

PAPER DETAILS

TITLE: Evaluation of attitude items in PISA 2006 student questionnaire in terms of differential item functioning

AUTHORS: Kübra ATALAY KABASAKAL,Hülya KELECIOGLU

PAGES: 77-96

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/508745>



Evaluation of attitude items in PISA 2006 student questionnaire in terms of differential item functioning^{*}

Kübra ATALAY KABASAKAL^{**}

Hülya KELECİOĞLU^{***}

ABSTRACT. Under the scope of this study, attitude items towards sciences were examined in PISA 2006 Student Survey is used to determine whether differential item functioning was observed in respect to gender and culture with ordinal logistic regression and poly-SIBTEST. Cultural differences and translational restrictions stand among the possible reasons of the observation of differential item functioning (DIF). Due to this reason, the research group of this study was determined as 5611 students from USA and 4942 students from Turkey whom participated in PISA application. In both methods, the number of items that showed DIF significantly and common is 7 in Turkey and 13 in the USA. In 30 items, with respect to culture, DIF was observed at a significant level through the analyses carried out in both countries utilizing both methods. When the above-mentioned two methods employed in the analyses, in terms of culture and gender, the data from USA and Turkey were compared, these two methods were found to be in compliance with each other, also with possible variations of their levels.

Key words: PISA, differential item functioning, ordinal logistic regression, poly-SIBTEST.

^{*} This study composed of part of the master thesis called "Evaluation of Attitude Items in PISA 2006 Student Questionnaire In Terms Of Differential Item Functioning"

^{**} Research Assist. Hacettepe University, Faculty of Education, Department of Educational Sciences, Ankara, Turkey. E-mail: katalay@hacettepe.edu.tr

^{***} Assoc. Prof. Dr., Hacettepe University, Faculty of Education, Department of Educational Sciences, Ankara, Turkey. E-mail: hulyaebb@hacettepe.edu.tr

SUMMARY

Purpose and Significance: In recent years, Differential Item Functioning (DIF) and item bias have received attention of many researchers. A test which not be biased is an important issue in the selection and the usage of any test. Test has to be fair for all participants and not to be biased against a certain group. Bias is a systematic error across to test results against a group of individual (Camilli, 2006; Camilli and Shepard, 1994). Researchers developed variety statistical methods to detect item bias and DIF has become popular technique to analyze item bias (Zumbo, 2007). DIF is an item bias index.

Item bias is a factor reducing validity not only cognitive tests but also psychological tests. Although there are many researches in DIF cognitive tests but there are few researchs in psychological tests. Under the scope of this study, attitude items towards sciences were examined in PISA 2006 Student Survey to determine whether differential item functioning was observed between the samplings of the USA and of Turkey in respect to gender and culture.

Method: In this research PISA 2006 student questionnaire was used. The research group of this study consists of the USA sample comprising 5611 students from 57 states participated in PISA application, and of the Turkey sample with 4942 students. These two countries were selected due to the fact that the cultural differences and translational restrictions stand among the possible reasons of the observation of differential item functioning (DIF).

Results: The analysis of the data collected in Turkey in terms of gender demonstrated DIF in 12 items when the OLR method was used, and in 13 items when poly-SIBTEST method was employed. The gender analysis of the data collected in the USA exhibited DIF in 15 items through both methods. In both methods, the number of items that showed DIF significantly and in common is 7 in Turkey and 13 in the USA. In 30 items, with respect to culture, DIF was observed at a significant level through the analyses carried out in both countries utilizing both methods.

Discussion and Conclusions: When OLR and poly-SIBTEST employed in the analyses, in terms of culture and gender, when the data from the USA and Turkey were compared, these two methods were found to be in compliance with each other, also with possible variations of their levels.



PISA 2006 Öğrenci Anketinde Yer Alan Maddelerin Değişen Madde Fonksiyonu Açısından İncelenmesi*

Kübra ATALAY KABASAKAL**

Hülya KELECİOĞLU***

ÖZ. Bu çalışma kapsamında PISA 2006 öğrenci anketinde yer alan fen bilimlerine ilişkin tutum maddeleri, cinsiyete göre ve kültüre göre değişen madde fonksiyonu gösterip göstermediğini incelenmektedir. Değişen madde fonksiyonu (DMF)'nin cinsiyetler arası farklılıklardan kaynaklanabileceği gibi, kültürel farklılıklardan ve testlerin başka bir dile çevrilmesinde karşılaşılan sorunlardan da kaynaklanabilmektedir. Bu nedenle araştırmanın çalışma grubu PISA uygulamasına katılan ABD'den 5611 öğrenci ve Türkiye'den 4942 öğrenci olarak belirlenmiştir. Cinsiyete göre yapılan analizlerde her iki yöntemde ortak ve önemli düzeyde DMF gösteren madde sayısı Türkiye'de 7, ABD'de 13'tür. İki ülke arasında kültüre göre her iki yöntemde yapılan analizlerde 30 maddede önemli düzeyde DMF görülmüştür. Analizler sonucu iki yöntem karşılaştırıldığında, yöntemlerin DMF belirlemede bir uyum gösterdiği; ancak düzeylerinde farklılıklar olabileceği sonucuna varılmıştır.

Anahtar Sözcükler: PISA, değişen madde fonksiyonu, ordinal lojistik regresyon, poly-SIBTEST.

* Bu makale "PISA 2006 Öğrenci Anketinde Yer Alan Maddelerin Değişen Madde Fonksiyonu Açısından İncelenmesi" adlı Yüksek Lisans tezinin bir kısmından oluşmaktadır.

** Arş. Gör. Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Ankara, Türkiye.

E-posta: katalay@hacettepe.edu.tr

*** Doç. Dr., Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Ankara, Türkiye.

E-posta: hulyaebb@hacettepe.edu.tr

GİRİŞ

Ölçme araçları uygulamalarındaki temel amaçlar arasında bireyleri tanımak, bireyler arasındaki farklılıkları ortaya çıkarmak, bireyleri yetenek düzeylerine göre ayırmak ve sıralamak yer almaktadır. Bazı durumlarda bireylerin ölçmek istediğimiz özelliklerine başka değişkenler karışabilir. Bu değişkenler test puanlarının güvenilirliğini düşürebileceği gibi, test puanların yanlış olmasına da yol açabilir. Yanlılık, ölçme sonuçları üzerinde belli bir gruba karşı sistematik hata anlamına gelmektedir (Camili,2006; Camili ve Shephard;1994). Psikolojik ve eğitsel testlerin belirli bir gruba karşı yanlış olduğu kararına varmak için nasıl bir kanıt gerekir? Yanlılığı belirlemek için çeşitli istatistiksel yöntemler geliştirilmiştir ve değişen madde fonksiyonları (DMF) tekniği psikometrik yanlılık analizlerinde bir standart haline gelmiştir. DMF, madde ile ölçülmek istenilen psikolojik yapının her bir yetenek düzeyinde maddeyi doğru yanıtlama olasılıklarının alt gruplara göre farklılık gösterip gösteremediğini belirleyen bir fonksiyondur (Zumbo, 1999). Değişen madde fonksiyonu ve madde yanlılığı belirlemeyle ilgili yöntemler yeni bir ölçme aracının geliştirildiği çalışmalarda ya da var olan bir ölçme aracının farklı dil veya kültür için uyarlama çalışmalarındaki madde analizi işlemlerinde kullanılır (Zumbo, 2007).

DMF iki farklı şekilde ortaya çıkabilir. İlki, yetenek düzeyindeki iki grubun ilgili maddeyi doğru cevaplama olasılığındaki farklılığın sabit olduğu tek biçimli DMF'dir. Tek biçimli DMF, bireylerin yetenek düzeyleri değiştikçe maddeyi doğru cevaplama olasılığındaki farklılığın da aynı düzeyde değişmesi durumudur (Jodoin ve Gierl, 2003). İkincisi ise tek biçimli olmayan DMF'dir. Bu da aynı yetenek düzeyindeki iki grubun ilgili maddeyi doğru cevaplama olasılığındaki farklılığın, tüm yetenek düzeylerinde tutarlı olmadığında görülür (Camili ve Shephard, 1994). DMF belirleme yöntemlerinin birçoğu tek biçimli DMF'yi belirlemek üzere tasarlanmıştır.(Jodoin ve Gierl, 2003).

DMF belirleme yöntemleri özellikle iki kategorili veriler için çok iyi yapılandırılmıştır. Tutum ölçeklerinde ve son yıllarda başarı testlerinde çok kategorili olarak puanlanabilen yapılandırılmış maddelerin kullanılması, bu tarz maddeler için de değişen madde fonksiyonu belirleme yöntemlerinin geliştirilmesine kaynak oluşturmuştur (Dorans ve Schmitt, 1993; Miller ve Spray, 1993; Welch ve Hoover,1993; Zwick ve diğerleri, 1993; Henderson, 1999). Hulin, Drasgow ve Parsons (1983) duyuşsal özelliği ölçen tutum maddelerinde DMF'yi, aynı özellikte farklı alt gruplarda bulunan bireylerin, bir maddeye olumlu tutum gösterme olasılığının farklı olması olarak tanımlamışlardır (Johanson ve Dodeen, 2003).

Çok kategorili veriler için kullanılan yöntemlerin çoğu iki kategorili veriler için kullanılan yöntemlerin bir uzantısı olmakla birlikte bu yöntemler daha karmaşıktır. Çok kategorili DMF belirleme yöntemleri arasında madde tepki kuramı (MTK) yöntemleri (Wainer, Sireci ve Thissen, 1991), Mantel (Mantel, 1963), Genelleştirilmiş Mantel-Haenszel (GMH, Zwick, Donoghue ve Grima, 1993), ordinal lojistik regresyon (OLR, Wilson, Spray ve Miller, 1993), Poly-SIBTEST (Chang, Mazzeo ve Raussos, 1993), lojistik diskriminant fonksiyon analizi (LDFA, Miller ve Sprey, 1993) ve standartlaştırma yöntemi (SD, Dorans ve Kulick, 1986) yer alır (Mellor, 1995; Henderson, 1996).

Araştırmada PISA 2006 öğrenci anketi verilerinin DMF analizleri yapılmıştır. Anket verileri çok kategorili olduğundan, bu verilere uygun olan klasik test kuramına dayalı yöntemlerden OLR ve Poly-SIBTEST kullanılmıştır.

Ordinal Lojistik Regresyon (OLR) : Lojistik regresyon iki kategorili verilerde DMF belirlemede oldukça etkili bir yöntemdir (French ve Miller, 1996; Swaminathan ve Rogers, 1990; Zumbo, 1999). French ve Miller (1996) lojistik regresyonun bir uzantısının çok kategorili maddelerde DMF belirlemede uygun olabileceğini önermiştir. Zumbo (1999), sıralama ölçeğindeki veriler için OLR yönteminin, genelleştirilmiş Mantel Haenszel (GMH) ve lojistik diskriminant fonksiyon analizine (LDFA) tercih edilmesi gerektiğini belirtmiştir. Çünkü OLR yöntemi hem iki kategorili hem çok kategorili verilerde kullanılabilir, yorumlaması kolaydır, etki büyüklüğü hesaplanabilir (Zumbo, 1999). Ordinal lojistik regresyon, Miller ve Spray (1993) tarafından çok kategorili veriler için önerilmekte birlikte Zumbo (1999) tarafından da geliştirilmiştir. Zumbo (1999) OLR yönetiminde madde cevaplarının bağımlı değişken, grup değişkeninin her birey için toplam puanla beraber grup toplam puan etkileşiminin bağımsız değişken olarak kullanıldığını belirtmiştir.

OLR yönteminde değişkenler modele hiyerarşik olarak sırayla eklenir. Modele ilk eklenen değişken toplam puan, ikinci değişken grup değişkeni (cinsiyet v.b.), son değişken ise toplam puan ile grup etkileşimidir. Üç aşamalı bu modelde, üç aşamada da DMF testinde kullanılacak olan ki-kare değeri elde edilir (Gelin ve Zumbo, 2003). OLR, lojistik regresyon gibi model temellidir; bundan dolayı tek biçimli ve tek biçimli olmayan DMF aynı denklemde modellenabilir ve katsayıları ayrı ayrı test edilebilir. Bu araştırmada, OLR yöntemi yukarıda belirtildiği gibi hem yorumlama kolaylığı açısından hem de tek biçimli olmayan DMF'yi de belirleyebildiği için tercih edilmiştir.

Her aşamada ki-kare değeri elde edildiği gibi, aynı zamanda etki büyüklüğü kestiricisi de elde edilir. Etki büyüklüğü R^2 değeri ile hesaplanır ve bu değer iki ve çok kategorili maddeler için kullanılabilir. Üçüncü

aşamada elde edilen R^2 değerinin, 1. Aşamada elde edilen R^2 değerinden çıkartılması ile etki büyüklüğü hesaplanır. Zumbo ve Thomas (1996) tarafından önerilen ΔR^2 'nin DMF'nin yorumlanmasında kullanılan ölçütler Tablo 1'de verilmiştir.

Tablo 1. ΔR^2 değerlerini yorumlama ölçütleri

A düzey DMF	$\Delta R^2 < 0,13$	DMF yoktur ya da ihmal edilebilir düzeydedir.
B düzey DMF	$0,13 \leq \Delta R^2 < 0,26$	Orta düzey DMF vardır
C düzey DMF	$\Delta R^2 \geq 0,26$	Yüksek düzeyde DMF vardır.

Poly-SIBTEST: SIBTEST, Shealy ve Stout (1993) tarafından iki kategorili verilerde DMF belirlemek için geliştirilmiş, parametrik olmayan örtük değişken modelidir (Potenza ve Dorans, 1995). SIBTEST DMF belirlemede, standartlaştırma yöntemi ile benzerdir; ancak standartlaştırma yönteminden farklı olarak 1. tip hatayı kontrol etmede kullanışlı olan regresyon temelli düzeltmeyi kullanılır (Clauser ve Mazor, 1998; Fang, 1999). Chang, Mazzeo ve Roussos (1996), SIBTEST yönteminin çok kategorili veriler için farklı formunu geliştirmişlerdir. Poly-SIBTEST örtük değişkenler için tanımlanmasına rağmen, bu yöntem MTK ile kestirilen yetenek ve madde parametrelerini kullanmaz. Bu yöntem, tek biçimli DMF'yi belirlemek için tasarlanmıştır. Bu araştırmada Poly-SIBTEST yöntemi, 1. tip hatayı daha kullanışlı bir yolla kontrol ettiği için tercih edilmiştir. Ayrıca, tek ve çok biçimli DMF belirleyebilen OLR yöntemi ile farklılığı da ortaya konmak istenmiştir.

SIBTEST yöntemi sadece bir maddenin DMF gösterip göstermediğini değil, aynı zamanda ne kadar DMF gösterdiğini de belirler. SIBTEST etki büyüklüğü β , her bir madde için DMF miktarı olarak yorumlanır. Pozitif β değeri o maddenin referans grubun lehine, negatif β değeri ise o maddenin odak grubun lehine çalıştığını gösterir. β değerlerini yorumlama ölçütlerine Tablo 2'de yer verilmiştir.

Tablo 2. β Değerlerini Yorumlama Ölçütleri

A düzey DMF	$ \beta < 0,059$	DMF yoktur ya da ihmal edilebilir düzeydedir.
B düzey DMF	$0,059 \leq \beta < 0,088$	Orta düzey DMF vardır
C düzey DMF	$ \beta \geq 0,088$	Yüksek düzeyde DMF vardır.

Son zamanlarda objektif ölçme sonuçları elde edebilmek için madde yanlılığı ve DMF ile ilgili yapılan çalışmalar sayıca artmıştır. Yanlılık, sadece bilişsel testlerde değil aynı zamanda psikolojik testlerde de geçerliği

düşüren bir faktördür. Bilişsel yetenek ölçen testlerde madde yanlılığı üzerine çok sayıda çalışma yapılmıştır; ancak psikolojik ölçmelerde bireylerin diğer özelliklerinin ölçme sonuçlarına etkisi bilişsel özellikleri ölçen testler kadar iyi bilinmemektedir. DMF belirleme ile ilgili hem duyuşsal alanda hem de çok kategorili veriler üzerinde yapılan çalışmaların azlığı, bu çalışmanın çıkış noktasını oluşturmaktadır.

DMF'nin olası nedenleri arasında cinsiyet, kültürel farklılıklar, çeviri sorunları v.b yer alır Yapılan araştırmalarda çoğunlukla bu değişkenlerin alındığı görülmektedir (Dancer, Anderson ve Derlin,1994; Mellor,1995; Allaouf, Hambleton ve Sireci,1999; Henderson,1999; Sireci ve Khaliq, 2002; Feng,2008; Smith,2009; Asil,2010) . Son yıllarda bilgi, yetenek ve tutum gibi özellikleri ölçmek amacıyla farklı dilleri konuşan bireylere çeviri testlerin uygulamasında artış görülmektedir. PISA, TIMSS ve PIRLS bu uygulamaların örnekleri arasındadır. Bu sınavlarda öğrencilerin bilgi ve beceri düzeylerinin yanı sıra ilgi, tutum ve öz yeterlilik gibi duyuşsal özellikleri de belirlenmektedir. Sınava katılan ülkeler, bu sonuçları değerlendirerek öğretim programlarında düzenlemelere gidebilmektedir. Bilişsel alandaki başarıyı açıklayan değişkenler arasında tutum, özyeterlik gibi duyuşsal özelliklerin de yer aldığı çeşitli araştırmalarla ortaya konmuştur (Atar, 2010; Özdemir, 2010; Anıl, 2009; Barış, 2009; Bloom, 1989). Bu nedenle, öğrencilerin duyuşsal özelliklerinin de en az bilişsel özellikler kadar geçerli ölçülmesi gerekir. PISA 2006'da öğrenci anketi ile ölçülen özellikler öğrencilerin cinsiyetine ve kültürüne göre değişiklik gösterebilir. Bu durum ise anketlerin geçerliğini düşürerek, öğrencilerin hem duyuşsal hem de bilişsel özelliklerinin hatalı olarak belirlenmesine yol açar.

Bu nedenle, PISA 2006 öğrenci anketinin ölçtüğü psikolojik yapının, kültüre ve cinsiyete göre değişen madde fonksiyonu gösterip göstermediğinin belirlenmesi, bu araçtan elde edilecek bilgilerle yapılan yorumların ve karşılaştırmaların geçerliğini artıracaktır. Bu çalışmanın amacı, PISA 2006 öğrenci anketlerindeki maddelerin, cinsiyete ve farklı ülkelere göre DMF gösterip göstermediğinin belirlenmesi poly-SIBTEST ve ordinal lojistik regresyon yöntemleri ile incelenmesidir.

YÖNTEM

Bu çalışma, PISA 2006 araştırmasında yer alan fen bilimlerine yönelik 32 tutum maddesinin cinsiyete ve farklı kültürlere göre değişen madde fonksiyonu içerip içermediğini OLR ve Poly-SIBTEST yöntemleri ile inceleyen betimsel bir çalışmadır.

Ölçme Aracı ve Çalışma Grubu

Araştırmanın verileri, Ekonomik İşbirliği ve Kalkınma Örgütü'nün 2006 yılında uyguladığı PISA 2006 öğrenci anketi PISA internet sitesinden elde edilmiştir (<http://pisa2006.acer.edu.au/>). Araştırmada kullanılan öğrenci anket 4'lü Likert tipi ölçektedir. Ölçek maddeleri “tamamen katılıyorum” için 1, “katılıyorum” için 2, “katılmıyorum için” 3, “tamamen katılmıyorum” için 4 olarak kodlanmıştır. Öncelikle, maddelerin kodları yorumlamayı kolaylaştırmak amacı ile ters çevrilmiş ve araştırmada kullanılan yöntemler için puanlamanın 0'dan başlaması gerektiğinden en küçük kod 0 olacak şekilde yeniden düzenlenmiştir. Yeni kodlama, “tamamen katılmıyorum” için 0, “katılmıyorum” için 1, “katılıyorum” için 2, “tamamen katılıyorum” için 3 şeklindedir.

PISA 2006 uygulamasında ülkelere ait örneklemeler PISA konsorsiyumu tarafından iki tabakalı örnekleme yoluyla seçilmiştir. Uygulamaya 57 ülkeden toplam 398750 öğrenci katılmıştır. Maddelerin değişen madde fonksiyonu göstermesinin nedenlerinden biri de testlerin geliştirildiği dilden yapılan çevirilerdir. Bu nedenle çalışmada ABD ve Türkiye alınmıştır. ABD örnekleminde 5611, Türkiye örnekleminde ise 4942 öğrenci uygulamaya katılmıştır.

PISA öğrenci anketinde yer alan 32 tutum maddesi beş farklı bölümde yer almaktadır. Bu bölümler ve bölümlerin içerdiği madde sayıları Tablo 3'te verilmiştir.

Tablo 3. PISA 2006 öğrenci anketi bölümleri ve madde sayıları

Bölümler	Madde sayısı
Fen bilimlerini karşı ilgi (ST16)	5
Fen bilimlerine verilen değer (ST18)	10
Fen bilimlerine karşı sorumluluk(ST26)	7
Fen bilimlerinin geleceği (ST29)	4
Fen bilimlerini öğrenme (ST37)	6

Tablo 3'te yer alan her bir bölüm için ayrı ayrı DMF analizleri gerçekleştirilmiştir. Verilerin analizinde OLR yöntemi için Zumbo (1999) tarafından yazılan syntax, Poly-SIBTEST için Sheally ve Stout (1993) tarafından hazırlanan bilgisayar programı kullanılmıştır.

BULGULAR

PISA öğrenci anketinde beş alt boyutu oluşturan 32 maddenin cinsiyete ve kültürlere göre farklı fonksiyonlaşıp fonksiyonlaşmadığı ordinal lojistik regresyon ve poly-SIBTEST yöntemleri ile incelenmiş ve yöntemlerin uyumu karşılaştırılmıştır.

PISA 2006 öğrenci anketlerinde yer alan tutum maddelerinin Türkiye uygulamasında cinsiyete göre OLR analizi sonuçları Tablo 4'te Poly-SIBTEST analizi sonuçları Tablo 5'de verilmiştir. Türkiye verisinde cinsiyete göre Poly-SIBTEST ve OLR yöntemlerinin uyumu ise Tablo 6'da verilmiştir.

Tablo 4. Türkiye verisinde cinsiyete göre OLR analizi sonuçları

Maddeler	DMF χ^2	p	Model1 R^2	Model2 R^2	Model3 R^2	DMF ΔR^2	Düzyey	DMF türü
ST16Q02	11,492	0,003	80,58	80,64	80,66	0,08	A	TB
ST16Q03	11,144	0,004	75,66	75,75	75,75	0,09	A	TB
ST18Q02	14,794	0,001	53,17	53,36	53,38	0,21	B	TB
ST18Q06	12,14	0,002	55,48	55,67	55,66	0,18	B	TB
ST18Q07	33,609	0,000	60,82	61,16	61,13	0,31	C	TB
ST18Q010	31,387	0,000	49,40	49,74	49,74	0,34	C	TB
ST26Q05	13,062	0,001	56,55	56,64	56,75	0,20	B	TBO
ST26Q07	24,283	0,000	50,56	50,77	50,84	0,28	C	TB
ST29Q02	9,267	0,010	89,23	89,25	89,29	0,06	A	TBO
ST29Q04	18,311	0,000	85,35	85,42	85,47	0,12	A	TB
ST37Q02	10,191	0,006	80,88	80,95	80,94	0,06	A	TB
ST37Q04	15,149	0,001	82,93	82,99	82,99	0,06	A	TB

Tablo 4'te 32 madde içinde DMF gösteren 12 madde bu maddelerdeki etki büyüklüğü ve DMF'nin türü verilmiştir. 12 maddeden 6 madde A düzeyinde DMF göstermesi nedeniyle göz ardı edilebilir. Fen bilimlerine verilen değer alt boyutunda (ST18) yer alan 2. ve 6. maddeler B düzeyinde, 7. ve 10. maddeler ise C düzeyinde DMF göstermektedir. Bu maddelerde Model3R² –Model2R² farkı incelendiğinde belirtilen maddelerin tek biçimli DMF gösterdiği görülmüştür. Fen bilimlerine karşı sorumluluk alt boyutunda (ST26) ise 5. madde B, 7. madde C düzeyinde DMF göstermektedir. Bu maddelerde Model3R² –Model2R² farkı incelendiğinde 5. maddenin tek biçimli olmayan, 7. maddenin ise tek biçimli DMF gösterdiği görülmüştür.

Tablo 5. Türkiye verisinde cinsiyete göre Poly-SIBTEST analizi sonuçları

Maddeler	Beta estimate	Standart e	p	Düzye	Cinsiyet
ST16Q02	-0.044	0.014	0.002	A	K
ST16Q03	0.069	0.018	0.000	B	E
ST16Q04	-0.038	0.013	0.005	A	K
ST18q02	-0.052	0.015	0.000	A	K
ST18q05	-0.058	0.015	0.000	B	K
ST18q06	0.116	0.019	0.000	C	E
ST18q10	0.101	0.022	0.000	C	E
ST26q02	-0.056	0.016	0.001	A	K
ST26q07	0.046	0.018	0.009	A	E
ST29Q04	-0.064	0.017	0.000	B	K
ST37Q03	-0.044	0.014	0.001	A	K
ST37Q04	0.049	0.015	0.001	A	E
ST37Q05	-0.049	0.014	0.001	A	K

Tablo 5’te 32 madde içinde DMF gösteren 13 madde, bu maddelerdeki etki büyüklüğü ve DMF’nin türü verilmiştir. Bu 13 madde incelendiğinde 8’nin A düzeyinde, 3’ünün B düzeyinde 2’sinin ise C düzeyinde DMF gösterdiği görülmektedir. B düzeyinde DMF gösteren maddelerden fen bilimlerine ilgi alt boyutunda yer alan ST16Q03 erkeler lehine, fen bilimlerine verilen değer alt boyutunda yer alan ST18Q06 ve fen bilimlerinin geleceği alt boyutunda yer alan ST29Q04 maddeleri kızlar lehine farklı fonksiyonlaşmaktadır. C düzeyinde DMF gösteren fen bilimlerine verilen değer alt boyutunda yer alan ST18Q07 ve ST18Q10 maddeleri erkekler lehine farklı fonksiyonlaşmaktadır.

Tablo 6. Türkiye verisinde cinsiyete göre Poly-SIBTEST ve OLR analizi karşılaştırması

	OLR			Poly-SIBTEST		
	A	B	C	A	B	C
OLR	A	ST29Q02		ST16Q02	ST16Q03	
		ST37Q02		ST37Q04	ST29Q04	
	B		ST26Q05	ST18Q02	ST18Q06	
	C			ST26Q07		ST18Q07 ST18Q10
Poly-SIBTEST	A			ST16Q04		
				ST26Q02		
				ST37Q03		
				ST37Q05		
	B					
	C					

Tablo 6 incelendiğinde poly-SIBTEST yöntemiyle 13 maddenin, OLR yöntemiyle 12 maddenin DMF gösterdiği görülmektedir. Her iki yöntemde birden DMF gösteren 9 madde bulunmaktadır. Bu 9 maddenin içinden ST16Q02, ST18Q06, ST18Q07, ST18Q10 ve ST37Q04 maddeleri ortaktır ve iki yöntemde de aynı düzeyde DMF göstermektedirler. Poly-SIBTEST yönteminde B düzeyinde DMF gösteren 2 madde (ST16Q03 ve ST29Q04) OLR yönteminde A düzeyinde DMF göstermektedir. Bununla birlikte OLR yönteminde C düzeyinde DMF gösteren ST26Q07 maddesi ve B düzeyinde DMF gösteren ST18Q02 maddesi poly- SIBTEST yönteminde A düzeyde DMF göstermektedir.

Sadece poly-SIBTEST yöntemiyle A düzeyde DMF gösteren 4 madde bulunmaktadır. Sadece OLR yöntemiyle DMF gösteren 3 madde bulunmaktadır. Bu maddelerden ST29Q02 ve ST37Q02 kodlu maddeler A düzeyinde, ST26Q05 maddesi ise OLR analizine göre B düzeyinde ve tek boyutlu olmayan DMF göstermektedir.

PISA 2006 öğrenci anketlerinde yer alan tutum maddeleri ABD uygulamasında cinsiyete göre OLR analizi sonuçları Tablo 7’de Poly-SIBTEST analizi sonuçları Tablo 8’de verilmiştir. ABD verisinde cinsiyete göre Poly-SIBTEST ve OLR yöntemlerinin uyumu ise Tablo 9’da verilmiştir.

Tablo 7. ABD verisinde cinsiyete göre OLR analizi sonuçları

Maddeler	DMF χ^2	p	Model1 R ²	Model2 R ²	Model3 R ²	DMF ΔR^2	Düzyey	DMF türü
ST16Q01	15,775	0,000	86,73	86,79	86,87	0,14	B	TB
ST16Q04	9,97	0,010	88,31	88,36	88,40	0,09	A	TB
ST18Q01	24,058	0,000	50,09	50,26	50,67	0,58	C	TBO
ST18Q02	13,667	0,001	59,52	59,64	59,80	0,28	C	TBO
ST18Q09	13,254	0,001	56,96	57,10	57,14	0,18	B	TB
ST26Q01	22,897	0,000	50,94	51,30	51,32	0,38	C	TB
ST26Q02	17,252	0,000	51,57	51,64	51,76	0,19	B	TBO
ST26Q03	40,416	0,000	54,71	55,14	55,16	0,45	C	TB
ST26Q04	13,039	0,001	59,61	59,76	59,77	0,16	B	TB
ST26Q05	51,518	0,000	54,86	55,46	55,48	0,62	C	TB
ST26Q06	25,314	0,000	57,98	58,28	58,28	0,31	C	TB
ST29Q01	25,809	0,000	89,38	89,44	89,45	0,07	A	TB
ST29Q02	42,036	0,000	89,54	89,64	89,67	0,13	B	TB
ST29Q03	16,198	0,000	89,53	89,60	89,59	0,06	A	TB
ST37Q02	30,089	0,000	78,68	78,87	78,87	0,19	B	TB

Tablo 7’de 32 madde içinden ki-kare değeri anlamlı olan 15 madde verilmiştir. Bu 15 maddede Model3R² –Model1R² farkı incelendiğinde 3

madde A düzeyinde, 6 madde B düzeyinde, 6 madde ise C düzeyde DMF içermektedir. DMF içeren maddelerde Model3R² –Model2R² farkı incelendiğinde bu maddelerden C düzeyinde DMF içeren ST18Q01 ve ST18Q02 maddeleri ile B düzeyinde DMF içeren ST26Q02 maddesi tek biçimli olmayan şekilde fonksiyonlaşmıştır.

Tablo 8. ABD verisinde cinsiyete göre Poly-SIBTEST analizi sonuçları

Maddeler	Beta estimate	Standart e	p	Düzye	Cinsiyet
ST16Q01	-0.069	0.013	0.000	B	K
ST16Q05	-0.038	0.013	0.003	A	K
ST18001	0.052	0.015	0.000	A	E
ST18q02	-0.041	0.013	0.002	A	K
ST18q05	-0.052	0.016	0.002	A	K
ST18q09	0.048	0.016	0.003	A	E
ST26q01	0.073	0.015	0.000	B	E
ST26q02	-0.050	0.018	0.006	A	K
ST26q03	0.121	0.019	0.000	C	E
ST26Q04	-0.057	0.016	0.000	A	K
ST26Q06	-0.114	0.016	0.000	C	K
ST26Q07	0.078	0.018	0.000	B	E
ST29Q01	-0.082	0.016	0.000	B	K
ST29Q02	-0.102	0.015	0.000	C	K
ST29Q04	0.074	0.018	0.000	B	E

Referans grup: erkek, Odak grup:kız

Tablo 8'e göre 32 maddenin 15'inde beta değeri anlamlıdır. Bu 15 madde incelendiğinde 7'sinin A düzeyinde, 5'inin B düzeyinde 3'ünün ise C düzeyinde DMF gösterdiği görülmektedir. B düzeyinde DMF gösteren maddelerden ST16Q01 erkeler lehine, ST26Q01, ST26Q07, ST29Q01 ve ST29Q04 kızlar lehine farklı fonksiyonlaşmaktadır. C düzeyinde DMF gösteren ST26Q03 maddesi erkeklerin, ST26Q06 ve ST29Q02 kodlu maddeler ise kızlar lehine farklı fonksiyonlaşmaktadır.

Tablo 9 incelendiğinde, her iki yöntemle de 15 maddenin DMF gösterdiği gözlenmektedir. Her iki yöntemde birden DMF gösteren 13 madde bulunmaktadır. Bu 13 madde içinden ST16Q01, ST26Q03, ST26Q06 ve ST29Q04 maddeleri ortaktır ve iki yöntemde de aynı düzeyde DMF göstermektedirler. Poly- SIBTEST yönteminde A düzeyinde DMF gösteren maddelerden ST18Q09, ST26Q02 VE ST26Q04 maddeleri OLR yöntemiyle B düzeyinde, ST18Q01 ve ST18Q02 maddeleri ise OLR yöntemiyle C düzeyinde DMF göstermektedir. OLR yönteminde A düzeyinde DMF gösteren ST29Q01 ve C düzeyinde DMF gösteren ST26Q01 ve ST26Q07 Poly-SIBTEST yöntemiyle B düzeyinde DMF göstermektedir.

Tablo 9. ABD verisinde cinsiyete göre Poly-SIBTEST ve OLR analizi karşılaştırması

	OLR			Poly-SIBTEST		
	A	B	C	A	B	C
OLR	A	ST16Q04 ST29Q03			ST29Q01	
	B			ST18Q09 ST26Q02 ST26Q04	ST16Q01 ST29Q04	ST29Q02
	C			ST18Q01 ST18Q02	ST26Q01 ST26Q07	ST26Q03 ST26Q06
Poly-SIBTEST	A			ST16Q05 ST18Q05		
	B					
	C					

Sadece Poly-SIBTEST yöntemiyle A düzeyinde DMF gösteren 2 madde sadece OLR yöntemiyle DMF gösteren 2 madde bulunmaktadır.

PISA 2006 öğrenci anketlerinde yer alan tutum maddeleri Türkiye ve ABD uygulamasında kültüre göre OLR analizi sonuçları Tablo 10'de Poly-SIBTEST analizi sonuçları Tablo 11'de verilmiştir. Türkiye ve ABD verisinde cinsiyete göre Poly-SIBTEST ve OLR yöntemlerinin uyumu ise Tablo 12'da verilmiştir.

Tablo 10. ABD-Türkiye verisinde kültüre göre OLR analizi sonuçları

Maddeler	DMF X ²	p	Model1 R ²	Model2 R ²	Model3 R ²	DMF ΔR^2	Düzy	DMF Türü
ST16Q01	13,29	0,001	87,02	87,05	87,05	0,03	A	TB
ST16Q02	481,79	0,000	84,90	86,17	86,40	1,5	C	TB
ST16Q03	119,64	0,000	80,78	81,03	81,04	0,26	C	TB
ST16Q04	110,39	0,000	87,50	87,82	87,85	0,35	C	TB
ST16Q05	42,98	0,000	87,54	87,64	87,72	0,18	B	TB
ST18Q01	119,721	0,000	49,81	50,59	50,70	0,89	C	TB
ST18Q02	29,124	0,000	56,33	56,63	56,65	0,32	C	TB
ST18Q03	28,174	0,000	50,99	51,17	51,17	0,18	B	TB
ST18Q04	16,836	0,000	46,24	46,36	46,40	0,16	B	TB
ST18Q05	42,159	0,000	63,82	63,99	64,06	0,24	B	TB
ST18Q06	224,037	0,000	63,18	64,35	64,32	1,14	C	TB
ST18Q07	109,44	0,000	64,88	65,40	65,40	0,62	C	TB
ST18Q08	37,353	0,000	65,42	65,63	65,62	0,20	B	TB
ST18Q09	529,56	0,000	54,06	57,35	57,34	3,28	C	TB
ST18Q010	874,963	0,000	52,52	57,06	57,05	4,53	C	TB
ST26Q01	101,807	0,000	59,02	59,26	60,19	1,17	C	TBO
ST26Q02	240,682	0,000	61,28	62,30	62,51	1,23	C	TBO
ST26Q03	942,624	0,000	65,69	69,80	69,97	4,28	C	TB
ST26Q04	71,590	0,000	64,06	64,12	64,62	0,56	C	TBO
ST26Q05	467,037	0,000	56,50	59,88	60,09	3,59	C	TB
ST26Q06	182,967	0,000	54,88	56,31	56,44	1,56	C	TB
ST26Q07	219,908	0,000	55,95	57,56	57,57	1,62	C	TB
ST29Q01	164,	0,000	88,21	88,38	88,45	0,24	B	TB
ST29Q02	443,74	0,000	88,55	89,50	89,48	0,93	C	TB
ST29Q03	116,32	0,000	89,05	89,29	89,35	0,30	C	TB
ST29Q04	85,68	0,000	82,94	83,77	83,81	0,87	C	TB
ST37Q01	348,11	0,000	80,97	81,62	82,34	1,37	C	TB
ST37Q02	134,67	0,000	82,29	82,81	82,81	0,52	C	TB
ST37Q04	29,16	0,000	78,09	78,15	78,24	0,15	B	TBO
ST37Q05	306,08	0,000	78,64	79,44	80,02	1,38	C	TB
ST37Q06	46,71	0,000	84,44	84,48	84,59	0,15	B	TBO

Tablo 10'a göre 32 maddenin 31'inde ki-kare değeri anlamlıdır. Bu 31 maddede Model3R²-Model1R² farkı incelendiğinde 1 madde A düzeyinde, 8 madde B düzeyinde, 22 madde ise C düzeyde DMF içermektedir. DMF içeren maddelerde Model3R²-Model2R² farkı incelendiğinde 5 tanesinde tek biçimli olmayan DMF gözlenmektedir.

Tablo 11. ABD-Türkiye verisinde kültüre göre Poly- SIBTEST analizi sonuçları

Maddeler	Beta estimate	Standart e	p	Düzy	Ülke
ST16Q01	0.055	0.010	0.000	A	Tr
ST16Q02	0.243	0.010	0.000	C	Tr
ST16Q03	-0.073	0.012	0.000	B	ABD
ST16Q04	-0.070	0.009	0.000	B	ABD
ST16Q05	-0.024	0.010	0.017	A	ABD
ST18Q01	0.110	0.011	0.000	C	Tr
ST18Q02	-0.053	0.010	0.000	A	ABD
ST18Q03	0.084	0.012	0.003	B	Tr
ST18Q04	-0.036	0.012	0.001	A	ABD
ST18Q05	0.077	0.012	0.000	B	Tr
ST18Q06	0.139	0.010	0.000	C	Tr
ST18Q07	-0.117	0.013	0.000	C	ABD
ST18Q08	-0.065	0.011	0.000	B	ABD
ST18Q09	0.280	0.012	0.000	C	Tr
ST18Q010	-0.368	0.014	0.000	C	ABD
ST26Q01	0.057	0.015	0.000	A	Tr
ST26Q02	0.238	0.015	0.000	C	Tr
ST26Q03	0.436	0.015	0.000	C	Tr
ST26Q04	-0.085	0.014	0.000	B	ABD
ST26Q05	-0.312	0.016	0.000	C	ABD
ST26Q06	-0.182	0.015	0.000	C	ABD
ST26Q07	-0.191	0.016	0.000	C	ABD
ST29Q01	-0.062	0.012	0.000	B	ABD
ST29Q02	-0.179	0.011	0.000	C	ABD
ST29Q03	0.194	0.011	0.000	C	Tr
ST29Q04	0.314	0.013	0.000	C	Tr
ST37Q01	0.144	0.010	0.000	C	Tr
ST37Q02	-0.107	0.010	0.000	C	ABD
ST37Q04	0.064	0.011	0.011	B	Tr
ST37Q05	-0.129	0.010	0.000	C	ABD
ST37Q06	0.051	0.010	0.000	A	Tr

Referans grup: Tr, Odak grup: ABD

Tablo 11'e göre 32 maddenin 31'inde beta değeri anlamlıdır. Bu 31 madde incelendiğinde 6'sının A düzeyinde, 8'inin B düzeyinde, 17'sinin ise C düzeyinde DMF gösterdiği görülmektedir. DMF gösteren maddelerin 15'i Türkiye lehine çalışırken, 16'sı ABD lehine çalışmaktadır.

Tablo 12 incelendiğinde her iki yöntemde de DMF gösteren 31 madde bulunduğu görülmüştür. ST37Q03 kodlu madde her iki yöntemde de DMF göstermeyen tek maddedir. Her iki yöntemde DMF gösteren maddelerden

21'i aynı düzeyde DMF göstermektedir. Poly-SIBTEST yönteminde A düzeyde DMF gösteren ST16Q05, ST18Q04, ST37Q06 maddeleri OLR yönteminde B düzeyinde, ST18Q02 ve ST26Q01 maddeleri ise OLR yönteminde C düzeyinde DMF göstermektedir. OLR yönteminde C düzeyde DMF gösteren maddeler ST16Q03, ST16Q04, ST18Q05, ST26Q04 ve ST37Q04 poly-SIBTEST yönteminde B düzeyinde DMF göstermektedir.

Tablo 12. ABD-Türkiye verisinde kültüre göre poly-SIBTEST ve OLR analizi karşılaştırması

OLR		Poly-SIBTEST			
		A	B	C	
OLR	A	ST16Q01			
	B	ST16Q05	ST18Q03		
		ST18Q04	ST18Q08		
		ST37Q06	ST29Q01		
	C	ST18Q02	ST16Q03	ST16Q02	ST26Q06
		ST26Q01	ST16Q04	ST18Q01	ST26Q07
			ST18Q05	ST18Q06	ST29Q02
			ST26Q04	ST18Q07	ST29Q03
			ST37Q04	ST18Q09	ST29Q04
				ST18Q10	ST37Q01
				ST26Q02	ST37Q02
				ST26Q03	ST37Q05
				ST26Q05	

TARTIŞMA VE SONUÇ

Bu araştırmada, OECD tarafından hazırlanan PISA 2006 uygulamasında yer alan fen bilimlerine yönelik 32 tutum maddesinde DMF olup olmadığı araştırılmıştır. PISA 2006 öğrenci anketinde yer alan 32 madde 5 alt boyutta yer almaktadır. DMF, öğrencilerin cinsiyet (kız-erkek) ve kültür (ABD-Türkiye) değişkenleri temel alınarak incelenmiştir. İncelemeler, OLR ve poly-SIBTEST yöntemleri kullanılarak yapılmıştır. Elde edilen bilgiler ışığında ulaşılan sonuçlar şöyle özetlenebilir:

Türkiye uygulamasında, iki yöntemle önemli düzeyde DMF gösteren toplam madde sayısı 8'dir. Bu maddelerin 7'sinin (%87,5'inin) her iki yöntemde de DMF gösterdiği ve 3'ünün (%42,9'unun) ise aynı düzeyde DMF gösterdiği görülmektedir.

ABD uygulamasında, iki yöntemle önemli düzeyde DMF gösteren toplam madde sayısı 13'tür. Bu maddelerin 13'ünün (%100'ü) her iki yöntemde de DMF gösterdiği ve 4'ünün (%30,8'inin) ise aynı düzeyde DMF gösterdiği görülmektedir.

PISA uygulamalarında yer alan bilişsel alanla ilgili maddelerde cinsiyete göre yapılmış çok sayıda DMF çalışması bulunmaktadır. Le (2009), PISA gibi uluslararası büyük ölçekli sınavlarda cinsiyete göre DMF gösteren maddelerin yer almasının kaçınılmaz olduğunu belirtmiştir. Bu araştırmada ulaşılan sonuçlar, cinsiyete göre DMF'nin tutum maddelerinde de ortaya çıkabileceğini göstermektedir.

ABD ve Türkiye uygulamasında, iki yöntemle önemli düzeyde DMF gösteren toplam madde sayısı 30'dur. Bu maddelerin 30'unun (%100'ünün) her iki yöntemde de DMF gösterdiği ve 20'sinin (%66,6'sının) ise aynı düzeyde DMF gösterdiği görülmektedir.

Türkiye ve ABD uygulamasında cinsiyete göre ve bu iki ülke arasında kültüre göre yapılan analizlerde iki yöntem karşılaştırdığında, yöntemlerin DMF belirlemede bir uyum gösterdiği; ancak düzeylerinde farklılıklar olabileceği sonucuna varılmıştır.

Büyük ölçekli sınavlarda yapılan DMF çalışmalarında, DMF'nin nedenlerine çeviri problemlerinin, kültürel farklılıkların ve eğitim programlarındaki farklılıkların neden olabileceği sonucuna varılmıştır (Yıldırım ve Berberoğlu, 2009; Le, 2009). Asil(2010) ise yaptığı çalışmada dilsel ve kültürel farklılıklar arttıkça DMF gösteren maddelerin arttığı sonucuna varmıştır. Bu nedenle çeviri testlerde DMF görülmesinin nedenlerinin kestirilmesi; test geliştirmede, test eşitlemede ve puanlamada önemli yararlar sağlayacaktır (Allaouf ve diğerleri, 1999).

PISA anketlerinden elde edilen veriler ile öğrenci başarısı ile ilgili faktörler belirlenmeye çalışılmakta, ülkeler arası kıyaslamalar yapılmakta ve belirli kararlar alınmaktadır. PISA 2006 fen bilimlerine karşı tutum maddelerinde cinsiyete ve kültüre karşı DMF gösteren maddeler yer alması nedeniyle bu anketlerde düzenli olarak DMF analizleri yapılması önerilebilir. Ayrıca cinsiyete ve kültüre göre DMF gösteren maddelerin uzmanlar tarafından değerlendirilerek bunların olası nedenlerinin tespit edilmesi ve benzer maddelerin ileriki yıllarda hazırlanırken dikkat edilmelidir.

Yapılan çalışmalar incelendiğinde (Kristjansson, Aylesworth, McDowell ve Zumbo 2005, Fang, 1999; Chang, Mazzeo ve Rousoss, 1996; Mellor, 1995), farklı DMF belirleme yöntemlerinin örneklem büyüklüğü, odak ve referans grup örneklem büyüklüğü, odak ve referans grup yetenek dağılımı, madde karakteristikleri gibi değişkenlere göre farklı çalıştığı görülmektedir. Ayrıca yöntemlerin 1. Tip hatayı kontrol güçleri ve tek biçimli olan ve olmayan DMF'yi belirlemedeki güçleri farklı koşullar altında değişmektedir. Bu araştırmada OLR ve poly-SIBTEST yöntemlerinden yararlanılmıştır. Öncelikli olarak çok kategorili veriler için simülasyon çalışmasıyla veriler üretilerek DMF belirlemede hangi yöntemin

hangi durumlar için daha uygun olduğu belirlendikten sonra LDFA, GMH gibi farklı DMF belirleme yöntemleri ile benzeri çalışmalar yapılabilir. Farklı çalışmalardan elde edilen sonuçlar karşılaştırılabilir.

Bu çalışmada iki farklı ülkede cinsiyet değişkenine göre DMF çalışması yapılmıştır. Cinsiyete dayalı yapılan DMF analizlerinde, ABD uygulamasında Türkiye uygulamasından daha fazla sayıda maddede cinsiyete göre DMF görülmüştür ve bu maddelerin çoğu ortak değildir. Diğer ülkelerde de cinsiyete göre DMF analizleri yapılarak, bu maddelerin karşılaştırılabilir.

Bu çalışmada kültürler ve dilsel farklılıkları olan iki ülke üzerinde çalışılmıştır. Yapılacak diğer çalışmalarda benzer kültüre sahip farklı diller kullanan, farklı kültürleri olan aynı dili kullanan ülkeler arasında DMF çalışması yapılabilir.

KAYNAKLAR

- Allaouf, A., Hambleton, R. & Sireci, S. (1999). Identifying the cause of DIF in translated verbal items. *Journal of Educational Measurement*, 36 (3), 185–198.
- Anıl, D. (2009). Uluslararası Öğrenci Başarılarını Değerlendirme Programı (PISA)'nda Türkiye'deki öğrencilerin fen bilimleri başarılarını etkileyen faktörler. *Eğitim ve Bilim*, 35 (152), 87-100.
- Atar, B. (2010). Basit doğrusal regresyon analizi ile hiyerarşik doğrusal modeller analizinin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 78-84.
- Asil, M. (2010). *Uluslararası Öğrenci Değerlendirme Programı (PISA) 2006 Öğrenci Anketinin Kültürler Arası Eşdeğerliğinin İncelenmesi*. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, yayınlanmamış doktora tezi.
- Barış, F. (2010). *TIMMS-R ve TIMSS-2007 sınavlarının öğrenci başarısını yordayan değişkenler açısından incelenmesi*. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, yayınlanmamış yüksek lisans tezi.
- Bloom, B. S. (1989). *İnsan Nitelikleri ve Okulda Öğrenme* (Çev. D. Ali Özçelik), Ankara: Milli Eğitim Basımevi.
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., Vol. 4, pp. 221-256). Westport: American Council on Education & Praeger Publishers.
- Camili, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. London: Sage publications.
- Chang, H., Mazzeo, J., & Roussos, L. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *Journal of Educational Measurement*, 33(3), 333-353.
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differential item functioningtest items. *Educational Measurement: Issues and Practice*, 17, 31-44.

- Dancer, L. S. , Anderson, A. J. , Derlin, R. L. (1994). Use of log-linear models for assesing differential item functioning in a measure of psychological functioning. *Journal of consulting and clinical psychology*, 62(4), 710-717.
- Dorans, N. J., & Schmitt, A. P. (1993). Constructed response and differential item functioning: A pragmatic approach. (ETS-RR-91-47). Princeton, NJ: Educational Testing Service.
- Jodion, M. G. & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure ith the logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14(4), 329-349.
- Johanson, G. A., & Dodeen, H. (2003). An Analysis of Sex-related Differential Item Functioning in Attitude Assessment. *Assesment &Evaluation in Higher Education*, 28(2).
- Fang, T. (1999). *Detecting DIF in polytomous item responses*. Unpublished doctor dissertation, University of Ottawa, Canada.
- Fang, T. (1999). *Detecting DIF in polytomous item responses*. Unpublished doctoral dissertation, University of Ottawa, Canada.
- Feng, Y. (2008). *Difference in gender differential item functioning patterns across item format and subject area on diploma examinations after change in administration procedure*. Unpublished doctoral dissertation, University of Alberta, Edmonton, Alberta, Canada.
- French, A. W., & Miller, T. R. (1996). Logistic regression and its use in detecting differential item functioning in polytomous items. *Journal of Educational Measurement*, 33(3), 315-332.
- Gelin, M. N., & Zumbo, B. D. (2003). DIF results may change depending on how an item is scored: An illustration with the Center for Epidemiological Studies Depression (CES-D) scale. *Educational and Psychological Measurement*, 63, 65-74.
- Henderson, D. L. (1999). *Investigation of differential item functioning in exit Examinations across item format and subject area*. Unpublished doctor dissertation, University of Alberta, Edmonton, Alberta, Canada.
- Kristjansson, E., Aylesworth, R., McDowell, I., & Zumbo, B. D. (2005). A comparison of four methods for detecting DIF in ordered response items. *Educational and Psychological Measurement*, 65, 935-953.
- Le, L.T. (2009). Investigation gender differential item functioning across countries abd test languages for PISA science items. *International Journal of Testing*, 9(2), 122-133.
- Mellor, T. L. (1995). *A comparison of four differantial item functioning methods for polytomously scored items*. Unpublished doctoral dissertation. The university of Texas Austin.
- Miller, T. R., & Spray, J. A. (1993). Logistic discriminant function analysis for DIF identification of polytomously scored items. *Journal of Educational Measurement*, 30(2), 107-122.
- Özdemir, F. (2009). *PISA 2003'de genel lise öğrencileri ve Kanuni Lisesi öğrencilerinin matematik başarısını etkileyen faktörlerin incelenmesi*. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, yayınlanmamış yüksek lisans tezi.

- Smith, F. W. (2009). *Language-related differential item functioning in the WASL mathematics test*. Unpublished doctor dissertation. The university of Washington.
- Sireci, G. S. & Khaliq, K. N. (2002). *An analysis of the psychometric properties of dual language test forms*. Annual Meeting of the National Council on Measurement in Education, New Orleans, LA, April 2-4.
- Shealy, R. T., & Stout, W. F. (1993). An item response theory model for test bias and differential test functioning. In P. W. Holland & H. Wainer (Eds.), *Differential itemfunctioning* (pp.197-239). Hillsdale, NJ: Erlbaum.
- Swaminathan, H. & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361- 370.
- Yıldırım, H.H. ve Berberoğlu, G. (2006). Judgementel and statistical analyses of the PISA 2003 mathematics literacy items. *International Journal of Testing*, 9(2),108-121.
- Welch, C., & Hoover, H. D. (1993). Procedures for extending item bias techniques to polytomously scored items. *Applied Measurement in Education*, 6, 1-19.
- 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.
- Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, Where it is now, and Where It Is Going. *Language Assessment Quarterly*, 4(2), 223–233.
- Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessing differential item functioning for performance tests. *Journal of Educational Measurement*, 30(3), 233-251.
- Zwick, R., Thayer, D. T., & Mazzeo, J. (1997).Descriptive and inferential procedures for assesing Differential Item Functioning in Polytomous Items. *Applied Measurement in Education*,10.