



An analysis of the differential item function through Mantel-Haenszel, SIBTEST and Logistic Regression Methods

Mantel-Haenszel, SIBTEST ve Lojistik Regresyon Yöntemleri ile değişen madde fonksiyonunu analizi¹

Süleyman Demir²
İbrahim Alper Köse³

Abstract

This study performs a Differential Item Function (DIF) analysis in terms of gender and culture on the items available in the PISA 2009 mathematics literacy sub-test. The DIF analyses were done through the Mantel Haenszel, Logistic Regression and the SIBTEST methods. The data for the gender variable were collected from the responses given by 332 students to the items in the mathematics literacy sub-test during the administration of the 5th booklet in the PISA 2009 application whereas the data for the culture variable were collected through the application of the 5th booklet in Turkey, Germany, Finland and the United States in the PISA 2009 application. As a result of DIF analysis according to gender, 4 items carried out in favor of men, only one item can be said to be advantageous in favor of girls. As a result of DIF analysis according to culture, 16 items for

Özet

Bu çalışmada, PISA 2009 uygulaması Matematik okuryazarlığı alttestinde, cinsiyete ve kültüre göre değişen madde fonksiyonu (DMF) içeren madde bulunup bulunmadığı incelenmiştir. DMF analizleri Mantel Haenszel, Lojistik Regresyon ve SIBTEST yöntemleriyle gerçekleştirilmiştir. Çalışmanın veri setini, cinsiyet değişkeni için PISA 2009 uygulamasında Türkiye’de 5. kitapçığındaki matematik okuryazarlığı alt testindeki maddelere 332 öğrencinin verdikleri cevaplar, kültür değişkeni için ise PISA 2009 uygulamasında Türkiye, Almanya, Finlandiya ve ABD’deki öğrencilerin 5. kitaptaki matematik testine verdikleri cevaplar oluşturmaktadır. Cinsiyet değişkenine göre 5 maddenin 4’ü erkekler lehine, bir maddenin kızlar lehine avantajlı olduğu söylenebilir. DMF analizi sonucunda Türkiye ile Almanya’dan katılan öğrencilerden elde edilen veri setlerinde

¹Bu çalışma “PISA 2009 Matematik Okuryazarlığı Alt Testinde Bulunan Maddelerinin Mantel-Haenszel, SIBTEST ve Lojistik Regresyon Yöntemleri ile Değişen Madde Fonksiyonunun İncelenmesi” isimli tez çalışmasından özetlenerek hazırlanmıştır.

²Research Assistant, Sakarya University, Faculty of Education, Educational Sciences Department, suleymand@sakarya.edu.tr

³Assist. Prof. Dr., Abant İzzet Baysal University, Faculty of Education, Educational Sciences Department, i.alper.kose@gmail.com

Turkish and German students, 14 items for Turkish and Finn students, 18 items for Turkish and United States students were determined.

Keywords: Differential item functioning, PISA, Mantel-Haenszel, logistic regression, SIBTEST

[\(Extended English abstract is at the end of this document\)](#)

16, Türkiye ile Finlandiya'dan katılan öğrencilerden elde edilen veri setlerinde 14, Türkiye ile ABD'den katılan öğrencilerden elde edilen veri setlerinde 18 maddede DMF bulunduğu görülmüştür.

Anahtar Kelimeler: Değişen madde fonksiyonu, PISA, Mantel-Haenszel, lojistik regresyon, SIBTEST

Giriş

Millî Eğitim Bakanlığı eğitim sistemiyle ilgili sağlıklı ve karşılaştırmalı veriler almak amacıyla ulusal ve uluslararası düzlemde periyodik olarak, ölçme-değerlendirme çalışmaları gerçekleştirilmektedir. Eğitimi Araştırma ve Geliştirme Daire Başkanlığı (EARGED) tarafından yürütülen ulusal düzeydeki “Öğrenci Başarılarını Belirleme Sınavı (ÖBBS)” çalışması ve uluslararası ölçme-değerlendirme çalışmaları olan TIMSS, PIRLS ve PISA projeleri eğitim sistemiyle ilgili veriler sunmaktadır. Bu çalışmalarda öğrencilere yönelik standart başarı testleri ile öğrenci, öğretmen ve okul anketleri uygulanmakta; buradan elde edilen bilgilerle temel eğitimi tamamlayan öğrencilerin neleri bildikleri, hangi becerileri ne derecede kazandıkları ölçülmeye çalışılmaktadır. Çalışmalardan elde edilen sonuçlarla eğitim sisteminin mevcut durumu tespit edilmekte; özellikle temel eğitimde eğitim politikaları, öğretim programları, öğretim yöntem ve teknikleri, öğretmenlerin yeterlilikleri, ders araç gereç ve materyalleri gibi unsurların gözden geçirilmesine yönelik bilimsel veriler elde edilmekte, diğer ülkelerin eğitim sistemleriyle de karşılaştırmalar yapılmaktadır. Elde edilen bilgiler, eğitimde politika belirleyicilerin, karar vericilerin, akademisyenlerin ve üniversitelerin, eğitim programcılarının, eğitim yöntem ve teknikleri ile eğitim materyali hazırlayanların, eğitimle ilgilenen STK'ların ve araştırmacıların hizmetine sunulmaktadır (EARGED, 2011).

Ekonomik işbirliği ve Kalkınma Teşkilatı-OECD (Organisation for Economic Co-Operation and Development) tarafından düzenlenen PISA; Matematik, Fen ve Okuma becerilerini ölçmeyi amaçlayan okuryazarlık alt testlerinden oluşmaktadır. Üç yılda bir yapılan bu araştırmalarla, OECD üyesi ülkeler ve diğer katılımcı ülkelerdeki (dünya ekonomisinin yaklaşık olarak %90'ı) 15 yaş grubu öğrencilerin modern toplumda yerlerini alabilmeleri için gereken temel bilgi ve becerilere ne ölçüde sahip oldukları değerlendirilmektedir (OECD, 2009).

PISA'nın değerlendirme çerçevesi ve kavramsal temelleri, projeye katılan ülkelerdeki uzmanlar tarafından belirlenmekte, yapılan görüşmeler sonrasında katılımcı ülkelerin görüş birliğiyle onaylanmaktadır. Bu çerçevede, “okuryazarlık” kavramına ilişkin yeni bir anlayış ortaya çıkmıştır. Okuryazarlık kavramı; öğrencilerin bilgilerini günlük yaşamda kullanmak, mantıksal çıkarımlar

yapmak, çeşitli durumlarla ilgili problemleri yorumlamak ve çözmek için öğrendiklerinden çıkarımlar yapma kapasitesi olarak tanımlanmaktadır (EARGED, 2010).

Başlangıçta OECD'ye üye ülkelerin katıldığı PISA, günümüzde dünya genelinde yapılan en önemli araştırmalardan biri haline gelmiştir. PISA 2009' a, 65 katılımcı ülkeden, 15 yaş grubu yaklaşık 26 milyon öğrenciyi temsilen; 475.460 öğrenci katılmıştır. Bu değerlendirmeye daha sonra ek olarak OECD üyesi olmayan 9 ülkeden, 15 yaş grubu yaklaşık 2 milyon öğrenciyi temsilen; 50.000 öğrenci daha katılmıştır (OECD, 2009).

Tüm ölçme araçlarında olduğu gibi PISA uygulamasında kullanılan ölçme araçlarında da bulunması gereken nitelikler geçerlik güvenirlik ve kullanılabilirlik. Geçerlik kavramıyla bir ölçme aracının, onunla ölçülmek istenen değişkeni ölçüp ölçmediği; ölçebiliyorsa onu başka değişkenlerden ne derece arınık olarak ölçtüğü ifade edilir (Turgut, 1997). Yapılan uygulamanın tüm ülkelerde geçerli ve güvenilir olmasını sağlamak, kültürel ve dilsel farklılıkları en aza indirmek için EOCED tarafından oldukça yoğun bir çaba ve kaynak sarf edilmektedir (EOCD, 2009). Ölçme aracından kaynaklanan ve geçerliği etkileyen önemli tehditlerden birisi de madde yanlılığıdır.

Genel anlamdaki tanımıyla madde yanlılığı; testin şartlarından ya da testteki maddelerin bazı özelliklerinden dolayı, maddeyi belirli bir grubun doğru yanıtlama olasılığının diğer gruba göre daha az ya da çok olmasıdır (Zumbo, 1999; Cleary, 1968; Mantel ve Haenszel, 1959; Camilli ve Shepard, 1994; Penfield ve Camilli, 2007). Madde yanlılığının belirlenmesi, test geçerliğini ve güvenirliğini artırmak için yapılan önemli çalışmalardan biridir. Madde yanlılığı çalışmaları madde fonksiyonunun aynı yetenek düzeyinde ve farklı gruplarda yer alan bireyler için farklılaşıp farklılaşmadığını belirleyen istatistiksel bir süreçle başlar. Aynı yetenek düzeyindeki alt gruplarda fonksiyonu farklılaşan maddeler, değişen madde fonksiyonu (DMF) gösteren maddeler olarak adlandırılır (Dorans ve Holland, 1993).

Madde Yanlılığı ve DMF

DMF içeren maddelerin belirli bir gruba veya gruplara avantaj sağladığı göz önüne alındığında DMF'li maddeler ölçme sonucunda sistematik hataya sebep olmaktadır. Bu bağlamda bir testte DMF içeren maddelerin bulunması o testin geçerli olmasını negatif olarak etkileyecektir. Ölçme sonuçları, sınava giren farklı gruplara karşı yansızlık özelliğini sağlamalıdır. Madde yansızlığı sadece teknik olarak geçerlikle yakından ilgili değil, aynı zamanda politik, felsefi, ekonomik, sosyal ve yasal boyutları da olan bir sorundur (Clauser ve Mazor, 1998). DMF maddelerin iki farklı grup için nasıl işleyiş gösterdiğini belirlemek için yol göstermektedir. İki grup arasındaki farklılığın grupların özelliklerinden mi yoksa maddelerin işleyişinden mi kaynaklandığını belirlemek ölçme ve değerlendirme için ayırt edilmesi gereken bir problemdir (Dorans ve Holland, 1993). DMF'nin tek

biçimli (uniform) ve tek biçimli olmayan (nonuniform) olmak üzere iki çeşidi vardır. Tek biçimli DMF yetenek düzeyleri ile grup arasında hiçbir etkileşim olmadığı durumlarda, tek biçimli olmayan DMF ise yetenek düzeyleri ile gruplar arasında bir etkileşim olduğu durumlarda ortaya çıkan DMF için kullanılmaktadır (Swaminathan ve Rogers, 1990). Diğer bir deyişle bir madde bütün yetenek düzeylerinde bir gruba avantaj sağlıyorsa tek biçimli DMF, eğer belirli bir yetenek düzeyine kadar bir gruba diğer yetenek düzeylerinde farklı gruba avantaj sağlıyorsa tek biçimli olmayan DMF olarak isimlendirilir.

Araştırmanın amacı

Testler veya anketler bir dilden başka bir dile uyarlandığında farklı dillerdeki formlarının eşdeğerliğinin sağlandığından emin olunmalıdır. Uyarlanan veya çevirisi yapılan her bir form farklı bir yapıyı ölçebilir, zorluk ya da aşinalık dereceleri bakımından farklılık gösterebilir (Asil ve Gelbal, 2012). Ayrıca uygulanacak ölçeklerin veya anketlerdeki maddeler öğrencilerin bireysel özelliklerini de dikkate alarak hazırlanmalıdır. Bu çalışmada 2009 yılı PISA uygulaması Matematik okuryazarlığı alt testindeki maddeler için cinsiyete göre DMF analizleri yapılmıştır. PISA uygulamasında kullanılan sorulara ulaşım sağlanamadığı için, çalışma sadece maddelerin farklı gruplar için değişen şartlarda işleyip işlemediği incelenecek madde yanlılığı çalışmasının ikinci aşaması olan uzman görüşü yapılmayacaktır. Bu çalışma ile testlerin geçerliğini artırmak amacıyla kullanabilecek bilgiler sağlanması amaçlanmıştır.

Yöntem

Araştırma Türü

Bu çalışma, değişen madde fonksiyonunu belirlemede kullanılan Mantel-Haenszel, lojistik regresyon ve SIBTEST tekniklerinin karşılaştırılmasına yöneliktir. PISA 2009 uygulaması Matematik testindeki maddeler üzerinden elde edilen veri cinsiyet ve kültür değişkenlerine göre inceleme gerçekleştirilecektir. Araştırma, var olan durumun ortaya koymaya yönelik bir çalışma olduğu için betimsel bir çalışmadır.

Çalışma grubu

Çalışmanın veri setini, cinsiyet değişkeni için PISA 2009 uygulamasında Türkiye’de 5. kitapçığın uygulandığı 332 öğrencinin matematik okuryazarlığı alt testindeki maddelere verdikleri cevaplar, kültür değişkeni için ise PISA 2009 uygulamasında Türkiye, Almanya, Finlandiya ve ABD’deki öğrencilerin 5. kitaptaki matematik testine verdikleri cevaplar oluşturmaktadır.

Veri toplama araçları

PISA uygulamasına katılan öğrenciler, uygulama kapsamında yer alan bütün maddeleri yanıtlamazlar. PISA 2009’da kullanılan üniteler, 13 madde demeti halinde gruplandırılmıştır. Her bir madde demetinin yanıtlanma süresi 30 dakikadır. Uygulamaya katılan her ülkede 7 okuma becerileri, 3 matematik ve 3 de fen madde demeti bulunmaktadır. Bu madde demetleri belirli bir döngü düzenine göre 13 kitapçığa yerleştirilmiştir. Her bir kitapçıkta 4 madde demeti yer alır ve her öğrenci seçkisiz yöntemle belirlenen 13 kitapçıktan birini yanıtlar. Bu şartlar altında, en fazla matematik sorusu bulunan ve 2 matematik, 2 okuma becerileri testi bulunan 5 numaralı kitapçık analiz edilmek için kullanılmıştır.

Mantel-Haenszel Yöntemi

Çalışma kapsamına alınan bireyler değişken açısından referans ve odak gruplarına ayrılırlar. Referans ve odak grupları yetenek veya yeterlik düzeylerine göre (genel olarak 4 veya 5 olmak üzere) tabakalara ayrılırlar (Osterlind ve Everson, 2009). Mantel ve Haenszel tarafından 1959 yılında oluşturulan olasılık tablosu tabakalara ayrılmış grupların karşılaştırılmasını kolaylaştırmıştır. Holland (1985) ve Holland ve Thayer (1986) tarafından yapılan çalışmalar sonucunda ise Mantel ve Haenszel tarafından oluşturulan bu olasılık tablosunun DMF belirleme çalışmalarında da kullanılabileceğini göstermişlerdir. Mantel-Haenszel yöntemi referans ve odak grupları için Odds-Oranının bir kestiricisi olarak kestirim katsayısını hesaplamaktadır. i yetenek düzeylerini göstermek üzere, kestirim katsayısı referans ve odak grubundaki her bir yetenek düzeylerindeki doğru ve yanlış yapan öğrenci sayılarına göre hesaplanır.

MH katsayısının daha kolay yorumlanabilmesi için logaritmik bir dönüşüm ile Δ_{MH} katsayısı elde edilmiştir. Bu katsayı ise; $\Delta_{MH} = 2,35 \ln(\alpha_{MH})$ şeklinde ifade edilmektedir. Oluşturulan bu yeni denklem için Δ_{MH} değerinin 0’a eşit olması DMF’nin olmadığını göstermektedir. Eğer Δ_{MH} değeri pozitif değer almışsa odak grubu için, negatif değer almışsa referans grubuna avantaj sağlayacak DMF bulunduğu tespit edilmiş olur (Holland ve Thayer, 1986; Holland ve Wainer, 1993; Osterlind ve Everson, 2009).

MH analizi sonucunda maddeler içerdikleri DMF oranına göre 3 kategoride incelenebilir (Zieky, 1993). Bu kategoriler Tablo 1 de özetlenmiştir. Tablo 1. Zieky (1993) tarafından düzenlenen sınıflama sistemi

DMF Düzeyi	Açıklama	Değer Aralığı
A	DMF yok veya ihmal edilebilir.	$ \Delta_{MH} < 1,0$
B	Orta Düzeyde DMF vardır.	$1,0 < \Delta_{MH} < 1,5$
C	Yüksek Düzeyde DMF vardır.	$ \Delta_{MH} \geq 1,5$

Lojistik Regresyon

Mantel-Haenszel yöntemi pratik ve kullanımının kolay olmasına rağmen Tek Biçimli ve Tek Biçimli olmayan DMF'leri belirleme konusunda yetersiz kalmaktadır. Swaminathan ve Rogers 1990 yılında yaptıkları çalışmada Mantel-Haenszel ve MTK'ya dayalı DMF belirleme yöntemlerine alternatif ve daha etkili olacak bir yöntem geliştirmişlerdir. Bu yöntem Lojistik Regresyon modeli üzerinden analiz yapmayı gerektirmektedir. Bu yöntem ile analizleri yapılan maddelerin hangi tür DMF içerdiklerine ilişkin bilgi de elde edilmiş olur. Lojistik regresyon analizinde madde bağımlı değişken, gruplar ve toplam puan bağımsız değişkenlerdir. DMF belirleme çalışmalarında değişkenler modele hiyerarşik olarak ilave edilirler. SPSS programında lojistik regresyon analizi işlemi yapılırken Model 1'de toplam puan, Model 2'de ise grup değişkeni dâhil edilir. Model 3'te ise toplam puan ile grup değişkenleri birlikte ilave edilmektedir.

Bakan Kalaycıoğlu ve Kelecioğlu (2011)'nin yapmış oldukları çalışmada yöntemlerin örneklem büyüklüklerinden etkileniyor olması ve farklı sınıflama sistemlerinde ortak bir görüş sağlanamamasından dolayı daha hassas bir sınıflama sistemi kullanmışlardır (Tablo 2). Bu çalışmada yapılan bu sınıflama sistemi kullanılmıştır.

Tablo 2. Bakan Kalaycıoğlu ve Kelecioğlu (2011) tarafından düzenlenen sınıflama sistemi

DMF düzeyi	Açıklama	Değer Aralığı
A	DMF yok veya ihmal edilebilir.	$\Delta R^2 < 0,010$
B	Orta Düzeyde DMF vardır.	$0,010 \leq \Delta R^2 < 0,020$
C	Yüksek Düzeyde DMF vardır.	$0,020 \leq \Delta R^2$

SIBTEST

Shealy ve Stout tarafından 1993 yılında geliştirilen SIBTEST yöntemi ile madde yanlılığı ve test yanlılığı aynı anda belirlenebilmektedir. Ayrıca bu yöntemle göre DMF belirlemesi maddelerin etkilerinden bağımsız olarak yapılmaktadır. Bunun için öncelikli olarak tüm maddeler kullanılarak karşılaştırılma yapılır. Bu karşılaştırma sonucunda maddeler DMF içerdiği belirlenen maddeler ve diğerleri olmak üzere iki ayrı gruba ayrılırlar. Maddelerin son olarak karşılaştırılması DMF içermeyen maddelerden elde edilen toplam puanlar üzerinden yapılmaktadır (Clauser ve Mazor, 1998). Roussos ve Stout (1996) tarafından hazırlanan sınıflama Tablo 8' de verilmiştir.

Tablo 3. Roussos ve Stout (1996) tarafından düzenlenen sınıflama sistemi

DMF düzeyi	Açıklama	Değer Aralığı
A	DMF yok veya ihmal edilebilir.	$\beta < 0,059$
B	Orta Düzeyde DMF vardır.	$0,059 \leq \beta < 0,088$
C	Yüksek Düzeyde DMF vardır.	$0,088 \leq \beta$

Bulgular ve Yorum

Bu bölümde cinsiyet ve kültür bakımından PISA 2009 matematik alt testinde bulunan maddelerde DMF olup olmadığı ele alınmıştır. Bu amaçla kullanılan MH, LR ve SIBTEST sonuçları verilmiş ve bu tekniklerden elde edilen sonuçlar ve sonuçlar arasındaki uyum incelenmiştir.

PISA 2009 Matematik okuryazarlığı alt testinde bulunan maddelerde cinsiyet açısından MantelHaenszel yöntemine göre DMF gösteren madde var mıdır?

Cinsiyet değişkeni için Mantel-Haenszel yöntemine göre maddelerin incelenmesi yapılırken öncelikli olarak %95 anlamlılık düzeyine göre p değerlerine bakılmıştır. $p < .05$ olarak bulunan maddelerin Δ -MH değerlerine bakılır ve Zieky (1993) tarafından düzenlenen sınıflama sistemine göre DMF düzeyleri belirlenmiştir. Maddelerin Δ -MH katsayısının almış olduğu değer negatif ise referans grubu için pozitif ise odak grubu için bir avantaj sağlamaktadır.

Tablo 4. DMF içeren maddeler, düzeyleri ve avantaj sağladığı grup ile ilgili bilgiler

Madde	Alfa	Δ -MH	SE (MH D-DIF)	DMF düzeyi	Avantaj sağladığı grup
1	0.609	1.167	0.588	B	Odak (Erkek)
11	2.263	-1.919	0.758	C	Referans (Kız)

MH yöntemine göre kız ve erkeklere farklı işleyen birer tane madde bulunmuştur. Kızlara avantajlı olacak şekilde işleyen madde C düzeyinde, erkeklere avantajlı olacak şekilde işleyen madde ise B düzeyindedir.

PISA 2009 Matematik okuryazarlığı alt testinde bulunan maddelerde cinsiyet açısından SIBTEST yöntemine göre DMF gösteren madde var mıdır?

PISA 2009 matematik okuryazarlığı alt testinde bulunan maddelerin lojistik regresyon ile analizi SPSS paket programı ile yapılmıştır. Lojistik Regresyon ile DMF belirlenirken Zumbo (1999) tarafından yazılmış olan kod sistemi (syntax) kullanılmıştır. Analiz sonuçlarının değerlendirilmesi yapılırken, analize Model 1'de ilave edilen toplam puan, Model 2'de ilave edilen grup ve model 3'de ilave edilen grup ile toplam puanın etkileşimi değişkenlerinden elde edilen R^2 (Nagelkerke R Square)

değerleri arasındaki farkı ($R_2^2 - R_1^2$ ile $R_3^2 - R_2^2$) temsil eden ΔR^2 'nin almış olduğu değere göre yapılmaktadır. Ancak yapılan çalışmada R_3^2 ile R_2^2 değerlerinin aynı olmasından dolayı değerlendirmeler sadece $R_2^2 - R_1^2$ farkından elde edilen ΔR^2 değeri ile yapılmıştır.

Tablo 5. DMF içeren maddeler, düzeyleri ve avantaj sağladığı grup ile ilgili bilgiler

Madde	β	p	R_1^2	R_2^2	R_3^2	ΔR^2	DMF Düzeyi	Avantaj sağladığı grup
1	0,244	0,030	0,256	0,267	0,267	0,011	B	Odak (Erkek)
11	-0,927	0,000	0,515	0,538	0,538	0,023	C	Referans (Kız)
17	3,016	0,000	0,218	0,250	0,250	0,032	C	Odak (Erkek)

Tablo 5 incelendiğinde Lojistik Regresyon yöntemine göre kız öğrencilere farklı işleyen 1 madde ve erkek öğrencilere farklı işleyen 2 tane madde bulunmuştur. Kızlara avantajlı olacak şekilde işleyen madde C düzeyinde, erkeklere avantajlı olacak şekilde işleyen maddeler ise sırasıyla B ve C düzeyindedir.

PISA 2009 Matematik okuryazarlığı alt testinde bulunan maddelerde cinsiyet açısından Lojistik Regresyon yöntemine göre DMF gösteren madde var mıdır?

SIBTEST yöntemi ile verilerin değerlendirilmesi yapılırken öncelikli olarak %95 anlamlılık düzeyine göre p değerlerine bakılmıştır. Daha sonrasında ise β değerine bakılır. β değerinin pozitif olduğu maddeler referans grubu için, negatif olduğu maddeler ise odak grubu için avantaj sağladığını göstermektedir. DMF içeren maddelerin DMF düzeylerinin belirlenmesinde Stout ve Roussos (1995) tarafından oluşturulan sınıflama sistemi kullanılmıştır. Cinsiyet değişkenine göre DMF'li olduğu belirlenen maddeler ile maddelere ait DMF düzeyleri ve avantajlı oldukları gruplar Tablo 6'da verilmiştir.

Tablo 6. Matematik testine ait SIBTEST yöntemine göre yapılan analiz sonuçları

Madde	β	Standart Hata	p	DMF düzeyi	Avantaj sağladığı grup
1	-0.117	0.055	0.032	C	Odak (Erkek)
5	-0.100	0.055	0.066	C	Odak (Erkek)
6	-0.066	0.049	0.183	B	Odak (Erkek)
11	0.130	0.044	0.003	C	Referans (Kız)

Tablo 6 incelendiğinde SIBTEST yöntemine göre kız öğrencilere farklı işleyen 1 madde ve erkek öğrencilere farklı işleyen 3 tane madde bulunmuştur. Kızlara avantajlı olacak şekilde işleyen madde C düzeyinde, erkeklere avantajlı olacak şekilde işleyen maddeler ise sırasıyla C, C ve B düzeyindedir.

Cinsiyet değişkenine ilişkin DMF analizlerinden elde edilen sonuçlara göre DMF gösteren madde sayısı bakımından yöntemler arasında fark vardır ancak maddenin avantajlı olduğu grup her bir yöntem için aynıdır (1., 5., 6. ve 11. maddeler erkekler lehine; 11. madde kızlar lehine). Bu açıdan araştırmadan elde edilen sonuç Yurdugül (2003), Bekci (2007) ve Gök ve diğerlerinin (2010) yapmış oldukları çalışmalarında elde ettiği sonuç ile tutarlılık göstermektedir.

Ayan (2011)'in PISA 2009 Fen Bilgisi maddeleri ile yapmış olduğu çalışmada hem MH yöntemi hem de LR yönteminde 4 adet DMF gösteren madde tespit edilmiştir. Ancak bu 4'er maddenin tamamı birbirinden farklıdır. Yapılan çalışmada ise DMF'li olduğu tespit edilen maddelerin yöntemlere göre bu kadar farklılık göstermediği görülmüştür. Bu açıdan elde edilen sonuçlar, Ayan (2011)'in yapmış olduğu çalışma ile uyumsuzdur.

Yıldırım (2008)'in PISA 2003 Matematik testindeki maddeler ile yapmış olduğu çalışma sonucunda MH yöntemi 13, Olabilirlik Oranı Analizi yöntemi 15 maddeyi farklı işleyen madde olarak tespit ederken Sınırlandırılmış Faktör Çözümlemesi yöntemi 7 maddeyi farklı işleyen madde olarak tespit etmiştir. Bu 7 madde her üç yöntem tarafından da farklı işleyen madde olarak tespit edilmiştir. Yine PISA 2009 Fen Bilgisi maddeleri ile yapılan çalışmada toplam da 8 maddede DMF tespit edilmiştir. Uluslararası düzeyde yapılmakta olan PISA çalışmalarında DMF gösteren maddelerin yıllara göre azalma gösterdiği söylenebilir.

PISA 2009 Matematik okuryazarlığı alt testinde bulunan maddelerde uygulandığı ülke açısından Mantel-Haenszel yöntemine göre DMF gösteren madde var mıdır?

PISA 2009 Matematik alttestinde Türkiye-Almanya, Türkiye-Finlandiya ve Türkiye-ABD ülkeleri arasında kültüre göre yapılan DMF analizine ilişkin MH yöntemine ait sonuçlar Tablo 7'de verilmiştir.

Tablo 7. Matematik testine ait Mantel-Haenszel yöntemine göre yapılan analiz sonuçları

	Almanya-Türkiye			Finlandiya-Türkiye			ABD-Türkiye		
	p	Δ -MH	Düzye	p	Δ -MH	Düzye	p	Δ -MH	Düzye
1	0,000	2,272	C	0,000	2,122	C	0,002	1,352	B
2	0,003	1,443	B	0,001	1,574	C	0,859	0,137	
4	0,613	0,257		0,015	1,061	B	0,127	0,646	
6	0,195	-0,631		0,743	-0,177		0,000	-1,882	
7	0,000	3,127	C	0,000	2,425	C	0,000	2,928	C
9	0,062	0,828		0,637	-0,237		0,000	2,928	C
10	0,128	0,891		0,029	1,150	B	0,040	1,111	B
11	0,000	-3,730	C	0,000	-2,552	C	0,547	-0,355	
12	0,003	-2,804	C	0,002	-2,615	C	0,010	-2,299	C
13	0,196	-0,687		0,015	-1,284	B	0,001	-1,424	B
14	0,529	-0,381		0,005	-1,495	B	0,000	-2,820	C
15	0,000	1,619	C	0,000	2,155	C	0,008	1,114	B

17	0,000	-3,809	C	0,082	-2,031		0,000	-4,853	C
19	0,000	-3,407	C	0,000	-1,831	C	0,000	-2,054	C
20	0,000	-1,741	C	0,001	-1,462	B	0,357	-0,411	
21	0,001	1,586	C	0,059	0,941		0,434	0,438	

Maddelerin incelemesi yapılırken öncelikli olarak %95 anlamlılık düzeyine göre p değerlerine bakılmıştır. Daha sonrasında ise Zikey (1993) tarafından düzenlenen sınıflama sistemine göre DMF düzeyleri belirlenmiştir. Maddelerin Δ -MH katsayısının almış olduğu değer negatif ise referans grubu için pozitif ise odak grubu için bir avantaj sağlamaktadır. Bu durumda tablo 7'ye göre Türkiye ile diğer üç ülke arasındaki DMF analizi sonuçlarında ülkeler arasındaki DMF'li madde sayıları Türkiye-Finlandiya arasında 12 madde, Türkiye-ABD 11 madde ve Türkiye-Almanya arasında ise 10 maddedir. Testteki 1., 7. ve 15. maddeler sürekli olarak Türkiye ile karşılaştırılan ülkelerdeki öğrencilere avantaj sağlamış; 12. ve 19. maddelerin ise sürekli olarak Türkiye'deki öğrencilere avantaj sağladıkları tespit edilmiştir. Diğer maddeler bazı yöntemler için Türkiye lehine, bazı yöntemler için ise Türkiye dışındaki ülkelerin lehine işlediği görülmüştür.

PISA 2009 Matematik okuryazarlığı alt testinde bulunan maddelerde uygulandığı ülke açısından lojistik regresyon yöntemine göre DMF gösteren madde var mıdır?

PISA 2009 Matematik alttestinde Türkiye-Almanya, Türkiye-Finlandiya ve Türkiye-ABD ülkeleri arasında kültüre göre yapılan DMF analizine ilişkin LR yöntemine ait sonuçlar Tablo 8'de verilmiştir.

Tablo 8. Matematik testine ait Lojistik Regresyon yöntemine göre yapılan analiz sonuçları

Madde	Almanya Türkiye		Finlandiya Türkiye		ABD Türkiye	
	ΔR^2	Düzeyi	ΔR^2	Düzeyi	ΔR^2	Düzeyi
1	0,037	C	0,026	C	0,017	B
2	0,011	B	0,014	B	0	
6	0,004		0		0,026	C
7	0,051	C	0,032	C	0,056	C
9	0,004		0,002		0,029	C
11	0,088	C	0,046	C	0,001	
12	0,041	C	0,050	C	0,030	C
13	0,005		0,014	B	0,018	B
14	0,002		0,014	B	0,049	C
15	0,017	B	0,028	C	0,011	B
17	0,078	C	0,020	C	0,135	C
19	0,077	C	0,025	C	0,028	C
20	0,028	C	0,022	C	0,001	
21	0,014	B	0,004		0,001	

Analiz sonuçlarının değerlendirilmesi yapılırken, analize Model 1’de ilave edilen toplam puan, Model 2’de ilave edilen grup ve model 3’de ilave edilen grup ile toplam puanın etkileşimi değişkenlerinden elde edilen R^2 (Nagelkerke R Square) değerleri arasındaki farkı ($R_2^2 - R_1^2$ ile $R_3^2 - R_2^2$) temsil eden ΔR^2 ’nin almış olduğu değere göre yapılmaktadır. Ancak yapılan çalışmada R_3^2 ile R_2^2 değerlerinin aynı olmasından dolayı değerlendirmeler sadece $R_2^2 - R_1^2$ farkından elde edilen ΔR^2 değeri ile yapılmıştır.

Tablo 8 incelendiğinde bazı maddelerin sürekli olarak Türkiye dışındaki ülkelere avantajlı olacak şekilde işleyen madde olduğu tespit edilmiştir. Sürekli olarak Türkiye dışındaki ülkelere avantaj sağlayacak şekilde işleyen maddeler 1., 7. , 12. ve 15. maddelerdir. Diğer maddeler bazı yöntemler için Türkiye lehine, bazı yöntemler için ise Türkiye dışındaki ülkelerin lehine işlediği görülmüştür.

PISA 2009 Matematik okuryazarlığı alt testinde bulunan maddelerde uygulandığı ülke açısından SIBTEST yöntemine göre DMF gösteren madde var mıdır?

PISA 2009 Matematik alttestinde Türkiye-Almanya, Türkiye-Finlandiya ve Türkiye-ABD ülkeleri arasında kültüre göre yapılan DMF analizine ilişkin SIBTEST yöntemine ait sonuçlar Tablo 9’de verilmiştir.

Tablo 9. Matematik testine ait SIBTEST yöntemine göre yapılan analiz sonuçları

	Almanya Türkiye		Finlandiya Türkiye		ABD Türkiye	
	β	Düzeyi	β	Düzeyi	β	Düzeyi
1	-0,17	C	-0,14	C	-0,129	C
2	-0,097	C	-0,1	C	-0,015	
3	-0,007		0,007		-0,084	B
4	-0,037		-0,09	C	-0,071	B
6	0,044		0,003		0,134	C
7	-0,216	C	-0,176	C	-0,202	C
8	-0,073	B	-0,015		-0,076	B
9	-0,085	B	0,018		-0,174	C
10	-0,039		-0,041		-0,063	B
11	0,234	C	0,159	C	0,016	
12	0,074	B	0,095	C	0,054	
13	0,041		0,071	B	0,11	C
14	0,021		0,096	C	0,144	C
15	-0,151	C	-0,185	C	-0,109	C
16	-0,006		-0,065	B	0,006	
17	0,104	C	0,025		0,184	C
19	0,238	C	0,167	C	0,154	C
20	0,146	C	0,103	C	0,031	
21	-0,117	C	-0,08	B	-0,029	

Verilerin değerlendirilmesi yapılırken β değerine bakılır. β değerinin pozitif olduğu maddeler referans grubu için, negatif olduğu maddeler ise odak grubu için avantaj sağladığını göstermektedir. DMF içeren maddelerin DMF düzeylerinin belirlenmesinde Roussos ve Stout (1995) tarafından oluşturulan sınıflama sistemi kullanılmıştır.

Tablo 9 incelendiğinde bazı maddelerin sürekli olarak Türkiye dışındaki ülkelere bazı maddelerin ise sürekli olarak Türkiye için avantajlı olacak şekilde işleyen madde olduğu tespit edilmiştir. Sürekli olarak Türkiye dışındaki ülkelere avantaj sağlayacak şekilde işleyen maddeler 1., 7. ve 15. maddeler; sürekli olarak Türkiye için avantajlı olacak şekilde işleyen madde ise 19. maddedir. Diğer maddeler bazı yöntemler için Türkiye lehine, bazı yöntemler için ise Türkiye dışındaki ülkelerin lehine işlediği görülmüştür.

Kültür değişkenine ilişkin DMF analizlerinden elde edilen verilerin tamamı düşünülecek olursa 1., 7. ve 15. maddelerin Türkiye dışındaki ülkeler lehine avantajlı, 19. maddenin Türkiye lehine avantajlı olduğu ve 12. maddenin MH ile SIBTEST yöntemlerine göre Türkiye lehine avantaj sağlarken, LR yöntemine göre ise Türkiye dışındaki ülkeler lehine avantaj sağladığı söylenebilir.

Farklı kültürleri ve farklı dilleri olan ABD, Almanya, Finlandiya ve Türkiye arasında yapılan analizler sonucunda, Matematik testinde bulunan maddelerin yaklaşık olarak yarısında DMF içeren madde bulunduğu görülmüştür. Bu sonuç, dilsel ve kültürel farklılıkların artıkça DMF gösteren maddelerin arttığını göstermektedir. Elde edilen bu sonuç, Asil (2010) ve Atalay (2010)'un yaptığı çalışma sonuçları ile tutarlılık göstermektedir.

Elde edilen analiz sonuçları incelendiğinde üç farklı DMF belirleme yönteminin madde DMF'li olan maddeleri belirleme konusunda yaklaşık olarak aynı sonuçları verdiği görülmektedir. Üç farklı yöntemden elde edilen DMF'li maddelerin DMF düzeyleri ise birbirlerine göre farklılık göstermektedir. Bu farklılığın ise DMF düzeylerini belirlemede kullanılan sınıflama sistemlerinden kaynaklandığı düşünülmektedir.

DMF analizleri yanlılık araştırmalarının ilk aşaması olarak değerlendirilir. DMF'li olduğu tespit edilen maddelerin yanlı olup olmadığına alt gruplardaki testin yapısı ve kapsamı dikkate alınarak yapılan incelemeler ve uzman görüşleri sonucunda karar verilir. Farklı alt gruplardaki bireylerin test puanlarının buldukları gruba bağlı olarak değişmesi diğer bir anlamda yanlı olması sistematik hatadır. Bu nedenle yanlılık bir geçerlik sorunudur.

Farklı ülkelerde uygulanan PISA verileri üzerinde elde edilen veriler üzerinden yapılan analizler sonucunda yaklaşık olarak testteki maddelerin %50'sinin DMF'li maddelerden oluştuğu görülmektedir. Bu amaçla MEB Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı'nın uygulamalarda görev alan çevirmenlerin seçimi, eğitimi ve PISA çalışmalarında çevirmenler, alan

uzmanları ve ölçme değerlendirme uzmanlarının ortak çalışmasının sağlanması konusuna daha fazla önem verilmesi önerilebilir.

Kaynakça

- Asil, M. (2010). Uluslararası Öğrenci Değerlendirme Programı (PISA) 2006 Öğrenci Anketinin Kültürler Arası Eşdeğerliğinin İncelenmesi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, yayınlanmamış doktora tezi.
- 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*, Yayınlanmamış Yüksek Lisans Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Ayan, C. (2011). *PISA 2009 Fen Okuryazarlığı Alt Testinin Değişen Madde Fonksiyonu Açısından İncelenmesi*, Yayınlanmamış Yüksek Lisans Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Bakan Kalaycıoğlu, D. ve Kelecioğlu, H. (2011). Öğrenci Seçme Sınavı'nın madde yanlılığı açısından incelenmesi. *Eğitim ve Bilim*, 36 (161), 3-12.
- Bekçi, B. (2007). Ortaöğretim Kurumları Öğrenci Seçme Sınavının Değişen Madde Fonksiyonlarının Cinsiyete ve Okul Türüne Göre İncelenmesi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Camilli, G. ve Shepard, L. A. (1994). *Methods for Identifying Biased Test Items*. Hollywood: Sage Publications.
- Clauser, B. ve Mazor, K. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issue and Practice*, 17, 31-44.
- Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and White students in integrated colleges. *Journal of Educational Measurement*, 5, 115-124.
- Dorans, N. J., ve Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. P. W. Holland ve H. Wainer (ed.), *Differential Item Functioning* (ss. 35-66). Lawrence Erlbaum: Hillsdale, NJ.
- EARGED, (2010) PISA 2009 Uluslararası Öğrenci Değerlendirme Programı Ulusal Ön Rapor, Ankara, MEB.
- EARGED, (2011) TIMSS 2007 Ulusal Matematik ve Fen Raporu 8. Sınıflar Ankara, MEB.
- Güler, N. (2011). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Pegem Akademi Yayınevi.
- Holland, P.W., ve Thayer, D.T. (1986). Differential item performance and the Mantel-Haenszel procedure (Technical Report No. 86-69). Princeton, NJ: Educational Testing Service.
- Holland, P. W. (1985). On the study of differential item performance without IRT. In *Proceedings of the 27th Annual Conference of the Military Testing Association* (Vol. 1, pp. 282-287). San Diego, CA: Navy Personnel Research and Development Center.
- Mantel, N. ve Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *Journal Natural Cancer Institute*, 22 (4), 719-748.
- OECD, (2009) Organization for Economic Cooperation and Development 2009. *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*. Paris: OECD. H. Wainer (ed.) *Differential item functioning* (ss. 337-347). Erlbaum: Hillsdale, NJ.
- Osterlind, S. ve H. Everson (2009). *Differential item functioning*. Londra: Sage Publications.
- Penfield, R. ve Camilli, G. (2007). Test fairness and differential item functioning. C.R. Rao (ed.) *Handbook of statistics: Psychometrics, volume 26* (ss. 125-167). Amsterdam: Elsevier.
- Roussos, L. ve Stout, W. (1996) A multidimensionality-based DIF analysis paradigm. *Applied Psychological Measurement* 20, 355-371.

- Shealy, R., ve Stout, W. F. (1993b). A model-based standardization approach that separates true bias/DIF from group differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika*, 58, 159-194.
- Swaminathan, H. ve Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361-370.
- Turgut, F. (1997). *Eğitimde Ölçme ve Değerlendirme Metotları*. Ankara: Yargıcı Matbaası.
- Yıldırım, D. (2008). Farklı İşleyen Maddelerin Belirlenmesinde Sınırlandırılmış Faktör Çözümlemesinin Olabilirlik-Oranı ve Mantel-Haenszel Yöntemleriyle Karşılaştırılması. Hacettepe Üniversitesi Eğitim Fakültesi, 34, 297-307.
- Yurdugül, H. (2003). *Ortaöğretim kurumları seçme ve yerleştirme sınavının madde yanlılığı açısından incelenmesi*. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi.
- Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. P. W. Holland ve H. Wainer (ed.) *Differential item functioning* (ss. 337-347). Erlbaum: Hillsdale, NJ.
- 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*. Directorate of Human Resources Research and Evaluation, Department of National Defense. Ottawa: Canada.

Extended English Abstract

Purpose

This study performs a Differential Item Function (DIF) analysis in terms of gender and culture on the items available in the PISA 2009 mathematics literacy sub-test. The DIF analyses were done through the Mantel Haenszel, Logistic Regression and the SIBTEST methods.

Method

The data for the gender variable were collected from the responses given by 332 students to the items in the mathematics literacy sub-test during the administration of the 5th booklet in the PISA 2009 application whereas the data for the culture variable were collected through the application of the 5th booklet in Turkey, Germany, Finland and the United States in the PISA 2009 application. Besides, the descriptive statistics and the reliability analyses were conducted for the mathematics literacy sub-test, and general information concerning the group distribution prior to the DIF analysis was obtained.

Results

Following the DIF analysis, which had been performed for the gender variable, 2 DIFs were specified in the MH technique while 3 were found in the LR technique and 4 were found in the SIBTEST technique. It may be said that 4 of the 5 items with DIF were in favour of male students, and only one was in favour of female students.

Based on the MH analysis results which had been performed in relation to the data obtained from the students in Turkey and Germany, the DIF was found in 10 items; and 5 of them were to the advantage of students in Germany and the other items were to the advantage of those in Turkey. With the LR analysis the DIF was found in 10 items, and 7 of them were to the advantage of German students and 3 were to the advantage of Turkish students. In consequence of the SIBTEST, the DIF was found in 12 items; 7 of them were found to be to the advantage of German students and 5 were found to be to the advantage of Turkish students.

According to the results of the MH analyses conducted for the data obtained from students in Turkey and Finland, the DIF was found in 12 items; and 6 of the items were to the advantage of Finnish students whereas the other 6 were to the advantage of Turkish students. In consequence of the LR analysis, the DIF was found in 11 items; and all of the items were to the advantage of Finnish students. The SIBTEST results, on the other hand, demonstrated that the DIF was available in 13 items; and that 7 of them were to the advantage of Finnish students while 6 were to the advantage of Turkish students.

Based on the MH analysis results for the data obtained from the students in Turkey and in the United States, the DIF was found for 10 items; and 5 of them were to the advantage of American Students whereas 5 were to the advantage of Turkish students. According to the LR analysis results, the DIF was found in 10 items; and 5 of them were to the advantage of American students while 5 were to the advantage of Turkish students. According to the results of the SIBTEST, however, the DIF was found in 13 items; and 8 of them were to the advantage of American students while 5 were to the advantage of Turkish students. This result shows that, when the language-related and cultural differences increase between countries, the number of items which have DIF increase.

Conclusion and Recommendation

When the obtained results of analysis are examined, observed that three kinds of DIF methods give approximately same results on the determination of substances with DIF. DIF levels of substance with DIF show differences for these three methods. It is thought that this difference is derived from classification methods which are used for detection of DIF levels.

DMF analysis is considered the first phase of studies on item bias. Whether the items detected to have DMF's are biased, is determined as conclusion of the investigations taking into account the structure and scope of the test with regard to sub-groups. The changes in the test scores of the individuals in different sub-groups depending on their groups, in other words, the test being biased, is a systematic error. Therefore, the test bias problem is also a problem of validity.

The analyses of PISA data when the culture is used as a variable, approximately half of test items have DIF. Consequently Ministry of Education researchers have to be more careful when they are translating and composing the Turkish version of PISA applications. They have to collaborate with experts of subject areas and experts on educational measurement and evaluation.