reading test. The distributions of between-groups items indicating DIF in accordance with the DIF method have been investigated with regards to item format, aspect and the advantageous group. The results indicate that there are less items with DIF in Booklet 6 (31% and 33%) and there are more items with DIF in Booklet 4 (36% and 49%) between Turkey and USA, Turkey and the UK. Also, the distribution of items with DIF is nearly equal between reference and focal groups. In the context of item format, it has been determined that the items with DIF being in favour of reference groups are open-ended. Besides, from the point of text analysis, the items developed with “integrate and interpret” aspect are mostly also in favour of reference groups. The sources of bias have been identified as hardly understood, being familiar with item format, translation errors and cultural differences by the experts.

Keywords

Differential item functioning, PISA 2009, Mantel Haenszel, SIBTEST, Item Response Theory-Likelihood Ratio, b parametres.

GİRİŞ

Testler, insan yaşamında önemli yere sahip olan, bazı psikolojik özellikleri ve davranışları ölçmek amacıyla kullanılan sistematik süreçler içeren araçlardır (Linn ve Gronlund, 2000). Eğitimde ve psikolojide kullanılan ölçme araçları, bireyleri seçme ve sınıflandırmada, eğitim süreçlerini değerlendirme ve bilimsel hipotezleri kabul etme veya reddetme amacıyla verilen kararlara yardımcı olmaktadır (Cronbach, 1990). Ayrıca, ölçme araçlarından elde edilen puanlar, bir işe veya eğitim programına başvuran bireyleri elemek, eğitim ve mesleki bağlamda bireyleri sınıflandırmak ve yerleştirmenin yanı sıra eğitim, mesleki ve kişisel konularda danışma amacıyla bireylere rehberlik etmek, öğrencileri yönlendirmek, çalışanları daha üst pozisyona getirmek ve program ya da yöntemlerin ne kadar etkili olduğunu değerlendirmek amacıyla kullanılabilir (Aiken, 1997).

İnsanların yaşamlarının belirli dönemlerinde sıklıkla karşılaştıkları testlerin, özellikle eğitim ve mesleki alanlara ilişkin önemli kararların alınmasında kullanılıyor olması, araçların güvenilirlik ve geçerliğine ilişkin çalışmaların yapılmasını gerektirmektedir. Test kullanıcılarına göre farklı özelliklere sahip gruplardan elde edilen puanlar birbirleriyle karşılaştırılabilir ve bu karşılaştırma sonuçları çeşitli kararlara temel olarak kullanılabilir. Ancak bir testteki maddelerin bir grup lehine yanlılık göstermesi, adil olmaması durumunda testin geçerliği olumsuz yönde etkilenir (Kane, 2006; Messick, 1989, 1988). Eğitimde ve psikolojide kullanılan bir testin, bir grubun (kız veya erkek gibi) haksız olarak lehine veya aleyhine olan maddeler içermesi, o testin yanlı olduğunu gösterir. *Test yanlılığı*, bir testin belli bir alt gruba ait bireyler için sistematik hata göstermesidir (akt. Camilli ve Shepard, 1994). Ayrıca, test yanlılığı, bir testin sonuçlarına göre alınacak bir kararın her alt grup için adil olmaması veya bir grubun üzerinde fark edilir bir şekilde etkisi olması şeklinde de tanımlanabilir. *Madde yanlılığı* ise yanlılığın madde düzeyine indirgenmesidir, bir maddenin belli bir alt grup lehine yanlılık göstermesidir (Osterlind ve Everson, 2009).

Maddenin bir grup lehine yanlı olması, bir testteki maddenin *değişen madde fonksiyonu* (DMF) göstermesi (*differential item functioning*) olarak adlandırılır (Dorans ve Holland, 1993; Holland ve Thayer, 1988; Osterlind ve Everson, 2009). Bir testteki maddelerin farklı fonksiyonlaşması, aynı yetenek düzeyinde ve eşit yeterlikte olan ancak cinsiyet, etnik köken, engelli olma, yaş gibi çeşitli özellikler bakımından farklı gruplarda bulunan bireylerin bir maddeyi doğru olarak yanıtlama olasılıklarının eşit olmamasıdır (Angoff, 1993; Camilli ve Shepard, 1994; Ellis ve Raju, 2003; Osterlind ve Everson, 2009).

Camilli ve Shepard (1994) ve Osterlind ve Everson'a (2009) göre alt grupların puanları arasında farklılık olması, her zaman yanlılık göstergesi değildir. Yanlılık

çalışmalarının amacı, test sonuçlarının testi alan alt gruplara göre değişiklik göstermesinin nedeninin gruplar arasındaki başarı ve yeteneğe dayalı farklılıkların veya ölçme işlemindeki sistematik bir adaletsizliğin sonucu olup olmadığını saptamaktır. Bu nedenle, *ölçme yanlılığı* (*measurement bias*) ve *madde etkisi* (*item impact*) kavramları arasındaki fark önemlidir. *Madde etkisi* kavramı, testlerle ve test maddeleri aracılığıyla ölçülmesi amaçlanan performans bakımından oluşan gerçek grup farklılıkları anlamına gelmektedir. *Ölçme yanlılığı* ise aynı yetenek düzeyindeki fakat farklı alt gruplara ait bireylerin bir maddeyi doğru yanıtlama olasılıklarının eşit olup olmama durumudur (Camilli, 2006; Camilli ve Shepard, 1994; Dorans ve Holland, 1993; Osterlind ve Everson, 2009).

Bir testte bir maddenin yanlı olması sorun oluşturmaz, ancak birkaç maddenin yanlı olması testten elde edilen puanlar üzerinde önemli bir fark yaratır. Yanlı madde sayısı arttıkça test, artık bireylere ait özellikleri adil olmayan bir biçimde ölçmeye başlar. Bir testteki yanlı maddelerin varlığı, geçerlik sorununu gündeme getirir. Belli bir alt gruba avantaj sağlayacak maddelerin bulunduğu bir testin sonuçlarına göre alınacak kararların doğru olmayacağı da açıktır. Bu duruma dayalı olarak, 1999 yılında Amerikan Psikologlar Derneği (American Psychological Association - APA), Eğitsel ve Psikolojik Testlerin Standartları (Standards for Educational and Psychological Testing), çalışmasında geçerliğin test geliştirmede önemli bir konu olduğunu, yanlılık belirleme çalışmaları dâhil, testlerin geçerliği konusunda kanıtlar içeren çalışmaların, puanların yorumlanmasına anlamlı katkı getireceğini belirtmiştir. Dolayısıyla, testlere ilişkin yanlılık çalışmaları ölçme ve değerlendirme sürecinde temel bir gereklilik olarak kabul edilmektedir (Osterlind ve Everson, 2009). Örneğin, bir testten elde edilen puan ile bir eğitim programına kabul, işe alınma ya da bir uzmanlığın belgelendirilmesi gibi önemli kararlar alınması söz konusu ise o testin geçerliğine ilişkin kanıt elde edilmesi gerekmektedir. Belirtilen nedenlerle, birçok araştırmacı testlerin geçerliğine ilişkin kanıt elde etmek için cinsiyet, yerleşim yeri, farklı diller, kültürler, ırklar arasında maddelerin yanlılık gösterip göstermediğini belirlemek üzere çalışmalar yürütmüştür.

Yanlılık ile ilgili yapılan araştırmalardan bir kısmı, ulusal düzeydeki sınavlara ilişkin madde yanlılığı çalışmalarıdır. Bu araştırmalarda, Öğrenci Seçme Sınavı (ÖSS), Öğrenci Yerleştirme Sınavı (ÖYS), Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı (OKÖSYS), Üniversitelerarası Yabancı Dil Sınavı (ÜDS), Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) ve Seviye Belirleme Sınavı (SBS) gibi ulusal düzeyde düzenlenen sınavlar üzerinde MTK ve Klasik Test Kuramı'na (KTK) dayalı çeşitli yöntemler aracılığıyla maddelerin cinsiyet, konu alanları, yerleşim yeri, okul türü, bölüm, bölge gibi değişkenlere ilişkin yanlılık içerip içermediği incelenmiştir (Acar, 2008; Berberoğlu, 1995; Berberoğlu, 1996; Çepni, 2011; Deveci, 2008; Kalaycıoğlu, 2008; Karakaya ve

Kutlu, 2012; Öğretmen, 1995; Öğretmen ve Doğan, 2004; Yenal, 1995; Yurdugül, 2003).

Ulusal düzeydeki sınavların yanısıra, farklı ülkeler, diller ve kültürleri karşılaştırma amacı ile yapılan sınavlara ilişkin DMF belirleme çalışmaları da bulunmaktadır. Bu çalışmalardan bazıları PISA (The Programme for International Student Assessment - Öğrenci Değerlendirme Programı) üzerinde yürütülmektedir. PISA, Ekonomik İşbirliği ve Kalkınma Teşkilatı (OECD – Organisation for Economic Cooperation and Development) tarafından yürütülen, öğrencileri uluslararası düzeyde başarı ve başka değişkenler açısından karşılaştırmayı amaçlayan kapsamlı bir çalışmadır. Çalışma kapsamında uygulanan ölçme araçlarının kırktan fazla dile uyarlaması yapılmaktadır. Kullanılan bu ölçme araçlarından elde edilen sonuçlara dayalı olarak ülkeler eğitim politikalarına yön verebilmekte, önemli kararlar alabilmektedirler. PISA kapsamında kullanılan ölçme araçlarında, bir alt gruba avantaj sağlayan maddelerin bulunması, elde edilen puanların geçerliğini olumsuz yönde etkileyecektir. Bu durumda bu sınavların sonuçlarına göre alınacak kararlar doğru olmayacaktır. Bu nedenle, farklı diller ve kültürleri karşılaştırmaya dayalı yapılan sınavlardaki maddelerin DMF içerip içermediğine dair birçok araştırma yapılmıştır. Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS - Trends International Mathematics and Science Study) ve PISA verileri üzerinde yürütülen araştırmalarda, farklı dil, kültür ve ülke grupları, MTK ve KTK'ya dayalı yöntemler kullanılarak, DMF içerip içermediği incelenmiştir. Yapılan araştırmalardan elde edilen sonuçlar, dilsel ve kültürel farklar artıkça DMF içeren maddelerin de arttığını göstermektedir. Kültürel denkleğin olmaması, çevirinin yetersiz olması, hedef dilde tam karşılığı olmayan sözcüklerin kullanılması, madde formatına aşına olmama ve eğitim programlarındaki farklılıklar DMF kaynakları olarak belirlenmiştir (Asil, 2010; Allouf ve Sireci, 1998; Atalay, 2010; Başusta, 2013; Schulz, 2003; Grisay ve Monseur, 2007; Ulutaş, 2012; Wu ve Ercikan, 2006; Yıldırım, 2006; Yıldırım ve Berberoğlu, 2009).

Yurtiçinde ve yurtdışında yapılan araştırmalar incelendiğinde, genellikle kız ve erkek öğrenci grupları, farklı dil ve kültürleri karşılaştırmak için DMF analizlerinin yapıldığı dikkat çekmektedir. DMF'li maddelerin incelendiği testlerin, daha çok matematik ve fen içerikli olduğu, okuma becerileri ile ilgili daha az araştırmanın yapıldığı saptanmıştır. Bu bağlamda, okuma becerileri testi ile ilgili yapılan bir araştırmanın, DMF çalışmalarına katkı sağlayacağı düşünülmektedir. Buna dayanarak bu araştırmanın problemi, uluslararası düzeyde öğrenci başarısını belirlemeye yönelik uygulanan PISA 2009 okuma becerileri testi maddelerinin DMF içerip içermediğini incelemektir. Bu probleme dayalı olarak, bu araştırmanın amacını PISA 2009 projesi kapsamında uygulanan okuma becerileri testindeki maddelerin Türkiye uygulaması ile ABD

ve Birleşik Krallık (BK) uygulamaları arasında değişen madde fonksiyonu gösterip göstermediğini incelemek ve olası madde yanlılığı nedenlerini ortaya koymak oluşturmaktadır.

YÖNTEM

Araştırma Modeli

PISA 2009 okuma becerileri testindeki maddelerin Türkiye ve ABD, Türkiye ve BK, ABD ve BK arasında DMF gösterip göstermediğinin incelenmesi ve DMF içeren maddelerin olası yanlılık nedenlerinin yargısal yaklaşımlar kullanılarak belirlenmesi amacıyla yapılmış olan bu araştırma, genel tarama modelinde betimsel bir araştırmadır.

Araştırma Grubu

Araştırma grubunu, PISA 2009 okuma becerileri testinde dört ve altı numaralı kitapçıklardaki maddeleri yanıtlayan Türk öğrenci grubundaki 767, ABD öğrenci grubundaki 796, BK öğrenci grubundaki 765 öğrenci oluşturmaktadır (Çizelge 1). DMF analizleri yapılırken karşılaştırma yapılacak grup büyüklüklerinin birbirine yakın olması dikkate alınmıştır.

Çizelge 1. PISA 2009 Okuma Becerileri Testinde Dört Ve Altı Numaralı Kitapçıklardaki Maddeleri Yanıtlayan Öğrencilerin Ükelere Göre Dağılımı

Ülke	Dört Numaralı Kitapçık	Altı Numaralı Kitapçık
Türkiye	382	385
ABD	400	396
BK	382	383

PISA 2009 Türkiye uygulamasında, öğrenciler uyarlaması yapılan testlerdeki maddeleri yanıtlamaları nedeniyle ABD ve BK'ya göre dezavantajlı grup, diğer bir deyişle odak grup olarak alınmıştır. ABD ve BK uygulamaları arasında DMF gösteren maddeler belirlenirken ise ABD odak grup olarak alınmıştır.

Veriler ve Toplanması

PISA 2009 okuma becerileri testinin 13 soru kitapçığına ilişkin veriler OECD tarafından açıklanmıştır. Bu araştırma kapsamında veriler, OECD'nin resmi sitesinden indirilmiş, örnek olarak internet sitesinde açıklanan maddelerin tümünün yer aldığı ve en fazla sayıda madde içeren dört ve altı numaralı kitapçık üzerinden analizler yürütülmüştür. Bu araştırma kapsamında ele alınan dört ve altı numaralı kitapçıkta dokuz açıklanan madde bulunmaktadır: R414Q02, R414Q06, R414Q09, R414Q11, R452Q03, R452Q04, R452Q07, R458Q01 ve R458Q07. OECD'nin resmi internet sitesinde açıklanan İngilizce

ve Türkçe okuma parçaları ve bu okuma parçalarına ilişkin belirtilen maddelerdeki ifadelerin her iki kültür tarafından anlaşılabilirliği, çevirinin niteliği, madde formatı, metnin içeriği açısından uygunluğunun belirlenebilmesi amacıyla araştırmacı tarafından bir anket geliştirilmiş ve uzman görüşüne sunulmuştur.

Verilerin Analizi

Verilerin analizinde, analiz yorumlarının bütün maddeler için aynı olabilmesi için, veriler 1-0 (Tam doğru ve kısmi doğru olan maddeler 1, yanlış ve boş bırakılan maddeler 0) puanlamaya dönüştürülmüştür. Verilerin tek boyutlu olup olmadığı, her grup için (Türkiye, ABD, BK), “Ağırlıklandırılmamış En Küçük Kareler” (Unweighted Least Squares) kestirim yöntemi kullanılarak doğrulayıcı faktör analizi ile test edilmiş ve her ülke için verilerin tek boyutlu olduğu saptanmıştır.

DMF belirleme yöntemleri seçilirken alanyazın taranmış, en sık kullanılan ve güvenilir olan yöntemler olan MH, SIBTEST, MTK-OO ve BILOG-MG DIF teknikleri kullanılmıştır. MH küçük gruplar da dâhil en güçlü DMF belirleme yöntemlerinden biridir. Ayrıca, tek biçimli madde fonksiyonlaşmasının belirlenmesinde etkilidir (Clauser ve Mazor, 1998; Gierl, Khaliq ve Boughton, 1999). SIBTEST birinci tip hatayı kontrol etmede, tek biçimli madde fonksiyonlaşmasını belirlemede uygun bir yöntemdir (Shealy ve Stout, 1993; Clauser ve Mazor, 1998). MTK’ya dayalı analizler, gözlenen puanlar yerine yetenek kestirimleri üzerinden yürütüldüğünden, hem referans hem de odak grup için madde parametreleri karşılaştırılabilir. MTK ve KTK’ya dayalı DMF belirleme yöntemlerinin birbirlerine kıyasla bazı üstünlükleri ve dezavantajları bulunması sebebiyle, DMF gösteren maddelerin belirlenmesi için birden fazla yöntem kullanılması önerilmektedir. Çünkü farklı yöntemler kullanıldığında DMF gösteren maddelerin de farklılaştığı belirlenmiştir. Bu nedenle, en az iki yöntemle DMF gösteren maddenin yanlış olup olmadığı araştırılması gerekir (Osterlind ve Everson, 2009; Camilli ve Shepard, 1994). Bu çalışmada ise dört DMF belirleme yöntemi kullanılmıştır ve en az iki yöntemde DMF gösteren madde, DMF’li madde olarak ele alınmıştır.

PISA 2009 okuma becerileri testinde OECD’nin resmi internet sayfasında örnek olarak açıklanan dokuz maddeden DMF’li olanlar araştırmacı tarafından geliştirilmiş anket ile olası yanlışlık nedenlerinin belirlenebilmesi için uzman görüşüne sunulmuştur. DMF içeren ancak örnek olarak açıklanmamış maddeler, uzman görüşüne sunulmamış, araştırmacı tarafından madde formatı, okurun metne yaklaşımı açısından yanlışlık içirme durumu incelenmiştir.

MH analizi DIFAS 4.0; SIBTEST analizi SIBTEST bilgisayar programı; MTK-OO analizi IRTLRDIF; BILOG-MG DIF analizleri BILOG-MG bilgisayar programı ile yapılmıştır.

BULGULAR

Türkiye - ABD Öğrenci Gruplarında DMF Gösteren Maddelere İlişkin Bulgular

PISA 2009 okuma becerileri testinin dört ve altı numaralı kitapçıklarında bulunan maddelerin Türkiye ve ABD uygulamaları arasında DMF gösterip göstermediği incelenmiş, dört numaralı kitapçıkta 45 maddeden 15 maddenin DMF içerdiği ve bunun kitapçıkta maddelerin %33'ünü oluşturduğu; altı numaralı kitapçıkta 59 maddeden 18 maddenin DMF içerdiği ve bunun kitapçıkta maddelerin %31'ini oluşturduğu sonucuna ulaşılmıştır. Dört numaralı kitapçıkta, bu maddelerin sekizinin ABD'li öğrencilerin, yedisinin ise Türk öğrencilerinin lehine; altı numaralı kitapçıkta ise sekizinin ABD'li öğrencilerin lehine, onunun ise Türk öğrencilerinin lehine olduğu saptanmıştır (Çizelge 2).

Çizelge 2. Türkiye Ve ABD Öğrenci Gruplarında DMF Gösteren Maddelere İlişkin Çizelge

		Dört Numaralı Kitapçık		Altı Numaralı Kitapçık	
		TUR	ABD	TUR	ABD
Madde Formatı	Çoktan Seçmeli	-	2	3	2
	Karmaşık Çoktan Seçmeli	1	1	3	2
	Açık Uçlu	3	4	2	4
	Kısa Cevaplı	-	1	1	-
	Kapalı Uçlu	3	-	1	-
Okurun Metne Yaklaşımı	Bilgiye Ulaşma ve Bilgiyi Hatırlama	3	2	2	1
	Bilgiyi Bir Araya Getirme ve Yorumlama	2	3	6	5
	Kendi Düşüncelerini Yansıtmaya ve Metni Değerlendirme	2	3	2	2

Çizelge 2'de DMF içeren maddelerin ülkelere olan dağılımı incelendiğinde, dört numaralı kitapçıkta, çoğunlukla açık uçlu maddelerin, “bilgiyi bir araya getirme ve yorumlama” ve “kendi düşüncelerini yansıtmaya ve metni değerlendirme” boyutlarında ABD lehine DMF gösterdiği saptanmıştır. Altı numaralı kitapçıkta ise çoktan seçmeli ve karmaşık çoktan seçmeli maddelerde Türk öğrencilerin lehine, açık uçlu maddelerde ise ABD'li öğrencilerin lehine DMF gösteren

maddeler olduğu saptanmıştır. Okurun metne yaklaşımı açısından, “bilgiye ulaşma ve bilgiyi hatırlama” boyutunda Türk öğrenciler lehine DMF içeren maddeler bulunmaktadır. En çok “bilgiyi bir araya getirme ve yorumlama” boyutunda DMF’li madde olduğu, bu maddelerin dört numaralı kitapçıkta ABD’li öğrenciler lehine iken, altı numaralı kitapçıkta Türk öğrenciler lehine olduğu görülmektedir.

Türkiye - BK Öğrenci Gruplarında DMF Gösteren Maddelere İlişkin Bulgular

PISA 2009 okuma becerileri testinin dört ve altı numaralı kitapçıklarında bulunan maddelerin Türkiye ve BK uygulamaları arasında DMF gösterip göstermediği incelenmiş, dört numaralı kitapçıkta 45 maddeden 22 maddenin DMF içerdiği ve bunun kitapçıkta maddelerin %49’unu; altı numaralı kitapçıkta ise 59 maddeden 21 maddenin DMF içerdiği ve bunun kitapçıkta maddelerin %36’sını oluşturduğu sonucuna ulaşılmıştır. Dört numaralı kitapçıkta DMF gösteren maddelerin 13’ünün BK; dokuzunun Türk öğrencilerinin lehine; altı numaralı kitapçıkta ise maddelerin 11’inin BK; onunun ise Türk öğrencilerinin lehine olduğu görülmektedir (Çizelge 3).

Çizelge 3. Türkiye Ve BK Öğrenci Gruplarında DMF Gösteren Maddeler

		Dört Numaralı Kitapçık		Altı Numaralı Kitapçık	
		TUR	BK	TUR	BK
Madde Formatı	Çoktan Seçmeli	2	4	3	2
	Karmaşık Çoktan Seçmeli	1	1	-	4
	Açık Uçlu	4	6	5	3
	Kısa Cevaplı	-	1	2	-
	Kapalı Uçlu	2	1	-	2
Okurun Metne Yaklaşımı	Bilgiye Ulaşma ve Bilgiyi Hatırlama	4	3	3	2
	Bilgiyi Bir Araya Getirme ve Yorumlama	3	6	3	7
	Kendi Düşüncelerini Yansıtırma ve Metni Değerlendirme	2	4	4	2

Çizelge 3’de DMF içeren maddelerin ülkelere olan dağılımı incelendiğinde, dört numaralı kitapçıkta çoktan seçmeli ve karmaşık çoktan seçmeli maddelerin BK lehine; altı numaralı kitapçıkta ise Türk öğrenciler lehine DMF içerdiği görülmektedir. Açık uçlu maddelerin ise dört numaralı kitapçıkta BK öğrencileri

lehine, altı numaralı kitapçıkta Türk öğrenciler lehine DMF gösterdiği saptanmıştır. Okurun metne yaklaşımı açısından ise, “bilgiyi bir araya getirme ve yorumlama” ve “kendi düşüncelerini yansıtmaya” boyutlarındaki maddelerin çoğunlukla BK öğrencileri lehine DMF içerdiği sonucuna ulaşılmıştır.

ABD - BK Öğrenci Gruplarında DMF Gösteren Maddelere İlişkin Bulgular

PISA 2009 okuma becerileri testinin dört ve altı numaralı kitapçıklarında bulunan maddelerin ABD ve BK uygulamaları arasında DMF gösterip göstermediği incelenmiş, dört numaralı kitapçıkta 45 maddeden sekizinin DMF içerdiği ve bunun kitapçıkta maddelerin %18’ini; altı numaralı kitapçıkta ise 59 maddeden yedi maddenin DMF içerdiği ve bunun kitapçıkta maddelerin %12’sini oluşturduğu saptanmıştır. Dört numaralı kitapçıkta bu maddelerin beşinin ABD lehine, üçünün BK lehine; altı numaralı kitapçıkta ise üçünün ABD lehine, dördünün BK lehine olduğu görülmektedir (Çizelge 4).

Çizelge 4. ABD Ve BK Öğrenci Gruplarında DMF Gösteren Maddeler

		Dört Numaralı Kitapçık		Altı Numaralı Kitapçık	
		ABD	BK	ABD	BK
Madde Formatı	Çoktan Seçmeli	2	-	1	-
	Karmaşık Çoktan Seçmeli	-	-	-	4
	Açık Uçlu	3	2	2	-
	Kısa Cevaplı	-	-	-	-
	Kapalı Uçlu	-	1	-	-
Okurun Metne Yaklaşımı	Bilgiye Ulaşma ve Bilgiyi Hatırlama	-	-	-	1
	Bilgiyi Bir Araya Getirme ve Yorumlama	2	2	1	3
	Kendi Düşüncelerini Yansıtmaya ve Metni Değerlendirme	3	1	2	-

Çizelge 4’de DMF içeren maddeler, format açısından incelendiğinde, dört numaralı kitapçıkta, DMF içeren sekiz maddeden beşinin açık uçlu, ikisinin çoktan seçmeli, birinin ise kapalı uçlu olduğu görülmektedir. Altı numaralı kitapçıkta DMF içeren altı maddeden dördünün karmaşık çoktan seçmeli, ikisinin açık uçlu, birinin ise çoktan seçmeli olduğu görülmektedir. Çizelge 4’de okurun metne yaklaşımı açısından DMF içeren maddeler incelendiğinde, dört numaralı kitapçıkta DMF içeren sekiz maddenin “bilgiyi bir araya getirme ve yorumlama” ve “kendi düşüncelerini yansıtmaya ve metni değerlendirme”

boyutunda olduğu; altı numaralı kitapçıkta ise dört maddenin “bilgiyi bir araya getirme ve yorumlama” boyutunda olduğu görülmektedir.

DMF içeren maddelere ilişkin uzman görüşleri

PISA 2009 okuma becerileri testinde örnek olarak açıklanan maddelerden R414Q02, R414Q06 ve R458Q07 kodlu olanların Türk öğrenciler ile BK ve ABD öğrenci grupları arasında DMF gösterdiği saptanmıştır. R414Q02 ve R414Q06 kodlu maddelerde DMF referans grubun lehine iken R458Q07 kodlu maddede DMF odak grubun lehinedir. Ayrıca, BK ve ABD öğrenci grupları arasında da DMF göstermiştir.

PISA 2009 okuma becerileri testindeki açıklanan sorulara ilişkin uzman görüşüne sunulan anket sonuçlarına göre, uzmanlar, açıklanan okuma metinlerinden biri olan “Cep Telefonu Güvenliği” başlıklı metnin Türk öğrencilerinin aşına olduğu bir yapıya sahip olmadığı belirtmişler, “gen dizilimi”, “manyetik”, “SAR değeri” gibi ifadelerle Türk öğrencilerinin aşına olmadığı yönünde ortak görüş bildirmişlerdir. Uzmanlar, ayrıca birebir çeviri niteliğinde olan ve Türk kültürüne uyarlaması yapılmayan bazı ifadelerin metnin okunmasını zorlaştırdığını, teknolojik gelişimin ülkeden ülkeye farklılık göstermesinin öğrencilerin artalan bilgilerinin farklı olmasına neden olacağını, bu durumun da yanlışlık kaynağı olabileceği yönünde görüş bildirmişlerdir.

“Cep Telefonu Güvenliği” başlıklı metne ait R414Q02 kodlu maddeye ilişkin olarak uzmanlar şu şekilde görüş bildirmişlerdir: “Metinde yer alan önemli noktaların amacı nedir?” sorusunun “Metinde yer alan önemli noktalar olarak belirtilen uyarıların amacı nedir?” gibi daha anlaşılır şekilde değiştirilebileceği önerilmiştir. Uzmanlar, Türk öğrencilerinin metin yapısına aşına olmama durumuna dayalı olarak, maddedeki “önemli nokta”nın işlevini anlamamış olabileceklerini, metinde öne sürülen önlemlerden biri olan “radyasyon koruyucu cihazlar”ın, Türkiye’deki kullanımı, satışı ve bilinme oranının, ABD ve BK’dan daha az olmasının ve belirtilen ülkelerdeki teknolojik gelişimin Türkiye’ye kıyasla daha hızlı olmasının bu madde ve seçenek için bir yanlışlık teşkil edebileceğini belirtmişlerdir. Ayrıca aynı metne ait DMF gösteren R414Q06 kodlu maddeye ilişkin uzmanlar şu şekilde görüş bildirmişlerdir: metin içinde, cep telefonuyla uzun konuşmalar yapanların, yorgunluk, baş ağrısı ve konsantrasyon kaybından şikayetçi oldukları, ancak bu şikayetlerin laboratuvar ortamında gözlenmediğini, modern yaşam tarzlarındaki “diğer nedenlerden” kaynaklanabileceğinden bahsedilmiştir. R414Q06 kodlu açık uçlu maddede ise öğrencilerden modern yaşam tarzındaki diğer nedenlerden birinin ne olabileceğinin açıklanması istenmiştir. Uzmanlar, maddenin çevirisinde hiçbir problem olmadığını belirtmişlerdir. Ancak ABD ve BK’ya benzer olarak modern yaşam tarzının Türkiye’deki insanlar üzerindeki etkisi başlamasına karşın, bu etkinin belirtilen diğer ülkelerin düzeyinde görülmediği ve modern

yaşam tarzının tüm gelir düzeylerinde aynı oranda görülmemesinin bir yanlılık kaynağı olabileceği yönünde uzmanlar tarafından görüş bildirmiştir. Ayrıca, “kendi düşüncelerini yansıtmaya ve metni değerlendirme” gibi üst düzey bilişsel beceriyi ölçen bir yaklaşıma göre hazırlanmış açık uçlu bir maddeye Türk öğrencilerinin aşına olmadığı belirtilmiştir. Bu durum da Türkiye’deki sınıf ortamında ve ders kitaplarında okuma metinlerinin bu biçimde yer almadığı, öğrencilerin metinleri bu tarzda incelemeye alışkın olmamaları şeklinde açıklanmıştır.

“Uzaktan Çalışma” metnine dayalı olarak verilen R458Q07 kodlu maddenin çevirisinin yeterli olduğu yönünde uzmanlar ortak görüş bildirmişlerdir. Uzmanların bir bölümü, “uzaktan çalışma”nın Türk kültürü için yabancı bir kavram olduğu, bu durumun da yanlılık yaratabileceğini belirtmişlerdir. Ancak, yapılan analizlerde bu maddenin Türk öğrenciler lehine DMF gösterdiği saptanmıştır.

TARTIŞMA VE SONUÇ

PISA 2009 okuma becerileri testinde Türk ve ABD’li, Türk ve BK’daki öğrenci grupları arasında DMF içeren maddelerin kitapçıklara dağılımı göz önünde bulundurulduğunda, en az altı numaralı kitapçıkta (%31 ve %33) en fazla ise dört numaralı kitapçıkta (%36 ve %49) DMF içeren madde bulunmaktadır. DMF içeren maddelerin odak ve referans gruplar arasında dağılımı incelendiğinde, yaklaşık olarak eşit dağıldığı belirlenmiştir.

DMF içeren maddeler madde formatı açısından incelendiğinde, ABD’li öğrencilerin lehine olan açık uçlu maddeler bulunmaktadır. Bu durum, Ulutaş’ın (2012) PISA 2006 fen okuryazarlığı testinde, ABD ve Türk öğrenci gruplarını karşılaştırdığı araştırmanın sonucu ile paralellik göstermektedir. Altı numaralı kitapçıkta, çoktan seçmeli maddelerin çoğunluğunun Türk öğrencilerin lehine olduğu görülmektedir. Çakan (2004) tarafından yapılan çalışmada, ilköğretimde öğretmenlerin en fazla çoktan seçmeli madde formatını kullandığı, bunun da öğrencilerin ortaöğretime geçiş sınavlarında en çok karşılaştıkları madde formatı olduğunu belirtilmiştir. Mertler (1999) ve Güven (2001) tarafından yapılan araştırmalar da benzer bir şekilde, öğretmenlerin en sık çoktan seçmeli madde formatını kullandıklarını göstermektedir. Bu araştırmaların sonuçları, PISA okuma becerileri testindeki çoktan seçmeli maddelerde Türk öğrencilerin avantajlı olma durumunu açıklamaktadır. Tek bir madde formatı veya değerlendirme şekli öğrencilerde tek yönlü bir çalışma alışkanlığı geliştirmekte, ölçme araçlarında sadece çoktan seçmeli madde formatının kullanılması eğitim sistemi üzerinde olumsuz etki yaratmaktadır (Gültekin ve Demirtaşlı, 2012; Berberoğlu, 2006). Geniş ölçekli testlerdeki madde formatlarının çeşitliliğinin ulusal sınavlarda tercih edilmemesinin bir

sonucu olarak, Türk öğrencilerin açık uçlu maddelerde dezavantajlı olması beklenen bir durum olabilir.

PISA 2009 okuma becerileri testindeki maddeler, metne yaklaşım açısından incelendiğinde, ikinci düzey olan “bilgiyi bir araya getirme ve yorumlama” yaklaşımına göre geliştirilmiş maddelerin çoğunluğunun ABD ve BK lehine; alt düzey olan “bilgiye ulaşma ve bilgiyi hatırlama” düzeyindeki maddelerin çoğunluğunun Türk öğrencilerin lehine DMF gösterdiği saptanmıştır. Alt düzey beceriler genellikle hatırlama ya da ezberleme gerektiren davranışlar içerirken, üst düzey beceriler bireyin birden fazla beceriyi kendi bireysel özellikleriyle birleştirerek kullanmasını kapsar (Kutlu, Doğan ve Karakaya, 2008). Berberoğlu (2009), Öğrenci İzleme Sistemi (ÖİS) çerçevesinde, öğrencilerin sınıf düzeyi arttıkça, daha çok ezberleme stratejisini kullandıklarını ortaya koymuştur. Öğrencilerin test çözmeyi bir öğretim etkinliği olarak görmelerinin sonucu olarak, konuları kavramak yerine, maddelerin çözüm algoritmalarını “ezberleme” yoluna gittikleri sonucuna ulaşmıştır. Bu sonuç, PISA okuma becerileri testinde, Türk öğrencilerin alt düzey bilişsel stratejileri içeren maddelerde neden avantajlı olduğunu açıklamaktadır.

BK ve ABD öğrenci grupları arasında DMF gösteren maddelerin dört numaralı kitapçığın %18’ini, altı numaralı kitapçığın ise %12’sini oluşturduğu belirlenmiştir. Türk öğrenci grubu ile BK ve ABD öğrenci grupları ile yapılan DMF analizleri kıyaslandığında, aynı dilde testi alan öğrenciler arasında DMF gösteren maddelerin oldukça az olduğu göze çarpmaktadır. Diğer bir anlatımla, dil ve kültür farklılığı arttıkça DMF gösteren maddeler artmaktadır. Elde edilen bulgular, Asil (2010), Atalay (2010) ve Başusta (2013) tarafından yapılan araştırmaların bulguları ile paralellik göstermektedir.

Uzman görüşleri incelendiğinde, örnek olarak açıklanan maddeler arasında DMF gösteren maddelerin yanlışlık kaynakları, sözcüklerin güç anlaşılması, madde formatına aşina olma durumu, çeviri hataları ve kültürel farklılıklar olduğu sonucuna ulaşılmıştır. Bu araştırmanın bulguları, daha önce farklı dil ve kültürleri karşılaştıran DMF çalışmalarının bulguları ile örtüşmektedir (Allaouf ve Sireci, 1998; Gierl ve Khaliq, 2000; Grisey ve Monseur, 2007; Yıldırım, 2008; Yıldırım ve Berberoğlu, 2009; Ulutaş, 2012).

Bu araştırmadan elde edilen bulgular, Türk öğrencilerin dezavantajlı olduğu açık uçlu ve üst düzey bilişsel görevler içeren maddeler olduğunu göstermektedir. Türk öğrencilerinin madde formatlarına aşina olmaması ve üst düzey bilişsel görevler içeren maddelerle çok fazla karşılaşmaması uluslararası düzeyde yapılan bir karşılaştırma sınavında Türk öğrencilerin dezavantajlı olarak görünmesine neden olmaktadır. Elde edilen bu bulgulara dayalı olarak, eğitimcilerin, derslerde ve sınavlarda öğrencilerin daha fazla üst düzeyde bilişsel görevleri içeren ve çeşitli madde formatları içeren farklı türlerdeki okuma metinleriyle karşılaşmalarını sağlamaları, bu yönde eğitim programlarının düzenlenmesi

yönünde çalışmalar yapmaları, ayrıca PISA projesi kapsamında uzmanların okuma parçaların yazımında eğitim programlarının ve kültürler arası farkları göz önüne almaları önerilebilir.

KAYNAKLAR

- Acar, T. (2008). *Maddenin farklı fonksiyonlaşmasını belirlemede kullanılan genelleştirilmiş aşamalı doğrusal modelleme, lojistik regresyon ve olabilirlik oran tekniklerinin karşılaştırılması*. (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Aiken, L.R. (1997). *Psychological testing and assessment*. Boston: Allyn and Bacon.
- Allouf A. ve Sireci, S. G. (Nisan-1998). Detecting sources of DIF in translated verbal items. *Annual Meeting of the American Educational Research Association'da Sunulan Sözlü Bildiri*. San Diego.
- Angoff, W. H. (1993). Perspectives on differential item functioning methodology. Holland P.W. ve Wainer H., (Ed.) *Differential item functioning*. Lawrence Erlbaum Associates Publishers: New Jersey.
- Asil, M. (2010). *Uluslararası öğrenci değerlendirme programı (PISA) 2006 öğrenci anketinin kültürler arası eşdeğerliğinin incelenmesi*. (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Atalay, K. (2006). *PISA 2006 öğrenci anketinde yer alan tutum maddelerinin değişen madde fonksiyonu açısından incelenmesi*. (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Başusta, B.U. (2013). *PISA 2006 fen başarı testinin madde yanlılığının kültür ve dil açısından incelenmesi*. (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Berberoğlu, G. (1995). Üniversite birinci basamak ÖSS testinin madde yanlılığı açısından incelenmesi. II. *Eğitim Bilimleri Kongresi'nde Sunulmuş Sözlü Bildiri*, Hacettepe Üniversitesi, Ankara.
- Berberoğlu, G. (1996). The university entrance examination in Turkey. *Studies In Educational Evaluation*. 21, 363 – 373.
- Berberoğlu, G. (2006). *Sınıf içi ölçme ve değerlendirme teknikleri*. İstanbul: Morpa Kültür Yayınları.
- Berberoğlu, G. (2009). CITO Türkiye Öğrenci İzleme Sistemi (ÖİS) Öğrenci Sosyal Gelişim Programı'na (ÖSGP) ilişkin ön bulgular. *CITO Eğitim: Kuram ve Uygulama Dergisi*, Kasım-Aralık, 32-42.
- Camilli, G. ve Shepard, L. A. (1994). *Methods for identifying biased test Items*. ABD: Sage Publications.
- Camilli, G. (2006). *Test fairness*. In R. L. Brennan (Ed.), *Educational Measurement*. Westport: American Council on Education & Praeger Publishers.
- Clauser, B. E. ve Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement Issues and Practice*, 17, 31 – 44.

- Cronbach, L. (1990). *Essentials of psychological testing*. (Fifth Edition). New York: Harper and Row.
- Çakan, M. (2004). Öğretmenlerin ölçme ve değerlendirme uygulamaları ve yeterlik düzeyleri: İlk ve orta öğretim. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*. 37, 99-114.
- Çepni, Z. (2011). *Değişen madde fonksiyonlarının SIBTEST, mantel haenzsel, lojistik regresyon ve madde tepki kuramı yöntemleriyle incelenmesi*. (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Deveci, N. A. (2008). *Üniversitelerarası kurul yabancı dil sınavının madde yanlılığı bakımından incelenmesi*. (Yayınlanmamış Doktora Tezi) Ankara Üniversitesi, Ankara.
- Dorans, N. J. ve Holland, P. W. (1993). DIF detection and description: Mantel haenzsel and standardization. P.W.Holland & H. Wainer (Ed.) *Differential item functioning: Theory and Practice*. Hillsdale, NJ: Erlbaum.
- Ellis, B. B. ve Raju, N. S. (2003). *Test and item bias: What they are, what they aren't and how to detect them*. Web: <http://www.eric.ed.gov/PDFS/ED480042.pdf> adresinden 8 Mart 2011 tarihinde alınmıştır.
- Gierl, M.J. ve Khaliq, S.N. (2000). *Identifying sources of differential item functioning on translated achievement tests: A confirmatory analysis*. The annual meeting of the National Council on measurement in education sunulmuş sözlü bildiri.
- Gierl, M., Khaliq, S. N. ve Boughton K. (1999). Gender differential item functioning in mathematics and science: Prevalence and policy implications. "Improving Large Scale Assessment In Education" Adlı Sempozyumda Sunulmuş Sözlü Bildiri. Sherbrook, Québec.
- Grisay, A. ve Monseur, C. (2007). Measuring the equivalence of item difficulty in the various versions of an international Test. *Studies in Educational Evaluation*, 33, 69-86.
- Gültekin, S. ve Demirtaşlı, N. Ç. (2012). Çoktan seçmeli, açık uçlu ve karma testlerden sağlanan bilginin madde tepki kuramına dayalı olarak karşılaştırılması. *İlköğretim Online*, 11(1), 251-263, 2012.
- Güven, S. (Haziran,2001). Sınıf öğretmenlerinin ölçme ve değerlendirmede kullandıkları yöntem ve tekniklerin belirlenmesi. 10. *Ulusal Eğitim Bilimleri Kongresinde sunulmuş bildiri*, Abant İzzet Baysal Üniversitesi, Bolu.
- Holland, P. W. ve Thayer, D. T. (1988). *Differential Item performance and the mantel-haenzsel procedure*. H. Wainer ve H.I. Braun (Ed.), *Test validity*. Hillsdale, NJ: Erlbaum.
- Kalaycıoğlu, D. B. (2008). *Öğrenci seçme sınavının madde yanlılığı açısından incelenmesi*. (Yayınlanmamış Doktora Tezi), Hacettepe Üniversitesi, Ankara.
- Karakaya, İ. ve Kutlu, Ö. (2012). Seviye belirleme sınavındaki Türkçe alt testlerinin madde yanlılığının incelenmesi. *Eğitim ve Bilim*, 37 (165), 348 – 362.
- Kutlu, Ö., Doğan C. D. ve Karakaya, İ. (2008). *Öğrenci başarısının belirlenmesi*. Ankara: Pegem Akademi Yayıncılık.
- Linn, R. L. ve Gronlund, N. E. (2000). *Measurement and assessment in teaching*. (Eighth Edition) Upper Saddle River, New Jersey.

- Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, 120 (2), 285-297.
- Osterlind, S. J., ve Everson, H. T. (2009). *Differential item functioning*. ABD: Sage Publications, Inc.
- Öğretmen, T. (1995). *Differential item functioning analysis of the verbal ability section of the first stage of the university entrance examination in Turkey*. (Yayınlanmamış Yüksek Lisans Tezi), Ortadoğu Teknik Üniversitesi, Ankara.
- Öğretmen, T. Ve Doğan, N. (2004). OKÖSYS matematik alt testine ait maddelerin yanlılık analizi. *Malatya İnönü Eğitim Fakültesi Dergisi*. 5 (8) 61 -76.
- Shealy, R. ve 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*, 58, 159-194.
- Schulz, W. (Nisan,2003). Validating questionnaire constructs in international studies. two examples from PISA 2000. *Annual Meetings of the American Educational Research Association (AERA)* Sunulmuş sözlü bildiri, Chicago.
- Ulutaş, S. (2012). *PISA 2006 fen okuryazarlığı testindeki maddelerin yanlılık bakımından araştırılması*. (Yayınlanmamış doktora tezi), Ankara Üniversitesi, Ankara.
- Yenal, E. (1995). *Differential item functioning analysis of the quantitative ability section of the first stage of the university entrance examination in Turkey*. (Yayınlanmamış Yüksek Lisans Tezi), Ortadoğu Teknik Üniversitesi, Ankara.
- Yıldırım, H. H. (2006). *The differential item functioning (DIF) analysis of mathematics items in the international assessment programs*. (Yayınlanmamış Doktora Tezi), Ortadoğu Teknik Üniversitesi, Ankara.
- Yıldırım, H. H. ve Berberoğlu, G. (2009). Judgmental and statistical DIF analyses of the PISA 2003 mathematics literacy items. *International Journal of Testing*, 9,108-121.
- Yurdugül, H. (2003). *Ortaöğretim kurumları seçme ve yerleştirme sınavının madde yanlılığı açısından incelenmesi*. (Yayınlanmamış Doktora Tezi), Hacettepe Üniversitesi, Ankara.
- Wu, A. D. ve Ercikan, K. (2006). Using multiple-variable matching to identify cultural sources of differential item functioning, *International Journal of Testing*, 6 (3), 287-300.

EXTENDED ABSTRACT

Educational and psychological tests are administered in a wide range of contexts in contemporary society such as at schools, industries, counseling centers and research contexts to evaluate skills, behaviours in order to assist in making judgements, predictions and decisions about people. The accuracy of these judgements depend on the validity and reliability of tests. For test users, the results of tests enable them to compare the test takers in terms of traits, such as achievement. However, bias in testing occurs when a decision grounded to the scores obtained from a test is considerably unfair for one group to another. Differential item functioning (DIF) is the difference in the way an item functions across different groups which are matched on the attribute measured by the item. There are many studies which aim at determining in the exams both at national and international level. In these research, the variables such as culture, language and gender have mainly been studied in the context of DIF. The aim of this research is to determine the items indicating DIF and its possible sources of bias among Turkey, the United States of America (USA) and the United Kingdom (UK) at Booklet 4 and Booklet 6 in PISA 2009 reading test. There are 13 booklets in PISA 2009, only nine items were revealed. For this research, the booklets including the most revealed items were selected. There are various statistical techniques for detecting the systematic differences in subgroup's performance. The literature on statistics was reviewed and the most commonly used DIF methods were selected for the research. The methods of Mantel Haenszel (MH), Simultaneous Item Bias Test (SIBTEST), Item Response Theory Likelihood Ratio (IRTLR) and comparison of b parameters have been used to identify DIF. The research group in this study involves 767 pupils from Turkey, 796 pupils from the USA and 765 pupils from the UK answering the questions in Booklet 4 and Booklet 6 of PISA 2009 reading test. The distributions of between-groups items indicating DIF in accordance with the DIF method have been investigated with regards to item format, aspect and the advantageous group.

The results indicate that there are less items with DIF in Booklet 6 (31% and 33%) and there are more items with DIF in Booklet 4 (36% and 49%) between Turkey and the USA, Turkey and the UK. Also, the distribution of items with DIF is nearly equal between reference and focal groups. However, there are less items with DIF between the USA and the UK (18% and 12%) when it is compared with Turkey and the reference groups. This significant finding indicates that the more there are differences in language and culture, the more items with DIF in a test occur. In the context of item format, it has been determined that there are some open-ended items with DIF being in favour of reference groups. Also, there are some multiple choice items with DIF being in favour of Turkish students. Besides, from the point of text analysis, the items

developed with “integrate and interpret” aspect are mostly also in favour of reference groups. The items developed with “access and retrieve” are mainly in favour of Turkish students. The sources of bias have been identified as hardly understood, being familiar with item format, translation errors and cultural differences by the experts. On the grounds of findings of this research, it can be concluded that educators must provide students more diversity of item types, especially open ended; and also tasks requiring higher order thinking skills.

YAZAR HAKKINDA

Dr. Deniz Tuğçe Özmen, Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Bölümünde araştırma görevlisi olarak çalışmaktadır. Yazar, doktora derecesini, 2013 yılında Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Ölçme ve Değerlendirme Anabilim Dalı'nda almıştır. / İletişim adresi: Ankara Üniversitesi Eğitim Bilimleri Fakültesi Cebeci, Ankara / Eposta: tozmen@ankara.edu.tr

ABOUT THE AUTHOR

Dr. Deniz Tuğçe Özmen has been working as a research assistant in the department of Measurement and Evaluation in the Faculty of Educational Sciences at Ankara University. The author got the PhD degree from the same department at the Institute of Educational Sciences in 2013. / Correspondence address: Ankara University Faculty of Educational Sciences Cebeci, Ankara / Email: tozmen@ankara.edu.tr
