

PISA 2012 BİLİŞSEL MADDELERİNİN KÜLTÜRE GÖRE DEĞİŞEN MADDE FONKSİYONU BAKIMINDAN İNCELENMESİ

Yrd. Doç. Dr. Şeyma Uyar
Mehmet Akif Ersoy Üniversitesi
syuksel@mehmetakif.edu.tr

Yrd. Doç. Dr. Gülden Kaya Uyanık
Sakarya Üniversitesi
guldenk@sakarya.edu.tr

Özet

Bu araştırmada PISA 2012 verilerinde yer alan matematik ve fen okuryazarlığı alt testindeki bilişsel maddeler ülkeler arasında değişen madde fonksiyonu (DMF) bakımından incelenmiştir. Araştırma, OECD ortalamasının üstünde puan alan iki ülke (Hong-Kong ve Finlandiya) ve altında puan alan iki ülke (Türkiye-Rusya) örnekleriyle gruplardaki öğrenci sayıları eşit olacak şekilde yürütülmüştür. Araştırmada yalnızca 1 ve 0 biçiminde puanlanan bilişsel maddeler dikkate alınmıştır. Değişen madde fonksiyonunu belirlemek amacıyla geliştirilmiş Mantel-Haenszel ve geliştirilmiş lojistik regresyon yöntemleri kullanılarak dört ülke arasında tek biçimli DMF içeren maddeler aynı anda belirlenmiştir. Sonuçlar orta ve yüksek düzeyde (B ve C) DMF içeren madde sayısının oldukça fazla olduğunu göstermiştir.

Anahtar Sözcükler: Değişen madde fonksiyonu, geliştirilmiş lojistik regresyon, geliştirilmiş Mantel-Haenszel.

INVESTIGATION OF PISA 2012 COGNITIVE ITEMS IN TERMS OF DIFFERENTIAL ITEM FUNCTIONING BASED ON CULTURE

Abstract

In this study, items in mathematics and science literacy sub-tests collected from PISA 2012 dataset were investigated in terms of differential item functioning (DIF). This study was carried out with equal sample sizes which means are high then OECD average scores (Hong-Kong and Finland) and which means are below then OECD average scores (Turkey and Russia). Only dichotomous variables were considered in this study. To determine differential item functioning across four countries generalized Mantel Haenszel and generalized logistic regression methods were used. Results showed that there were too many items including high level DIF (B and C).

Keywords: Differential item functioning, generalized logistic regression, generalized Mantel-Haenszel.

GİRİŞ

Uluslararası eğitim uygulamalarının en geniş kapsamlı olanlarından biri Ekonomik İşbirliği ve Kalkınma Teşkilatı-OECD tarafından düzenlenen Uluslararası Öğrenci Başarılarını Değerlendirme Programı olarak isimlendirilen PISA sınavıdır. İlk olarak 2000 yılında başlayan uygulama üç yılda bir yapılmaktadır ve Türkiye ilk olarak 2003 yılında katılmıştır (MEB,2013).

PISA sınavında öğrencilerin, matematik, fen ve okuma gibi farklı alanlarındaki bilgi ve becerileri değerlendirilir. Sınava OECD üyesi ve diğer katılımcı ülkelerdeki 14-15 yaş aralığında olan öğrenciler katılır. PISA' da öğrencilerin matematik, fen ve okuma temel konu alanı bilgi ve becerilerini değerlendirmenin yanında öğrenme stratejileri, problem çözme becerileri, ilgi ve tutumları da araştırılmaktadır. Yapılan her uygulamada farklı bir alana ağırlık

verilmektedir. 2012 yılında yapılan PISA sınavına 65 ülkeden yaklaşık olarak 510 bin öğrenci katılmıştır ve ağırlıklı alanı matematik olmuştur(MEB, 2013).

Uluslararası düzeyde yapılan PISA sınavı uygulandığı tarihten itibaren ülkelerin genel durumları ve sıralamaları hakkında bilgi verir niteliktedir. Bu bilgiler ülkelerin eğitim sisteminde var olan eksikliklerin giderilmesi ve alınması gereken önlemlerin belirlenmesi açısından önemlidir. Bu nedenle uygulamada kullanılan ölçme araçlarının psikometrik özelliklerinin yeterli düzeyde olması gerekmektedir. Aksi halde alınan sonuçların doğruluğu tartışılır hale gelir. Yeterli geçerlik ve güvenilirlikte olmayan ölçme araçları ile yapılan ölçmelerden elde edilen sonuçlara dayanarak verilecek kararlar hatalara sebep olabilir (Clauser ve Mazor, 1988; Horst, 1966)

Ölçme araçlarında bulunması gereken en önemli özelliklerden biri geçerliktir. Geçerlik en kısa tanımı ile ölçme aracının ölçmesi istenilen özelliği ne derece ölçebildiğinin derecesidir. Elde edilen ölçme sonuçları değerlendirilirken geçerlik sağlanmazsa, bu sonuçlara dayalı alınan kararlar da hatalı olacaktır. Madde ve test yanlılığı geçerliliği etkileyen en önemli tehditlerin arasında sayılabilir (Clauser ve Mazor, 1998; Zumbo, 1999).

İlk olarak Cleary (1968) tarafından ortaya atılan test yanlılığı kavramı testle ölçülen özelliklerden farklı olarak belli bir gruba ait ölçme sonuçlarının farklılaşması olarak tanımlanabilir. Bir diğer deyişle testin kapsamı alt gruplarda farklılık gösterip, bu farklılıktan kaynaklı avantajlı ya da dezavantajlı bir durum oluşuyorsa, oluşan bu durum nedeniyle testin yanlı olduğu söylenir (Angoff, 1982). Test yanlılığı genel olarak farklı alt gruplar veya farklı cinsiyet grupları için söz konusudur. Test yanlılığı test dışı faktörlerden kaynaklanabileceği gibi testte yer alan maddelerden de kaynaklanabilir (Camilli ve Shepard, 1994). Maddelerden kaynaklanan yanlılık madde yanlılığı olarak adlandırılır.

Madde yanlılığı, testin özelliği, uygulandığı koşullar veya maddenin yapısından kaynaklı olarak bir maddeyi doğru yanıtlama olasılığının farklı gruplar arasında farklılık göstermesi durumu olarak tanımlanır (Tittle,1988; Zumbo, 1999). Maddenin doğru yanıtlama durumu ölçülen yetenek düzeyine bağlı olması gerekirken madde yanlılığı söz konusu olduğunda bireyin herhangi bir gruba ait olup olmadığına bağlıdır (Osterlind, 1983). Madde yanlılığı belirleme çalışmaları iki aşamalı olarak yürütülür. İlk aşama istatistiksel bir süreçtir ve bu aşamada yetenek düzeyi aynı olan ancak farklı gruplarda yer alan bireyler için madde fonksiyonunun farklılaşması belirlenir. Farklılaşmanın olduğu maddeler için değişen madde fonksiyonu (DMF) gösteren maddeler denir (Camilli ve Shepard, 1994). DMF birçok kaynakta “yetenek düzeyleri aynı olmasına rağmen cinsiyet, ırk, sosyoekonomik düzey, inanç vb. gibi farklı gruplarda yer alan kişilerin test maddelerine doğru yanıt verme olasılıklarının farklılaşması” biçiminde tanımlanmıştır (Devine ve Raju, 1982; Mellenberg, 1989; Osterlind, 1983; Raju, 1990; Rodney ve Drasgow, 1990; Shepard, Camilli, ve Williams, 1985). Bir maddede DMF bulunmasının iki temel sebebi olabilir. Bunlardan biri ölçülen özelliğin gruplarda gerçekten farklı olması diğeri ise maddenin yanlı olmasıdır. Yanlı olan maddede DMF gözlenir ancak aksi her zaman doğru değildir. DMF gösteren maddelerin tümü yanlı olmayabilir (Kamata ve Vaughn, 2004). Maddenin yanlı olduğuna ikinci aşamada karar verilir. Bu aşamada maddenin yanlı olduğuna eğitim ve psikolojiye dayanan bulgularla, uzman kanısı, içerik analizi, olgusal değerlendirme vb. yöntemlerle karar verilir.

Değişen madde fonksiyonunun, tek biçimli (uniform) ve tek biçimli olmayan (non-uniform) fonksiyonlar olmak üzere iki farklı türü vardır (Mellenbergh, 1982). Tek biçimli DMF’de maddenin doğru yanıtlanma olasılığı tüm yetenek düzeyleri için bir grubun lehine işlerken tek biçimli olmayan DMF’de madde farklı yetenek düzeyleri için farklı gruplara avantaj sağlamaktadır (Camilli ve Shepard, 1994; Zumbo, 1999).

Madde yanlılığının sinyalini veren DMF belirleme için birçok yöntem vardır. Bu yöntemler genel olarak klasik test kuramına (KTK) ve madde tepki kuramına (MTK) dayalı teknikler olarak ayrılır. Lojistik regresyon (LR) (Swaminathan ve Rogers, 1990), Mantel-Haenszel (MH) (Holland ve Thayer, 1988) ve Sibtest yöntemi (Shealy ve Stout, 1993) sıklıkla kullanılan KTK’ya dayalı tekniklerdir. MTK’ya dayalı teknikler arasında ise işaretli ve işaretli alan indeksleri(Raju, 1990), Lord’un Kikare’si(Lord, 1980) ve olabilirlik oran teknikleri sayılabilir (Thissen, Steinberg, ve Wainer, 1988). Tüm bu yöntemler referans ve odak grup olarak adlandırılan iki grubu karşılaştırmak için kullanılır. Ancak birçok durumda madde fonksiyonlarının ikiden fazla grup arasındaki durumunun incelenmesi gerekebilir(Kim, Cohen ve Park, 1995).

PISA gibi uluslararası ölçekli sınavlarda yer alan maddeler DMF açısından incelendiğinde genel olarak ikiden fazla ülkenin sonuçlarının karşılaştırılması daha uygun sonuçlar verebilir. İki den fazla grubun karşılaştırılmasının yapıldığı birçok çalışmada gruplar ikiye bölünmüş şekilde karşılaştırılmışlardır (Ellis ve Kimmel, 1992; Schmitt ve Dorans, 1990; Zwick ve Ercikan, 1989; Bakan Kalaycıoğlu ve Kelecioğlu, 2011; Gök, Atalay Kabasakal ve Kelecioğlu, 2015). Ancak tek bir durumun birden çok hipotez şeklinde test edilmesi I. tür hataya sebep olabilmektedir (Kanje, 2007, Penfield, 2001).

Günümüzde I. tür hataya sebep olmadan birden çok grubun DMF açısından karşılaştırılmasını sağlayan yöntemler de bulunmaktadır. Bu yöntemlerden ikisi genelleştirilmiş Mantel Haenszel (Mantel ve Haenszel, 1959; Somes, 1986; Penfield, 2001; Fidalgo ve Madeira, 2008) ve genelleştirilmiş Lojistik regresyondur (Kanje, 2007; Magis , Raïche, Béland ve Paul, 2011). Genelleştirilmiş Mantel Haenszel (GMH) yöntemi ile çoklu gruplar arası tek biçimli DMF belirlenebilir (Magis , Raïche, Béland ve Paul, 2011). Mantel Haenszel (MH) yönteminde çok kategorili maddelerde kategorileri sıralı ele alırken GMH’de kategoriler sınıflamalı olarak ele alınır (Su ve Wang,2005). Benzer şekilde MH yönteminde referans ve odak grupların madde ortalamaları karşılaştırılırken GMH’de referans ve odak grupların madde dağılımları karşılaştırılır (Su ve Wang,2005). Genelleştirilmiş Lojistik regresyon (GLR) lojistik regresyon yönteminin genişletilmiş halidir ve bu yöntem ile hem tek hem de tek biçimli olmayan DMF kestirilebilir (Magis , Raïche, Béland ve Paul, 2011). Yöntem “birleştirilmiş” odak grubun örneklem büyüklüğünün geniş olduğu durumlarda daha güçlüdür (Magis , Raïche, Béland ve Paul, 2011).

Yapılan bu çalışmada amaç OECD ortalamasının üstünde puan alan iki ülke (Hong-Kong ve Finlandiya) ve altında puan alan iki ülke (Türkiye-Rusya) olmak üzere dört ülkeden elde edilen 2012 PISA matematik ve fen test maddelerinde yer alan değişen madde fonksiyonlarını belirlemektir. Çalışmada genelleştirilmiş Mantel-Haenszel ve genelleştirilmiş lojistik regresyon yöntemleri kullanılarak dört ülke arasında tek biçimli DMF içeren maddeler aynı anda belirlenmiştir. Çalışmanın çoklu grup DMF çalışmalarında kullanılacak yöntemleri açıklaması açısından önemli olduğu düşünülmektedir. Belirtilen amaçlar çerçevesinde çalışmada aşağıdaki sorulara yanıt aranmıştır.

1. Matematik okuryazarlığı testinde genelleştirilmiş Mantel Haneszel (GMH) ve genelleştirilmiş lojistik regresyon (GLR) yöntemlerine göre DMF gösteren madde var mıdır?
2. Fen okuryazarlığı testinde genelleştirilmiş Mantel Haneszel (GMH) ve genelleştirilmiş lojistik regresyon (GLR) yöntemlerine göre DMF gösteren madde var mıdır?

YÖNTEM

Araştırmanın Türü

Bu çalışmada, genelleştirilmiş Mantel-Haenszel ve genelleştirilmiş lojistik regresyon analizleriyle PISA 2012 matematik ve fen testindeki maddeler kültüre göre değişen madde fonksiyonu bakımından incelenmiştir. Araştırma, var olan durumun ortaya koymaya yönelik bir çalışma olduğu için betimsel bir çalışma niteliğindedir.

Evren ve Örneklem

PISA 2012 çalışmasının evrenini 15 yaşındaki öğrenciler oluşturmaktadır. Bu çalışmanın veri setini, PISA 2012 uygulamasında matematik ve fen alt testlerinde OECD ortalamasının üzerinde puan alan Hong-Kong ve Finlandiya ülkeleriyle başarı bakımından OECD ortalamasının altında kalan Türkiye ve Rusya ülkeleri oluşturmaktadır. PISA 2012 uygulamasına Türkiye’den toplam 4848 katılmıştır. Finlandiya örnekleme 8829, Hong-Kong örnekleme 4670 ve Rusya örnekleme 5231 öğrenciden oluşmaktadır. Bu çalışmada matematik ve fen testi için yedinci kitapçığa yanıt veren öğrenci sayıları dikkate alınmıştır. Matematik testini Finlandiya’dan 580, Hong-Kong’ tan 311, Rusya’dan 333 ve Türkiye’den 326 öğrenci cevaplamıştır. Gruplardaki öğrenci sayılarını eşitlemek amacıyla ülkelerden 326 öğrenci rastgele alınarak analize devam edilmiştir. . Fen okuryazarlığı testine yedinci kitapçıkta Türkiye’den 372, Finlandiya’dan 667, Hong-Kong’ tan 372 ve Rusya’dan 449 öğrenci yanıt vermiştir. Referans ve odak grup sayılarını eşitlemek amacıyla bu ülkelerden toplamda 372 öğrenci rastgele olarak seçilmiştir.

Veri Toplama Aracı

PISA uygulamasına katılan öğrenciler, uygulama kapsamında yer alan bütün maddeleri yanıtlamazlar. PISA’ da kullanılan üniteler, 13 madde demeti halinde gruplandırılmıştır. Her bir madde demetinin yanıtlanma süresi 30

dakikadır. 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, öğrenci sayılarının ve madde sayılarının fazla olduğu matematik ve fen testi için yedinci kitapçık seçilmiştir. Yedinci kitapçıkta öğrencilerin fen ve matematik testindeki ikili puanlanan (1 ve 0) bilişsel maddelere verdikleri yanıtlar dikkate alınmıştır.

Verilerin Analizi

Verileri analiz etmek için R 3.1.2 programında "difR" kütüphanesi kullanılmıştır. R, internet aracılığı ile ücretsiz olarak dağıtılan ve hemen hemen bütün işletim sistemlerinde çalışabilen açık kaynak kodlu bir programdır. "difR" kütüphanesi iki kategorili verilerin çeşitli yöntemler ile DMF bakımından incelenmesi amacı ile geliştirilmiştir. Burada bulunan hazır kodlar aracılığıyla genelleştirilmiş Mantel Haenszel ve genelleştirilmiş lojistik regresyon yöntemleriyle tek biçimli DMF belirlenmiştir. Bu yöntemlerle bir grup referans ve diğer üç grup odak grup olarak seçilmiştir. Referans grup odak gruba göre daha avantajlı gruptur. Bu çalışmada Finlandiya referans grup, diğer üç ülke odak grup olacak şekilde aynı anda analize alınmıştır. GMH ve GLR analiz sonucunda eğer referans ve odak gruplar arasında DMF gösteren madde bulunmuşsa, ülkeler MH ve LR yöntemleriyle ayrıca ikili olarak karşılaştırılmıştır. Bu aşamada ülkeler arasında DMF gösteren maddeler ve DMF'nin düzeyi ortaya konmuştur.

Mantel-Haenszel Yöntemi

Mantel-Haenszel (MH), tek biçimli DMF'yi belirlemek için aynı yetenek düzeyinde yer alan iki grubun (odak ve referans grup) ikili puanlanan maddelerden aldıkları puanlardan elde edilen, bir olayın gerçekleşme olasılığının gerçekleşmeme olasılığına bölümü olarak adlandırılan, "od" değerleri arasındaki farka dayanan bir yöntemdir (Mertler ve Vannatta, 2005). Toplam test puanını dikkate alarak aynı puanı alan bireyler için yetenek grupları oluşturulur. Her bir yetenek düzeyi için 2x2 şeklinde çapraz tablo oluşturulur (Holland ve Thayer, 1988). Referans ve odak grupları içerisinde maddeyi doğru ve yanlış yanıtlayanların sayısı ile maddeyi yanıtlayan toplam kişi sayıları dikkate alınarak olabilirlik oranı elde edilir. Elde edilen bu istatistik olasılıksal bir oran olması nedeniyle 0 ile ∞ arasında değerler alabilmektedir. Camilli ve Shepard (1994), yorumlamalara kolaylık getirmesi açısından olabilirlik oranının doğal logaritmasının -2.35 katı alınarak Δ_{MH} (delta) istatistiğinin kullanılmasını önermektedir. Elde edilen Δ_{MH} istatistiği DMF derecesini belirten DMF etki büyüklüğü olarak yorumlanmaktadır. Bu çalışmada MH yöntemi için Dorans ve Holland (1993) tarafından önerilen etki büyüklüğü aralığına göre DMF düzeyine karar verilmiştir. Eğer $|\Delta_{MH}| < 1$ ise maddenin A düzeyinde ihmal edilebilir DMF içerdiğini ya da DMF içermediği, eğer $1 \leq |\Delta_{MH}| < 1.5$ ise maddenin B düzeyinde ya da orta düzey DMF içerdiğini ve $|\Delta_{MH}| \geq 1.5$ ise C düzeyinde yani yüksek DMF içerdiğini belirtilmiştir. Maddelerin Δ_{MH} katsayısının almış olduğu değer negatif ise referans grubu için pozitif ise odak grubu için bir avantaj sağlamaktadır.

Lojistik Regresyon Yöntemi

DMF'yi belirlemek amacıyla kullanılan MH yöntemi pratik olmasına rağmen I. tip hata yapma olasılığının yüksek olması ve tek biçimli olmayan DMF'yi belirleyememesi sebebiyle zayıf kalmaktadır. Lojistik regresyon yöntemi tek biçimli olmayan DMF'yi tanımlaması açısından MH istatistiğine göre daha üstündür. Bu teknik kurulan modeller sayesinde maddeye verilen cevaplar ve toplam puan üzerinden DMF'nin sınıflandırmasını sağlar. Lojistik regresyon analizinde madde bağımlı değişken, gruplar ve toplam puan bağımsız değişkenlerdir. DMF belirleme çalışmalarında değişkenler modele hiyerarşik olarak ilave edilirler. 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. Modellerden elde edilen R^2 değerleri karşılaştırılarak DMF varlığına ve türüne karar verilir. Hangi düzeydeki R^2 değerlerinin önemli sayılacağına ilişkin çeşitli ölçütler geliştirilmiştir. Literatürde örneklem büyüklüklerinden etkileniyor olması nedeniyle farklı sınıflama sistemlerine rastlanmaktadır. Bu çalışmada Gierl, Khaliq ve Boughton (1999) tarafından önerilen R^2 aralıkları dikkate alınmıştır. Eğer $0 < \Delta R^2 < 0.035$ ise maddede DMF'nin olmadığına ya da ihmal edilebilir düzeyde olduğuna, $0.035 \leq \Delta R^2 < 0.07$ ise orta düzey (B) DMF içerdiğine, $\Delta R^2 \geq 0.07$ ise yüksek düzeyde (C) DMF içerdiğine karar verilmiştir.

BULGULAR

Bu bölümde PISA 2012 matematik ve fen okuryazarlığı alt testlerindeki maddelerde kültür bakımından DMF olup olmadığı incelenmiştir. Bu amaçla kullanılan genelleştirilmiş Mantel Haenszel, genelleştirilmiş lojistik regresyon sonuçlarına göre ikili MH ve LR sonuçları arasındaki benzerlik ve farklılıklar karşılaştırılmıştır.

1. Matematik okuryazarlığı testinde genelleştirilmiş Mantel Haneszel (GMH) ve genelleştirilmiş lojistik regresyon (GLR) yöntemlerine göre DMF gösteren madde var mıdır?

Bu alt problemin çözümü için Finlandiya referans grup alınmış, Türkiye, Hong-Kong ve Rusya ülkeleri aynı anda odak grup olarak analize dahil edilmiştir. Öncelikle GMH analizinde %95 anlamlılık düzeyine göre p değerlerine bakılmıştır. Anlamlı bulunan p değerleri için ikili olarak MH değerleri elde edilmiştir. Bu analiz sonucunda da $p < .05$ olarak bulunan maddelerin ΔMH değerlerine bakılmış ve DMF düzeyleri Tablo 1'de verilmiştir. GLR analizinde %95 anlamlılık düzeyine göre p değerlerine bakılmış ve anlamlı bulunan p değerleri için ikili olarak LR değerleri elde edilmiştir. Bu analiz sonucunda da anlamlı ΔR^2 değerlerine göre maddelerin DMF düzeyleri tablo 2'de verilmiştir.

Tablo 1: Matematik Okuryazarlığı Testinde GMH ve MH ile Elde Edilen Sonuçlar

Madde	GMH	1,2		1,3		1,4		2,3		2,4		3,4	
		ΔMH	Düze-yi	ΔMH	Düze-yi	ΔMH	Düze-yi	ΔMH	Düze-yi	ΔMH	Düze-yi	ΔMH	Düze-yi
1	Var	1,73	C	2,25	C	1.02	B	-	-	-	-	-	-
2	Var	-	-	-	-	-1.01	B	-	-	-1.22	B	-1.74	C
3	Var	-1,12	B	-	-	-	-	-	-	1.98	C	1.08	B
4	Yok	-	-	-	-	-	-	-	-	-	-	-	-
5	Var	-	-	-1.75	C	-1.04	B	-1.83	C	-	-	1.11	B
6	Var	3.21	C	-	-	-2.83	C	-3.92	C	-6.14	C	-2.15	C
7	Var	-	-	2.16	C	-	-	1.63	C	-	-	-1.64	C
8	Var	-3.05	C	1.23	B	-1.88	C	5.79	C	1.53	C	-4.17	C
9	Var	7.88	C	2.26	C	-1.37	B	-5.84	C	-10.12	C	-3.67	C
10	Var	-10.48	C	-1.83	C	-2.16	C	9.06	C	10.10	C	-	-
11	Var	5.28	C	-	-	-	-	-5.23	C	-4.69	C	-	-
12	Var	1.88	C	-	-	-	-	-1.48	B	-1.53	C	-	-
13	Var	-	-	1.13	B	-	-	1.35	B	-	-	-2.0	C
14	Yok	-	-	1.70	C	-	-	-	-	-	-	-	-
15	Var	1.61	C	-	-	-	-	-1.40	B	-1.92	C	-	-
16	Var	-1.42	B	-	-	-	-	-	-	-	-	-	-
17	Yok	-	-	-	-	-	-	-	-	-	-	-	-
18	Var	1.61	C	-	-	1.79	C	-	-	-	-	1.16	B
19	Var	1.35	B	-2.13	C	1.31	B	-3.3	C	-	-	3.27	C
20	Var	-1.02	B	-	-	-	-	-	-	1.42	B	-	-
21	Yok	-	-	-	-	-	-	-	-	-	-	-	-
22	Var	-1.63	C	-1.42	B	-	-	-	-	-	-	-	-
23	Var	-3.51	C	-1.27	B	2.01	C	1.37	B	5.39	C	3.14	C
24	Var	-	-	-	-	2.65	C	-	-	2.41	C	3.02	C
25	Var	-1.25	B	-1.46	B	-	-	-	-	2.28	C	2.51	C
26	Var	-4.1	C	-2.75	C	-1.45	B	-	-	3.21	C	2.21	C
27	Var	-1.61	C	1.00	B	-	-	2.54	C	1.92	C	-	-
28	Var	-	-	-	-	2.18	C	-	-	3.47	C	2.09	C

1: Finlandiya, 2:Rusya, 3:Hong-Kong, 4:Türkiye

Tablo 1’de %95 güven düzeyine göre anlamlı bulunan GMH sonuçları incelendiğinde 4, 14, 17 ve 21. maddeler dışında kalan 24 maddenin dört ülkeden en az iki arasında DMF sergilediği görülmektedir. İkili MH analizi sonuçlarına göre Finlandiya ve Rusya arasında 13 madde C düzeyinde DMF göstermektedir. B düzeyinde 4 madde DMF olduğuna işaret etmektedir. Finlandiya ile Hong-Kong arasında ikili MH sonuçları incelendiğinde 9 maddenin C ve 6 maddenin B düzeyinde DMF gösterdiği görülmektedir. GMH ile 14. maddede DMF bulunmazken bu iki ülke arasında MH ile maddenin DMF gösterdiği gözlenmektedir. Finlandiya ile Türkiye arasında DMF’li bulunan madde sayısı 7 tanesi B ve 7 tanesi C düzeyinde olmak üzere 14’tür. Rusya ve Hong-Kong ülkeleri arasında 9 madde B düzeyinde ve 5 madde C düzeyinde DMF sergilemektedir. Türkiye ve Rusya örneklemelerinde toplamda 16 madde DMF’li olarak gözlenmiştir. Bunlardan 14’ü C ve 2’si B düzeyindedir. Son olarak MH yöntemine göre Türkiye ve Hong-Kong örneklemeleri karşılaştırıldığında 12 maddenin C düzeyinde 3 maddenin B düzeyinde DMF’ye işaret ettiği görülmektedir. Yapılan analiz sonucunda sürekli olarak Türkiye ile karşılaştırılan ülkelere avantaj sağlayan maddeler 2, 6 ve 9. maddelerdir. Türkiye lehine sürekli olarak işleyen maddelerin 23, 24 ve 25. maddeler olduğu söylenebilir.

Tablo 2: Matematik Okuryazarlığı Alt Testinde GLR ve LR ile Elde Edilen Sonuçlar

Madde	GLR	1,2		1,3		1,4		2,3		2,4		3,4	
		Δr^2	Düze-yi	Δr^2	Düze-yi	Δr^2	Düze-yi	Δr^2	Düze-yi	Δr^2	Düze-yi	Δr^2	Düze-yi
1	Var	0.05	B	0.11	C	-	-	-	-	-	-	-	-
2	Var	-	-	-	-	0.04	B	-	-	0.07	C	0.06	B
3	Var	0.04	B	-	-	-	-	-	-	0.11	C	-	-
4	Yok	-	-	-	-	-	-	-	-	-	-	-	-
5	Yok	-	-	-	-	-	-	-	-	-	-	-	-
6	Var	0.15	C	-	-	0.07	C	0.14	C	0.35	C	0.03	B
7	Var	-	-	0.12	C	-	-	0.05	B	-	-	0.06	B
8	Var	0.18	C	0.07	C	0.06	B	0.40	C	-	-	0.25	C
9	Var	0.40	C	0.05	B	-	-	0.18	C	0.57	C	0.07	C
10	Var	0.79	C	-	-	-	-	0.73	C	0.53	C	-	-
11	Var	0.31	C	-	-	-	-	0.18	C	0.36	C	-	-
12	Var	0.05	B	-	-	-	-	-	-	-	-	-	-
13	Var	-	-	0.04	B	-	-	-	-	-	-	0.08	C
14	Yok	-	-	-	-	-1.14	B	-	-	-	-	-	-
15	Var	-	-	-	-	-	-	-	-	0.04	B	-	-
16	Var	0.04	B	-	-	-	-	-	-	-	-	-	-
17	Var	-	-	-	-	-	-	-	-	-	-	-	-
18	Var	0.04	B	-	-	0.07	C	-	-	-	-	-	-
19	Var	-	-	-	-	-	-	0.04	B	-	-	-	-
20	Var	-	-	-	-	-	-	-	-	-	-	-	-
21	Var	-	-	-	-	-	-	0.04	B	-	-	-	-
22	Var	0.08	C	0.05	B	-	-	-	-	-	-	-	-
23	Var	0.1	C	-	-	-	-	-	-	0.20	C	0.03	B
24	Var	-	-	-	-	0.06	B	-	-	0.07	C	0.05	B
25	Var	-	-	-	-	-	-	-	-	0.05	B	-	-
26	Var	0.22	C	0.06	B	-	-	-	-	0.14	C	-	-
27	Var	0.04	B	-	-	-	-	0.13	C	0.06	B	-	-
28	Var	-	-	-	-	-	-	-	-	0.07	C	-	-

1: Finlandiya, 2:Rusya, 3:Hong-Kong, 4:Türkiye

Tablo 2’de yer alan GLR sonuçları incelendiğinde 4, 5, ve 14. Madde dışında kalan 25 maddenin en az üç ülke arasında DMF gösterdiği görülmektedir. GLR sonuçlarında anlamlı olarak DMF gösteren maddeler ülkeler bakımından ikili olarak karşılaştırıldığında Finlandiya ve Rusya arasında 8’i C düzeyinde ve 6 tanesi B düzeyinde olmak üzere toplam 14 madde DMF sergilemektedir. Bu yöntemle Finlandiya ve Hong-Kong arasında yalnızca 7

madde DMF'li olarak gözlenmiştir. Türkiye ve Finlandiya arasında DMF gösteren madde sayısı 6'dır ve bunlardan 4'ü B düzeyindedir. Rusya ve Hong-Kong ülkeleri arasında 6 madde C düzeyinde ve 3 madde B düzeyinde DMF sergilemektedir. Türkiye ve Rusya arasında DMF gösteren madde sayısı 13'tür ve 10 tanesi C düzeyinde DMF'ye sahiptir. Türkiye ve Hong-Kong arasında DMF gösteren madde sayısı 8'dir. Bu maddelerden 5 tanesi B düzeyinde DMF sergilemektedir. Ayrıca 14. maddede GLR ile DMF bulunmamasına rağmen Finlandiya ve Türkiye arasında MH ile DMF'li madde bulunmuştur.

2. Fen okuryazarlığı testinde genelleştirilmiş Mantel Haneszel (GMH) ve genelleştirilmiş lojistik regresyon (GLR) yöntemlerine göre DMF gösteren madde var mıdır?

Bu alt problemin çözümü için alt problem 1'de yer alan adımlar incelenmiştir. Fen okuryazarlığı testinde yer alan 34 maddeye ilişkin ülkeler arasındaki DMF düzeyleri GMH ve MH sonuçlarına göre Tablo1'de, GLR ve LR sonuçlarına göre Tablo 4'te verilmiştir.

Tablo 3: Fen Okuryazarlığı Testinde GMH ve MH ile Elde Edilen Sonuçlar

Madde	GMH	1,2		1,3		1,4		2,3		2,4		3,4	
		ΔMH	Düze yi	ΔMH	Düze yi	ΔMH	Düze yi	ΔMH	Düze yi	ΔMH	Düze yi	ΔMH	Düze yi
1	Var	0.53	-	-	-	1.66	C	-1.08	B	1.46	B	2.54	C
2	Var	-2.39	C	-2.7	C	-1.59	C	-	-	1.10	B	-	-
3	Var	-	-	-5.93	C	-1.48	B	-4.72	C	-	-	4.38	C
4	Var	-	-	-2.39	C	-	-	-2.14	C	-	-	1.32	B
5	Var	-	-	-1.10	B	-1.64	C	-1.60	C	-2.14	C	-	-
6	Var	1.35	B	-2.268	C	-2.06	C	-3.43	C	-3.70	C	-	-
7	Var	-	-	1.16	B	-	-	1.65	C	-	-	-1.12	B
8	Var	-	-	-1.10	B	1.15	B	-	-	-	-	-	-
9	Yok	-	-	-	-	-	-	-	-	-	-	-	-
10	Yok	-	-	-	-	-	-	-	-	-	-	-	-
11	Var	-	-	2.45	C	1.58	C	1.51	C	-	-	-	-
12	Var	-	-	-	-	1.70	C	1.79	C	2.52	C	-	-
13	Var	2.15	C	1.98	C	2.19	C	-	-	-	-	-	-
14	Var	-1.08	B	1.01	B	-1.15	B	1.94	C	-	-	-1.69	C
15	Var	-2.29	C	-	-	-2.11	C	1.73	C	-	-	-1.55	C
16	Var	-	-	-2.76	C	-1.73	C	-2.40	C	-1.33	B	-	-
17	Yok	-	-	-	-	-	-	-	-	-	-	-	-
18	Var	-	-	1.86	C	-	-	1.30	B	-	-	-1.56	C
19	Var	-	-	-1.59	C	-1.62	C	-1.95	C	-2.25	C	-	-
20	Var	1.29	B	2.14	C	1.18	B	-	-	-	-	-	-
21	Var	-	-	1.56	C	2.13	C	-	-	-1.09	B	-	-
22	Var	-	-	-2.23	C	-	-	-2.69	C	-1.07	B	1.48	B
23	Var	1.27	B	1.36	B	-	-	-	-	-	-	-	-
24	Var	-2.09	C	-	-	-	-	2.75	C	2.15	C	-	-
25	Var	-	-	-2.69	C	2.00	C	-1.74	C	2.69	C	4.34	C
26	Var	-	-	-	-	-1.50	C	-	-	-1.44	B	-1.71	C
27	Yok	-	-	-	-	-	-	-	-	-	-	-	-
28	Var	-	-	1.16	B	1.44	B	1.00	B	1.53	C	-	-
29	Var	1.78	C	-1.86	C	-	-	-	-	1.49	B	-	-
30	Var	1.29	B	-1.53	C	-	-	-2.45	C	-1.83	C	-	-
31	Var	-2.0	C	-1.7	C	2.3	C	3.67	C	3.16	C	-1.26	B
32	Var	-	-	1.93	C	1.79	C	1.68	C	-1.27	B	-	-
33	Var	1.78	C	1.69	C	-	-	-	-	-1.21	B	-	-
34	Var	-	-	2.04	C	-	-	2.28	C	1.43	B	-1.27	B

1: Finlandiya, 2:Rusya, 3:Hong-Kong, 4:Türkiye

Tablo 3 incelendiğinde, GMH sonuçlarına göre 9, 10, 17 ve 27. Maddeler dışında kaşan 30 maddenin en az iki ülke arasında DMF gösterdiği görülmektedir. Ülkeler arasında ikili MH analiz sonuçları Finlandiya ve Rusya arasında 12 maddede, Finlandiya ile Türkiye arasında 10 maddede, Rusya ile Hong-Kong arasında 21, Rusya ile Türkiye arasında 19 maddede ve Türkiye ile Hong-Kong arasında 12 maddede DMF bulunduğu işaret etmektedir. Türkiye ile karşılaştırılan ülkeler arasında 26. Madde sürekli diğer ülkelere avantaj sağlayan madde olarak gözlenmektedir. Bunun yanında Tablo 3 aracılığıyla tüm ikili karşılaştırmalarda C düzeyinde DMF sayısının daha fazla olduğu söylenebilir.

Tablo 4: Fen Okuryazarlığı Testinde GLR ve LR ile Elde Edilen Sonuçlar

Madde	GLR	1,2		1,3		1,4		2,3		2,4		3,4	
		Δr^2	Düzeği	Δr^2	Düzeği	Δr^2	Düzeği	Δr^2	Düzeği	Δr^2	Düzeği	Δr^2	Düzeği
1	Var	-	-	-	-	0.07	C	0.06	B	0.10	C	0.16	C
2	Var	0.14	C	1.10	C	0.04	B	-	-	-	-	-	-
3	Var	-	-	0.32	C	-	-	0.21	C	-	-	0.11	C
4	Var	-	-	0.10	C	-	-	0.08	C	-	-	-	-
5	Var	-	-	-	-	0.05	C	0.04	B	0.11	C	-	-
6	Var	-	-	0.07	B	0.07	C	0.13	C	0.20	C	-	-
7	Var	-	-	-	-	-	-	0.05	B	-	-	-	-
8	Var	-	-	0.04	B	-	-	-	-	-	-	-	-
9	Yok	-	-	-	-	-	-	-	-	-	-	-	-
10	Yok	-	-	-	-	-	-	-	-	-	-	-	-
11	Var	-	-	0.23	C	0.05	B	0.12	C	-	-	-	-
12	Var	-	-	-	-	0.07	C	0.10	C	0.22	C	-	-
13	Var	0.11	C	0.08	C	0.07	C	-	-	-	-	-	-
14	Var	-	-	0.04	B	0.03	B	0.13	C	-	-	0.13	C
15	Var	0.10	C	-	-	0.07	C	0.05	B	-	-	0.04	B
16	Var	-	-	0.13	C	0.06	B	0.12	C	0.06	B	-	-
17	Yok	-	-	-	-	-	-	-	-	-	-	-	-
18	Var	-	-	0.11	C	-	-	0.04	B	-	-	0.05	B
19	Var	-	-	0.07	C	0.10	C	0.06	B	0.16	C	-	-
20	Var	0.03	B	0.11	C	0.04	B	-	-	-	-	-	-
21	Var	0.04	B	0.10	C	0.14	C	-	-	0.06	-	-	-
22	Var	-	-	0.15	C	-	-	0.22	C	0.09	C	0.04	B
23	Var	0.06	B	0.12	C	-	-	-	-	-	-	-	-
24	Var	0.05	B	-	-	-	-	0.15	C	0.08	C	-	-
25	Var	-	-	0.11	C	0.07	B	0.05	B	0.15	C	0.19	C
26	Var	-	-	-	-	0.05	B	-	-	0.05	B	0.09	C
27	Yok	-	-	-	-	-	-	-	-	-	-	-	-
28	Var	-	-	0.06	B	0.07	B	0.05	B	0.13	C	-	-
29	Var	0.13	C	0.11	C	-	-	-	-	0.06	B	-	-
30	Var	0.05	B	0.10	C	-	-	0.22	C	0.13	C	-	-
31	Var	0.32	C	0.49	C	0.26	C	0.24	C	0.12	C	-	-
32	Var	0.00	-	0.10	C	0.04	B	0.04	B	-	-	-	-
33	Var	0.11	B	0.18	C	-	-	-	-	0.07	B	0.08	C
34	Var	-	-	0.14	C	-	-	0.23	C	0.09	C	0.05	B

1: Finlandiya, 2:Rusya, 3:Hong-Kong, 4:Türkiye

Tablo 4'te fen okuryazarlığı testinde yer alan 34 maddenin GLR ile dört ülke arasındaki DMF durumu incelendiğinde yine dört madde dışında tüm maddelerde DMF olduğu görülmektedir. Finlandiya ile Rusya arasında 5 maddenin C düzeyinde ve 6 maddenin B düzeyinde DMF içermektedir. Finlandiya ile Hong-Kong arasında 23 maddenin, Finlandiya ile Türkiye arasında 18 maddenin, Rusya ile Hong arasında 21 maddenin,

Rusya ile Türkiye arasında 16 maddenin ve Türkiye ile Hong-Kong arasında 10 maddenin DMF sergilediği gözlenmektedir. GLR sonuçları fen okuryazarlığı testinde C düzeyinde DMF'li madde sayısının genel olarak daha fazla olduğuna işaret etmektedir.

TARTIŞMA VE SONUÇ

Farklı kültürleri ve farklı dilleri olan Türkiye, Finlandiya, Hong-Kong ve Rusya arasında yapılan analizler sonucunda, matematik ve fen okuryazarlığı testlerinde maddelerin yaklaşık olarak yarısında DMF içeren madde bulunduğu görülmüştür. Buna göre kültürel farklılıklar arttıkça değişen madde fonksiyonu gösteren maddelerin sayısının arttırdığı görülmektedir. Bu sonuç Asil (2010) , Demir ve Köse (2014)'ün bulguları ile tutarlılık göstermektedir. Bunun yanında özellikle fen okuryazarlığı testinde Finlandiya ve Hong-Kong arasında DMF'li madde sayısının fazla olduğu sonucuna varılmıştır. Ayrıca Türkiye karşılaştırıldığı ülkeler arasında matematik ve fen testi için en az Hong-Kong ile DMF'li madde sergilemiştir. Bu durum, DMF'li bulunan madde sayılarının ülkelerin başarı sırasından etkilenmediğini ortaya koymaktadır (Gök, Atalay Kabasakal ve Kelecioğlu, 2014).

Analiz sonuçları incelendiğinde DMF belirleme yöntemlerinin DMF'li olan maddeleri belirleme konusunda yaklaşık olarak aynı sonuçları verdiği, ancak MH yönteminin daha fazla sayıda DMF'li madde bulunduğu sonucuna varılmıştır. Bu durumun MH yönteminin I. tip hata yapma olasılığının fazla olmasından kaynaklandığı söylenebilir. Yöntemlerden elde edilen DMF'li maddelerin DMF düzeyleri ise birbirlerine göre farklılık göstermektedir. Bu farklılık DMF düzeylerini belirlemede kullanılan sınıflama sisteminden kaynaklanabilmektedir (Demir ve Köse, 2014). GMH ve GLR ile DMF'li bulunmamasına rağmen MH ve LR yöntemleri bu maddelerde DMF bulunduğu işaret etmiştir. Bu sonuç Penfield (2001)'in GMH ve MH yönteminin karşılaştırıldığı ve üç odak grubun bulunduğu durumda I. tip hata oranının MH yönteminde daha fazla olduğu bulgusu ile tutarlılık göstermektedir.

Bu araştırmada ortaya çıkan en önemli sonuç farklı ülkelerde uygulanan PISA verileri üzerinde elde edilen veriler üzerinden yapılan analizler sonucunda yaklaşık olarak testteki maddelerin yarısının DMF'li maddelerden oluşmasıdır. Bu nedenle PISA, PIRLS ve TIMSS gibi uluslararası uygulamalarda yapılacak karşılaştırmalarda maddelerinin DMF gösterip göstermediğinin incelenmesi ve maddelerin DMF göstermesinin nedenleri araştırılmalıdır. Bununla birlikte, çeviri ya da kültürden kaynaklanan sorunları gidermek için uluslararası sınavlarda görev alan kişilerin bu konuda uzman olması ve dikkatli çalışmaları gerekmektedir. Özellikle bu tür sınavlarda görev alan çevirmenlerin seçimine ve çevirmenler ile ölçme değerlendirme uzmanlarının işbirliği içinde çalışmasına büyük önem verilmelidir. Öte yandan araştırmacılara odak grup sayısının ikiden fazla olduğu durumda GMH ve GLR yöntemlerini kullanmaları önerilebilir.

Not: Bu çalışma 13- 15 Mayıs 2016 tarihlerinde Antalya'da 10 Ülkenin katılımıyla düzenlenen 7th International Congress on New Trends in Education – ICONTE'de sözlü bildiri olarak sunulmuştur.

KAYNAKÇA

Adams, R. J., & Rowe, K. J. (1988). Item bias. In J.P.Keeves (ed.) *Educational research, methodology, and measurement: An international handbook*. Oxford: Pergamon Press.

Angoff, W. H. (1982). Use of difficulty and discrimination indices for detecting item bias. *Handbook of methods for detecting test bias*, 96-116.

Asil, M. (2010). Uluslararası Öğrenci Değerlendirme Programı (PISA) 2006 Öğrenci Anketinin Kültürler Arası Eşdeğerliğinin İncelenmesi. *Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü*, yayınlanmamış doktora tezi.

Bakan Kalaycıoğlu, D. ve Kelecioğlu, H. (2011). Öğrenci Seçme Sınavı'nın madde yanlılığı açısından incelenmesi. *Eğitim ve Bilim*, 36 (161), 3-12.

- Camilli, G. & Shepard, L. A. (1994). *Methods for identifying biased test items*. London: Sage Publications
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedure to identify differential item functioning test items. *Educational Measurement: Issues 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(2), 115-124.
- Devine, P. J., & Raju N. S. (1982). Extent of overlap among four item bias methods. *Educational and Psychological Measurement*, 42, 1049–1066.
- Ellis, B. B., & Kimmel, H. D. (1992). Identification of unique cultural response patterns by means of item response theory. *Journal of Applied Psychology*, 77(2), 177.
- Fidalgo, A. M., & Madeira, J. M. (2008). Generalized Mantel-Haenszel methods for differential item functioning detection. *Educational and Psychological Measurement*, 68, 940–958.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gök, B., Atalay Kabasakal, K. ve Kelecioğlu, H. (2015). PISA 2009 Öğrenci Anketi Tutum Maddelerinin Kültüre Göre Değişen Madde Fonksiyonu Açısından İncelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 5/1, Yaz 2014, 72-87.
- Kamata, A. and Vaughn, B. (2004). An introduction to differential item functioning analysis. *Learning Disabilities: A Contemporary Journal*, 2(7), 49-69.
- Kanje, A. (2007). Using logistic regression to detect bias when multiple groups are tested. *South African Journal of Psychology*, 37, 47–61.
- Kim, S.-H., Cohen, A. S., & Park, T.-H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement*, 32, 261–276.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Magis, D., Raîche, G., Béland, S., & Gérard, P. (2011). A generalized logistic regression procedure to detect differential item functioning among multiple groups. *International Journal of Testing*, 11(4), 365-386.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719–748.
- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics*, 7, 105–118.
- Mellenberg, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research: Applications of Item Response Theory*. 13, 123–144.
- MEB (2013). PISA 2012 Uluslararası Öğrenci Değerlendirme Programı Ulusal Ön Raporu. MEB, Ankara.
- Osterlind, J. S. (1983). *Test item bias*. London: Sage Publications.
- Penfield, R. D. (2001). Assessing differential item functioning among multiple groups: a comparison of three Mantel-Haenszel procedures. *Applied Measurement in Education*, 14, 235–259.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, 14, 197–207.

Rodney, G. L., & Drasgow, F. (1990). Evaluation of two methods for estimating item response theory parameters when assessing differential item functioning. *Journal of Applied Psychology*, 75, 164-174.

Shealy, R.T., & Stout, W. (1993). A model based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika*, 58, 159–194.

Shepard, L. A., Camilli, G., & Williams, D. M. (1984). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement*, 22, 77–105.

Somes, G.W. (1986). The generalized Mantel–Haenszel statistic. *The American Statistician*, 40, 106–108.

Su, Y. H., & Wang, W. C. (2005). Efficiency of the Mantel, generalized Mantel–Haenszel, and logistic discriminant function analysis methods in detecting differential item functioning for polytomous items. *Applied Measurement in Education*, 18(4), 313-350.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361–370.

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group difference in trace lines. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 147–170). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tittle, C. K. (1988). Test Bias. In J.P. Keeves, (ed.). *Educational research, methodology, and measurement: An international handbook*. Oxford: Pergamon Press.

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Zwick, R., & Ercikan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. *Journal of Educational Measurement*, 26(1), 55-66.