

Değişen Madde Fonksiyonunu Belirlemede Mantel - Haenszel, Ki-Kare ve Lojistik Regresyon Tekniklerinin Karşılaştırılması*

The Comparison of Mantel – Haenszel, Chi-Square and Logistic Regression Techniques For Identifying Differential Item Functioning

Nuri DOĞAN **
Hacettepe Üniversitesi

Tuncay ÖĞRETMEN ***
Ege Üniversitesi

Oz

Bu çalışmanın amacı, değişen madde fonksiyonunu (DMF) belirleme tekniklerinden ki-kare, Mantel-Haenszel ve lojistik regresyon tekniklerini karşılaştırarak uygulamada ortaya çıkan benzerlik veya farklılıklarını belirlemektir. Çalışma, 2003 yılında Ortaöğretim Kurumları Seçme ve Yerleştirme Sınavı'na (OKÖSYS) katılan yaklaşık 600.000 öğrenci arasından yansız olarak seçilen 3345 öğrenciden oluşan bir örneklem üzerinde yürütülmüştür. Çalışmanın verileri, OKÖSYS fen bilgisi alt testine öğrencilerin verdiği yanıtlardan oluşmaktadır. Değişen madde fonksiyonu analizi sadece cinsiyet grupları üzerinde yürütülmüştür. Araştırmanın sonuçları, söz konusu tekniklerin bazı açılardan birbirlerine göre benzerlik ve farklılıklar sağladığını göstermiştir.

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

Abstract

The purpose of this study was to determine to similarity and differences in practice by comparing Chi-square, the Mantel-Haenszel and logistic regression techniques, and differential item functioning (DIF) determination techniques. The study was carried out with a sample of 3345 students selected from approximately 600.000 students who Participated Selection and Placement Examination for Secondary Education Institutions in 2003. The data of study was composed of students' answers to science sub-test in Selection and Placement Examination for Secondary Education. The DIF analysis was carried out with gender groups only. The results of the study indicated that these techniques provided many similarities and differences.

Keywords: Differential item functioning; Chi-square; Mantel-Haenszel; logistic regression.

Summary

Purpose

In this study, Differential Item Functioning (DIF) values obtained through Mantel-Haenszel (MH), Chi-Square and logistic regression (LR) were compared according to gender (males were the reference group while females were the focal group) to test whether these procedures yielded similar results. The study intended to make unique contribution to the relevant literature by comparing Chi-Square procedures with LR; and Chi-Square procedures with MH. These comparisons can provide evidence for the fact that these techniques can be used interchangeably. Thus, the study sought answers to the following questions:

* 15. Ulusal Eğitim Bilimleri Kongresinde Değişen Madde Fonksiyonunun Belirlemede Kullanılan Tekniklerin Karşılaştırılması Başlığı ile Sunulmuştur.

** Dr. Nuri DOĞAN, Hacettepe Üniversitesi, Eğitim Fakültesi, nurid@hacettepe.edu.tr

*** Yrd. Doç. Dr. Tuncay ÖĞRETMEN, Ege Üniversitesi, Eğitim Fakültesi, togretmen@gmail.com

1. Do items yielding to DIF values and their numbers vary significantly according to MH, Chi-Square and logistic regression procedures for each gender group?
2. What are relationships between DIF values of gender groups obtained through MH, Chi-Square and logistic regression procedures?

Method

Participants of the study were 3344 students randomly selected among 60000 students who took The Elementary School Student Selection and Placement Examination (ESSSPE) held in 2003 by the Turkish Ministry of National Education. SPSS 12, MH, EZDIF and Microsoft Excel 2003 software programs were used in data analyses. There were two steps to data analysis. First, descriptive statistics were determined. Then, using the respective software programs, the DIF values on the science subscale of the ESSSPE and correlations among these values were calculated.

Results

The MH resulted in a total of 15 items with DIF values. The remaining 10 items had DIF at A (neglegable) level. Of the 15 items, 8 favored males and 7 favored females. Non-signed Chi-Square resulted in 11 items with DIF values while signed Chi-Square yielded to no items with DIF values. Logistic regression resulted in 9 items with DIF values, of which only the 9th item had non-uniform DIF while the rest of the items had uniform DIF. However, effect sizes show that none of the levels of DIF were significant.

Spearman correlation coefficients between all the Chi-Square values obtained through logistic regression, MH, Non-signed Chi-Square and Signed Chi-Square procedures were significant at $\alpha = 0,01$. The correlations ranged between 0,934 (between LR and MH) and 0,789 (between MH and Non-signed Chi-Square).

Conclusion

The results of MH, LR and Chi-Square procedures showed that the number of items yielding to DIF values ranged between 0 and 15. In other words, the number of items varied significantly according to the procedure used. The biggest number of items with DIF values was obtained through MH which was followed by non-signed Chi-Square (9 items). On the other hand, signed Chi-Square and LR resulted with no items with DIF values. Thus, LR and signed Chi-Square techniques were similar in terms of the resulting items with DIF values. These procedures can be compared by the size of their respective Chi-square values. The fact that their Spearman correlation coefficients were significant could be taken as evidence for the parallel between the two procedures.

Based on the results, one can attempt to answer which procedures are more advantageous than the others, which procedure results in superior outcomes depending on the frame of reference of the viewer. For instance, by examining the correlations, one can see that all these procedures are Chi-square based and do not differ significantly. On the other hand, considering the items with DIF values and their number, logistic regression seems to provide results that are more sensitive and consistent with findings of prior research. Furthermore, given the number of common items with DIF values they generated, perhaps MH and Non-signed Chi-Square can be used interchangeably.

In sum, MH, Chi-Square and logistic regression procedures were similar with respect to the size of their respective Chi-square values and were different in terms of the number of items with DIF values they generated.

Giriş

Bir testin psikometrik bakımdan en önemli özelliği olan geçerlik düzeyini saptamak amacıyla kullanılabilecek birçok teknik bulunmaktadır. Bu tekniklerden biri de test ve maddelere ilişkin değişen madde fonksiyonlarının - DMF - (differential item functioning – DIF) belirlen-

mesidir. Test ve testteki maddeler için DMF belirleme çalışmaları son yıllarda yaygınlaşmakta, bu çalışmalarda elde edilen bilgilerden madde yanlılığını belirlemede yararlanılmaktadır.

DMF, birçok kaynakta benzer şekilde tanımlanmıştır. Bu tanımlara dayanarak DMF, 'aynı yetenek düzeyinde olan, fakat cinsiyet, sosyoekonomik düzey, etnik köken, inanç vb. gibi farklı gruplardan gelen bireylerin, test maddelerine doğru cevap verme olasılıklarının değişmesi' biçiminde sentezlenebilir (Rodney & Drasgow, 1990; Raju, 1990; Mellenberg, 1989; Tittle, 1988; Adams & Rowe, 1988; Shepard, Camilli, & Williams, 1985; Mellenberg, 1983; Osterlind, 1983; Devine & Raju, 1982). Aynı yetenek düzeyine sahip bireylerin bir maddeye doğru cevap verme olasılığının değişmesinin iki temel kaynağı vardır: madde yanlılığı veya gerçek bilgi, beceri vb. farklılığı. Maddelerin DMF verip vermediğini belirleme, yanlılık için daha objektif bir yaklaşım olarak görüldüğünden daha sık kullanılan bir tekniktir. DMF veren bir maddenin DMF kaynağının ne olduğunu belirlemek için içerik analizi veya uzman kanısına başvurma yollarından yararlanılabilir.

DMF'de kendi içinde tek biçimli (uniform) ve tek biçimli olmayan (nonuniform) şeklinde ikiye ayrılmaktadır. Bir maddenin doğru cevaplanma olasılığı, tüm yetenek düzeyleri için bir grubun lehine işlediğinde tek biçimli, farklı yetenek düzeylerinde farklı gruplar lehine işlediğinde tek biçimli olmayan DMF'den söz edilmektedir (Zumbo, 1999). Tekniklerin hemen hepsi tek biçimli DMF'yi belirleyebilmesine rağmen, tek biçimli olmayan DMF'yi belirlemek her teknik için olanaklı değildir.

DMF belirleme teknikleri, klasik test kuramına, örtük özellikler (madde tepki) kuramına veya bazı istatistiksel tekniklere dayanarak açıklanabilir. Klasik test kuramı kapsamında madde ayıricılık gücü, madde güçlüğü, faktör analizi, varyans analizi, madde güçlük dönüşümü (MGD) vb.; örtük özellikler kuramı kapsamında işaretli ve işaretsiz alan indeksleri, Lord'un Ki-kare'si, madde parametreleri ya da en çok olabilirlik oranları farklarının karşılaştırılması vb.; bazı istatistiksel teknikler dayananlar arasında ise Ki-kare, Mantel-Haenszel (MH) test istatistiği ve lojistik regresyon (LR) gibi teknikler sıralanabilir. (Zumbo, 1999; Hambleton, Swaminathan, & Rogers, 1991; Rodney & Drasgow, 1990; Raju, 1990; Mellenberg, 1989; Adams & Rowe, 1988; Seong & Subkoviak, 1987; Holland & Thayer, 1986; Shepard, Camilli, & Williams, 1985; Hills, 1984; Devine & Raju, 1982; Rudner, Getson & Knight, 1980). Bu çalışmada, DMF sonuçlarının karşılaştırmasını yapmak amacıyla Mantel – Haenszel (MH), Ki-kare ve lojistik regresyon (LR) teknikleri ele alınmıştır.

Mantel – Haenszel Tekniği (MH)

Bir Ki-kare tekniği olan MH yaklaşımında, herhangi bir maddeye verilen cevaplar iki grup için Tablo 1'deki gibi gösterilebilir. MH Ki-kare istatistiği denklem (1) ve MH istatistiğinin doğal logaritması alınarak türetilmiş delta değeri denklem (2) kullanılarak hesaplanabilir.

Denklem (1)'den elde edilen sonuçlar, $MH > 1$ ise referans (reference) grup lehine DMF'den; $MH < 1$ ise odak (focal) grup lehine DMF'den ve $MH \approx 1$ ise DMF'nin yokluğundan söz edilir. MH istatistiğinin daha kolay yorumlanabilmesi için logaritmik dönüşüm yapılmaktadır. Logaritmik dönüşüm formülü denklem (2)'de görülmektedir. Denklem (2)'ye göre elde edilen sonuçlar, $\Delta_{MH} > 0$ ise DMF odak grup lehine, $\Delta_{MH} < 0$ ise DMF referans grup lehine ve $\Delta_{MH} \approx 0$ ise DMF yoktur şeklinde yorumlanmaktadır (Holland & Thayer, 1986). Ayrıca, MH büyüklüğüne göre DMF'nin düzeyi hakkında da yorum yapılabilmektedir. Eğer $|\Delta_{MH}| < 1$ ise A (önemsenmeyen) düzeyde; $1 \leq |\Delta_{MH}| < 1,5$ ise B (orta) düzeyde ve $|\Delta_{MH}| \geq 1,5$ ise C (yüksek) düzeyde DMF olduğundan söz edilmektedir (Dorans & Holland, 1993). Bu tekniğin en zayıf yönlerinden birisi tek biçimli veya tek biçimli olmayan DMF'yi ayıramamasıdır.

Ki-Kare Tekniği

Scheuneman (1979) ve Camilli (1979) tarafından geliştirilmiştir. Uygulaması kolay bir tekniktir. Bu teknikte, gözlenen puanlar birkaç kategoriye ayrılır. Gruplar her kategori için, maddeyi doğru cevaplama oranları açısından karşılaştırılır. Grupların kategorilerde verdiği cevaplar oranından Ki-kare değeri elde edilir. Hesaplanan Ki-kare değeri, ilgili serbestlik derecesi ve alfa düzeyindeki Ki-kare dağılımındaki tablo değeri ile karşılaştırarak manidarlık testi yapılmaktadır.

Tablo 1.
Madde Cevaplarının İki Grup İçin Dağılımı

Gruplar	Doğru (1)	Yanlış (0)	Toplam
Grup 1	A	B	A+B
Grup 2	C	D	C+D
Toplam	A+C	B+D	T

$$(1) \quad MH = \frac{AD}{BC}$$

$$(2) \Delta\Omega = \Delta MH = -(4/1,7) \ln MH = -2,35 * \log it$$

Ki-kare ile DMF belirleme, işaretli ve işaretli olmak üzere iki teknik altında incelenebilmektedir. Denklem (3) işaretli Ki-kare tekniğinin formülünü göstermektedir. Bu istatistiği J serbestlik derecesinde, (1-α) düzeyinde Ki-kare dağılımı ile karşılaştırarak manidarlık testi yapılmaktadır. Bu formülde J toplam aralık (kategori) sayısı, P_{1j} j puan aralığında birinci grupta bulunan bireylerin maddeyi doğru cevaplama oranıdır. İkinci grup içinde aynı yöntemle P_{2j} hesaplanmaktadır. Gruplar için P._j değerleri P_{1j} = D_{1j}/ N_{1j} formülüyle hesaplanmaktadır. Bir kategoride tüm grubun maddeyi doğru cevaplama oranı olan P._j ise, P._j = (D_{1j}+D_{2j}) / (N_{1j}+N_{2j}) formülü ile bulunur. D_{1j} ve D_{2j} sembolleri j inci kategoride maddeye doğru cevap veren cevaplayıcı sayısının alt gruplara göre gösterimidir.

$$(3) \quad \chi_c^2 = \sum_{j=1}^j \frac{N_{1j} N_{2j} (P_{1j} - P_{2j})^2}{(N_{1j} + N_{2j}) P_j (1 - P_j)} = \sum \chi_j^2$$

$$(4) \quad \chi_s^2 = \sum_{j=1}^j \frac{(D_{1j} - P_j N_{1j})^2}{P_j N_{1j}} + \sum_{j=1}^j \frac{(D_{2j} - P_j N_{2j})^2}{P_j N_{2j}}$$

İşaretli Ki-kare istatistiği, (D_{1j}-P._jN_{1j})² ve (D_{2j}-P._jN_{2j})² hesaplanarak pozitif ya da negatif işaretler D_{1j}'nin P._jN_{1j}'den büyük ya da küçük oluşuna göre elde edilmektedir. Bu istatistiğe Scheuneman istatistiği adı verilmektedir ve denklem 4'teki formülle hesaplanmaktadır. Bu tekniğin avantajı işaretine bakılarak DMF veren maddenin hangi grubun lehine işlediğini belirlemesidir. Scheuneman buradaki dağılımın (J-1) serbestlik dereceli χ² dağılımı olduğunu ileri sürmektedir.

Ki-kare yaklaşımı kolay olmasına rağmen, yetenek düzeylerinin (kategori sayısı) seçimi ve gruplar için madde ayrıncılıklarına duyarlı olmanın dezavantajlarını beraberinde getirmesi nedeniyle eleştirilmektedir. Kesme puanlarının seçimi, elde edilecek χ² değerlerinin büyüklü-

ğünü etkilemektedir. Diğer taraftan, her grup ve aralık için, doğru cevap sayısının en az 5 olması gerektiği vurgulanmaktadır (Crocker ve Algina, 1986, pp. 384-386; Ironson ve Craig, 1982).

Lojistik Regresyon Tekniği

Lojistik regresyon genellikle bir test veya alt test puanı ölçüt alınarak, aynı yetenek düzeyindeki farklı grup üyelerinin bir maddeye doğru cevap verme olasılıklarının istatistiksel olarak modellenmesine dayanır. Lojistik regresyon tekniği 1 ve 0 biçiminde puanlanan madde cevaplarını bağımlı değişken; grup değişkeni, bireylerin ölçek puanı ve grupla ölçek puanı etkileşimini bağımsız değişken olarak kullanır. Bu teknik kurulan modeller sayesinde maddeye verilen cevaplar ve toplam puan üzerinden DMF'nin sınırlanmasını sağlar. Regresyon denkleminde grupların etkisinin manidarlığı tek biçimli DMF'yi, grup ve toplam puan etkileşiminin manidarlığı tek biçimli olmayan DMF'yi gösterir. Lojistik regresyonun genel eşitliği denklem 5'teki gibi gösterilebilir. Elde edilen değerler doğal logaritması alınarak odss oranı da denklem 6'da verilen formülle elde edilmektedir. Bu eşitlikte P_i , maddeye doğru cevap verenlerin oranı ve bir maddenin doğru cevaplanma olasılığıdır.

Lojistik regresyon kullanmanın MH gibi diğer DMF belirleme tekniklerine göre üç üstünlüğü olduğu iddia edilmektedir. Bunlar, sürekli bir değişken olan ölçüt değişkenin (test puanlarının) kategorileştirilmesini gerektirmemesi, tek biçimli ve tek biçimli olmayan DMF'yi modelleyebilmesi, ikili puanlama yanında sıralama ölçeğindeki veriler için de kullanılabilmesi olarak belirtilebilir. Bununla beraber lojistik regresyonda grup etkisi için hesaplanan Ki-kare değerlerinin MH tekniğinde hesaplanan Ki-kare değerlerine çok yakın olduğu da söylenmektedir (Zumbo, 1997; Camilli & Shepard 1994). Lojistik regresyonla elde edilen DMF'nin önemliliğini belirlemek, modellemenin yapısıyla ilişkilidir. Lojistik regresyonla DMF'nin modellenmesi modele, alınan değişkenlerin doğal bir hiyerarşisine dayanır. Bu hiyerarşi aşağıdaki gibi gösterilebilir (Zumbo, 1997).

1. Öncelikle toplam puan modele alınır.
2. İkinci olarak grup değişkeni modele alınır.
3. Son olarak etkileşim terimi modele alınır.

Tablo 2.

Lojistik Regresyon Tekniğinde DMF Düzeyi İçin R² Ölçütleri

DMF Düzeyi	Gierl ve Arkadaşlarının Ölçütleri	Zumbo ve Thomas'ın Ölçütleri	Yorumlar
A	$R^2 < 0,035$	$R^2 < 0,13$	DMF yoktur.
B	$0,035 \leq R^2 < 0,070$	$0,13 \leq R^2 < 0,26$	Orta düzeyde DMF vardır.
C	$R^2 \geq 0,070$	$R^2 \geq 0,26$	Yüksek düzeyde DMF vardır.

Lojistik regresyonda bu üç aşama için Ki-kare değerleri hesaplanır. Hesaplanan bu Ki-karelerin manidarlık düzeyi bize DMF ve önemi hakkında bilgi verebilir. Tek biçimli DMF sınaması ikinci ve birinci aşamalar arası farkı gösterir. Tek biçimli olmayan DMF'nin varlığına ilişkin kanıt, üçüncü adım için elde edilen Ki - kare değeriyle birinci adımdaki Ki-kare değerinin farkını alarak elde edilebilir. Ortaya çıkan bu yeni Ki-kare değeri 2 serbestlik dereceli Ki-kare dağılım fonksiyonu (tablo değeri) ile karşılaştırılır. İki serbestlik derecesi birinci adımdaki modelin 1 serbestlik derecesi ve 3. adımdaki modelin 3 serbestlik derecesinin farkının bir sonucu olarak ortaya çıkar. İki serbestlik dereceli Ki - kare testi sonuçları tek biçimli ve tek biçimli olmayan DMF'nin eşzamanlı olarak sınırlanmasını sağlar (Waller 1998; Zumbo, 1997; Swaminathan ve Rogers 1990). Zumbo (1997) tarafından sözü edilen ardışık modelleme stratejisi, tek biçimli DMF için toplam puanlar (ölçüt değişken) üzerinde grup farklarını gösteren tek değişkenlik ölçüsü için 2. aşamadaki R² değeri ve 1. aşamadaki R² değerini karşılaştırmayı sağlar. Ayrıca 2. ve 3. adımdaki R² değerini karşılaştırmak, tek biçimli olmayan DMF'nin miktarını bulmak amacıyla etkileşim için kullanılacak değişkenlik ölçüsüdür.

Bununla beraber DMF'nin önemliliği için R2 değerleri kullanılması önerilmektedir. Zumbo ve Thomas (1997, 1998) göstermiştir ki DMF belirlemek için etki büyüklüğü ölçüsü ve lojistik regresyondaki 2 serbestlik dereceli Ki-kare (olabilirlik oran istatistiği) istatistiklerinin sınanmasına gereksinim vardır. Bir maddeyi DMF gösteriyor olarak sınıflamak için, lojistik regresyondaki 2 serbestlik dereceli Ki-kare istatistiğinin 0,01 den küçük p olasılığına ve Zumbo – Thomas (1997) etki büyüklüğü en az 0,13 değerine sahip olmalıdır. Büyük örneklemelerde etki büyüklüğüne bakmaksızın DMF çalışıldığı zaman, önemsiz bir etki istatistiksel olarak önemli bulunabilir. Bu nedenle R2 sınaması üzerinde önemle durulmaktadır. Çünkü R2 değeri DMF'nin derecesini vermektedir. Hangi düzeydeki R2 değerlerinin önemli sayılacağına ilişkin çeşitli ölçütler geliştirilmiştir. Gierl, Khaliq ve Boughton (1999) ile Zumbo ve Thomas (1997) tarafından önerilen ölçütler Tablo 2'de verilmiştir. Gierl ve arkadaşları tarafından önerilen ölçütler Zumbo (1999:27) tarafından daha tutucu değerler olarak kabul edilmekle beraber kullanılabilmesi yönünde değerlendirilmiştir. Bu çalışmada da Gierl ve arkadaşları tarafından önerilen değerler ölçüt alınmıştır.

Problem Durumu

Türkiye'de DMF çalışmalarının son yıllarda yapılmaya başlandığı söylenebilir. DMF çalışmalarında genellikle klasik test kuramına dayanan teknikler, lojistik regresyon, Ki-kare MH veya örtük özellikler kuramına dayanan teknikler kullanılarak DMF veren maddeler belirlenmeye çalışılmaktadır (Doğan & Öğretmen 2006; Öğretmen & Doğan, 2004; Yurdugül, 2003; Özdemir, 2003; Öğretmen, 1995; Yenal, 1995). Ancak, örneklem büyüklüğü, verilerin yapısı, maddelerin puanlanış biçimi, DMF belirlemede kullanılan teknikler vb. maddelerin DMF düzeylerinin farklı araştırmalarda farklı şekilde belirlenmesine neden olabilmektedir. Bu nedenle de hangi tekniğin kullanılması gerektiğine karar verebilmek için DMF belirleme tekniklerinin karşılaştırılmasına ihtiyaç duyulmaktadır. Uluslararası alanda bu çalışmalara rastlanmakla birlikte, genellikle burada verilen tekniklerin sadece ikisinin karşılaştırılması ile yetinildiği; üçünün birden karşılaştırılmadığı söylenebilir. DMF belirleme tekniklerini karşılaştırma çalışmalarına, Türkiye'deki nüfusun büyük kısmını ilgilendiren sınav sonuçları kullanılarak katkı sağlamak, hem bilimsel birikim hem de ülkemizdeki ölçme araçlarının geçerliğinin belirlenmesi bakımlarından önemlidir.

Bu araştırmada, MH, Ki-kare ve lojistik regresyon kullanılarak elde edilen DMF sonuçları cinsiyet gruplarına göre karşılaştırılarak, tekniklerin benzer sonuç verip vermediklerine bakılmıştır. Ki-kare teknikleri ile LR ve Ki-kare teknikleri ile MH arasında yapılacak karşılaştırmaların alanyazına sağlayacağı katkılar araştırmanın özgün yanını oluşturmaktadır. Tekniklerin birbiriyle karşılaştırılmasıyla elde edilen sonuçlardan hangi tekniklerin birbiri yerine kullanılabilmesi, hangi tekniklerin benzer sonuçlar verdiği vb. durumlara ilişkin kanıtlar ortaya çıkarılabilir. Bu amaçla aşağıdaki sorulara cevap aranmıştır.

1. Cinsiyet grupları için, MH, Ki-kare ve lojistik regresyon tekniklerine göre DMF veren maddeler ve sayısı değişmekte midir?
2. Cinsiyet gruplarına göre, MH, Ki-kare ve lojistik regresyon tekniklerine göre elde edilen madde DMF değerleri arasındaki ilişkiler nasıldır?

Yöntem

Araştırma, DMF belirleme tekniklerinden MH, Ki-kare ve lojistik regresyon tekniklerini karşılaştırma amacı taşıdığından dar da olsa kuramsal özellik taşımaktadır. Aynı zamanda veri toplama aracındaki maddelerin DMF düzeylerine ilişkin bilgi verdiği için uygulama alanına ilişkin çıkarımlar da elde edilebilir. Bu nedenle araştırmanın hem kuramsal hem de uygulamaya dönük özellikler taşıdığı söylenebilir.

Örneklem

Araştırma MEB 2003 Ortaöğretim Kurumları Seçme ve Yerleştirme Sınavı'na katılan yaklaşık 600.000 öğrenci arasından yansız olarak seçilmiş 3344 kişi üzerinden yürütülmüştür. Örneklem seçilirken tesadüfi örnekleme tekniği kullanılmıştır. Örneklem seçiminde SPSS paket programından yararlanılmıştır. Ancak evrenden örneklem seçilmeden önce evrende sapan değer (outlire) analizi yapılmıştır. Bireylerin fen puanları z standart puanına dönüştürülmüş ve testten sıfır puan alanlar ile -3 ile +3 standart puanın dışında kalan bireyler evrenden çıkarılmıştır. Örneklem ve alt gruplara ilişki betimsel istatistikler Tablo 3'te görünmektedir.

Tablo 3.

OKS 2003 Örneklemine Ait Betimsel İstatistikler

İstatistikler	Kız+Erkek	Kız	Erkek
N	3344	1660	1684
Ortalama	11,02	10,87	11,13
Medyan	10,00	10,0	10,00
Mod	8,00	8,00	10,00
Std. sapma	5,44	5,22	5,65
Çarpıklık	0,43	0,53	0,35
Basıklık	-0,63	-0,50	-0,75
Güvenirlilik (KR-20)	0,85	0,83	0,86

Tablo 3'e göre örneklem alt gruplara ilişkin hesaplanan KR-20 güvenirliklerinin yeterli olduğu söylenebilir. Ayrıca alt grup dağılımının birbirine ve örneklem dağılımına benzediği betimsel istatistiklerden anlaşılmaktadır. Örneklemin ve alt grupların betimsel istatistikleri incelendiğinde, çok küçük farklarla birbirine yaklaştığını söylemek olanaklı görünmektedir. Betimsel istatistikler alt gruplar için DMF analizi yapılmasını engelleyen önemli bir fark bulunmadığına ilişkin bir fikir verebilmektedir.

Araştırma Verileri

Veriler, MEB 2003 Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı'ndan (OKÖSYS) alınmıştır. OKÖSYS'nin alt testlerinden Fen Bilgisi alt testine verilen cevaplara ilişkin puanlar üzerinde çalışılmıştır. Fen Bilgisi alt testi 25 sorudan oluşmaktadır. Sorular çoktan seçmeli ve dört seçeneklidir. Soruların puanlanması doğru cevaba 1, doğru olmayan cevaplara 0 puan verilmesi şeklindedir. Cevapsız sorular için herhangi bir işlem yapılmamaktadır.

Verilerin Analizi

Verilerin analiz edilmesi sırasında SPSS 12, MH (Dorans & Kulick, 1983), EZDIF (Waller, 1998) ve Microsoft Excel 2003 programlarından faydalanılmıştır.

Verilerin analizi iki aşamada gerçekleştirilmiştir. Öncelikle betimsel istatistikler elde edilmiştir (Tablo 3). İkinci aşamada, uygun paket programlar kullanılarak araştırma kapsamındaki tekniklere göre, Fen Bilgisi alt testine ait maddelerin DMF değerleri hesaplanması; elde edilen DMF değerleri arasındaki korelasyonların bulunması gerçekleştirilmiştir. Korelasyon hesaplamalarında, sıra farkları tekniğinden yararlanılmıştır. Korelasyonlar her bir teknikte DMF ölçüsü olarak hesaplanan Ki-kare değerlerinin büyüklük sırası arasındaki ilişki düzeyini belirlemek amacıyla kullanılmıştır. Analizler sırasında erkekler referans (reference), kızlar ise odak (focal) grup olarak ele alınmıştır.

Bulgular ve Yorum

Farklı Tekniklere Göre Yanlılık Gösteren Maddeler

Maddelerin güçlük düzeyleri Tablo 4'te 2. ve 3. sütunlarda verilmiştir. Bu değerlerle ilgili olarak vurgulanması gereken nokta, maddelerin kız ve erkek gruplarındaki güçlük düzeyleri arasındaki fark büyüdükçe DMF verme olasılıklarının yükselmesidir. Tablo 4 incelendiğinde, güçlük düzeyleri arasındaki farklar büyüdükçe hesaplanan Ki-kare değerlerinin ve bu değerlerin manidar olma olasılıklarının da yükseldiği görülebilir.

Tablo 4'te 8. ve 9. sütunlarda yer alan MH tekniği sonuçlarına göre 4, 5, 9, 10, 12, 13, 15, 16, 19, 20, 22, 23 ve 25. maddeler C düzeyinde; 2 ve 3. maddeler B düzeyinde DMF göstermektedir. Toplam 15 madde DMF gösterirken geriye kalan 10 madde A (önemsenmeyecek) düzeyde DMF göstermektedir. DMF veren maddelerden 2, 3, 4, 5, 9, 10, 12 ve 15. maddeler erkekler; 13, 16, 19, 20, 22, 23 ve 25. maddeler kızlar lehine işlemiştir. DMF veren maddelerden 8'i erkekler lehine işlerken 7'si kızlar lehine işlemektedir.

Maddelere ilişkin olarak elde edilen işaretli ve işaretsiz Ki-kare yanlılık değerleri Tablo 4'te son iki sütunda verilmiştir. İşaretsiz Ki-kare istatistiği için 5 serbestlik derecesinde tablo değeri alfa 0,01 alındığında 15,0863; işaretli Ki-kare istatistiği için 9 serbestlik derecesinde tablo değeri alfa 0,01 alındığında 21,666 olarak tespit edilmiştir. Bir madde için hesaplanan Ki-kare istatistiği tablo değerlerinden büyükse madde DMF' ye sahip şeklinde yorumlanabilir. Buna göre, Tablo 4'te işaretsiz ve işaretli Ki-kare değerleri incelendiğinde, işaretsiz Ki-kare sonuçları 11 maddenin DMF gösterdiği, işaretli Ki-kare sonuçları ise hiçbir maddenin DMF göstermediği şeklinde açıklanabilir.

Lojistik regresyon tekniğine göre elde edilen sonuçlar Tablo 4'te 4. ile 7. sütun aralığında verilmiştir. Tablo 4'teki 4. sütun lojistik regresyon Ki-kare değerlerini, 5. sütun ise bu değerlerin olasılıklarını göstermektedir. Ki-kare değerlerinin manidarlığı için 0,01 değerinden küçük olasılıkları dikkate alındığında 9 maddenin DMF verdiğini söylemek mümkündür. 6. sütundaki etkileşim terimlerinin olasılığı incelendiğinde, sadece 9. maddenin tek biçimli olmayan DMF, diğer 8 maddenin ise tek biçimli DMF verdiği söylenebilir. Ancak 7. sütundaki ΔR^2 etki büyüklüğü değerleri incelendiğinde, hiçbir maddenin DMF düzeyinin önemli sayılamayacağı, ΔR değerlerinin çok düşük olduğu görülmektedir. Bu sonuçlara dayanarak lojistik regresyon sonuçlarına göre hiçbir madde DMF vermemektedir yorumuna ulaşılabilir.

Tablo 4 incelendiğinde, üç tekniğe göre ayrı ayrı belirlenen DMF veren maddelerin ve sayısının değiştiği görülmektedir.

Tablo 4.

Madde Güçlükleri ve Farklı Tekniklerle Elde Edilen DMF Değerleri

No	Pe	Pk	$\Delta \chi^2$	P(G)	P(TG)	ΔR^2	Δ_{MH}	MH	χ^2_c	χ^2_{si}
1	0,55	0,52	4,38	0,34	0,081	0,002	-0,24	2,25-A	5,13	1,49
2	0,40	0,35	9,03	0,01	0,033	0,003	-0,38	4,75-B	8,45	3,07
3	0,31	0,26	5,62	0,06	0,674	0,002	-0,50	6,07-B	5,73	4,04
4	0,41	0,36	7,34	0,03	0,717	0,002	-0,45	7,86-C	10,80	6,32
5	0,30	0,26	6,31	0,43	0,835	0,002	-0,51	7,66-C	11,05	7,13
6	0,24	0,22	0,23	0,89	0,913	0,000	-0,21	0,86-A	4,04	3,19
7	0,17	0,17	0,50	0,78	0,568	0,000	0,00	0,00-A	6,86	4,20
8	0,34	0,33	1,78	0,41	0,194	0,000	0,01	0,00-A	7,45	3,42
9	0,39	0,34	19,11	0,00	0,001	0,006	-0,55	8,28-C	16,33	6,34
10	0,60	0,51	33,84	0,00	0,540	0,009	-1,06	32,82-C	31,48	15,09
11	0,53	0,51	1,88	0,39	0,175	0,000	-0,03	0,01-A	5,92	0,69
12	0,48	0,39	24,84	0,00	0,622	0,007	-0,95	23,94-C	22,04	12,19
13	0,39	0,42	23,36	0,00	0,029	0,006	0,78	13,82-C	28,40	9,68
14	0,44	0,44	2,41	0,30	0,418	0,001	0,23	1,56-A	5,63	1,90
15	0,55	0,48	13,93	0,00	0,284	0,003	-0,67	11,63-C	15,57	5,64
16	0,44	0,47	10,28	0,01	0,867	0,004	0,55	9,04-C	15,26	9,08
17	0,16	0,16	1,16	0,56	0,287	0,000	0,02	0,00-A	4,40	3,55
18	0,38	0,37	4,70	0,10	0,041	0,001	0,08	0,17-A	18,67	4,40
19	0,71	0,75	10,53	0,01	0,493	0,003	0,65	11,81-C	19,88	5,69
20	0,44	0,49	14,69	0,00	0,389	0,005	0,61	14,30-C	14,55	8,22
21	0,46	0,44	0,31	0,86	0,760	0,000	-0,13	0,62-A	2,30	1,30
22	0,74	0,78	12,95	0,00	0,110	0,004	0,76	13,16-C	15,98	4,36
23	0,60	0,67	28,53	0,00	0,580	0,009	0,98	30,30-C	29,67	12,39
24	0,52	0,54	1,78	0,41	0,913	0,001	0,22	2,20-A	4,23	2,38
25	0,60	0,66	31,86	0,00	0,014	0,009	0,97	26,33-C	30,09	10,24

Pe: Erkekler için madde güçlük değerleri

 $\Delta \chi^2$: Lojistik regresyon Ki-kare değerleri

Pk: Kızlar için madde güçlük değerleri

P(G): LR Ki-kare değerleri olasılığı

 ΔR^2 : LR için etki büyüklüğü değerleri

MH: MH odds oranı

 Δ_{MH} : Dönüştürülmüş MH değerleri χ^2_c : İşaretsiz Ki-kare istatistiği χ^2_{si} : İşaretili Ki-kare istatistiği

P(TG): LR için toplam puan x grup etkileşim değerlerinin olasılığı

Yanlılık Değerleri Arasındaki İlişki

Lojistik Regresyon, MH, işaretsiz Ki-kare ve işaretili Ki-kare teknikleriyle elde edilen Ki-kare değerleri arasındaki sıra farkları korelasyonları Tablo 5'te görünmektedir.

Tablo 5.

LR, MH, İşaretsiz Ki-Kare ve İşaretili Ki-Kare Tekniklerine Göre Elde Edilen Ki-Kare Değerleri Arasındaki Sıra Farkları Korelasyonları

Teknikler	LR χ^2	MH χ^2	χ^2_c	χ^2_{si}
LR χ^2	1			
MH χ^2	0,934**	1		
χ^2_c	0,906**	0,789**	1	
χ^2_{si}	0,838**	0,812**	0,867**	1

** p<0,01

Tabloya göre bütün korelasyonlar $\alpha = 0,01$ düzeyinde manidar bulunmuştur. Korelasyonların değerleri 0,934 (LR ve MH teknikleri arasında) ile 0,789 (MH ile işaretli Ki-kare teknikleri arasında) değerleri arasında değişmektedir. LR tekniğinin MH ve işaretli Ki-kare tekniğinden elde edilen Ki-kare değerlerinin büyüklük sırası bakımından benzer sonuçlar verdiği söylenebilir.

Sonuçlar ve Tartışma

Tekniklere göre DMF veren madde sayısı 0 ile 15 arasında değişmektedir. Söz konusu tekniklere göre elde edilen yanlı madde oranları arasında manidar bir fark bulunduğu söylenebilir. DMF'li madde sayısı en fazla MH tekniğinde elde edilmiştir. İşaretsiz Ki-kare tekniği 9 madde ile ikinci sıradadır. İşaretsiz Ki-kare ve LR tekniğinde ise hiçbir madde için DMF ortaya çıkmamıştır. LR ve işaretli Ki-kare tekniklerinin DMF'li madde kararı bağlamında benzer sonuçları ürettiği söylenebilir.

LR ve MH tekniklerinin karşılaştırıldığı çok sayıda araştırma bulunmaktadır. Gierl, Khaliq ve Boughton (1999) tarafından yapılan Matematik ve Fen Bilgisi başarı testindeki maddelerin DMF verip vermediğini kontrol ettikleri çalışmada, bu araştırma ile uyumlu sonuçlara ulaşılmıştır. Söz konusu çalışmada LR ve MH tekniği ile aynı testler için farklı sayıda maddenin DMF verdiği bulunmuştur. Bu çalışmada LR tekniğinin MH tekniğinden daha az DMF'li madde vermesi, Hidalgo ve Lopez-pina'nın (2004) benzetim (simülasyon) yoluyla her biri 75 soruya sahip 25 test üreterek LR ve M-H tekniklerini karşılaştırdıkları çalışmadaki sonuçlarla da uyum göstermektedir. Söz konusu araştırma sonucunda, LR tekniği kullanılarak yapılan hesaplarda, MH tekniğine göre tek biçimli olmayan DMF'ye sahip daha fazla maddenin ortaya çıktığı sonucuna ulaşılmıştır. Kısmen benzer başka bir çalışmada ise, Hambleton ve Rogers (1990) LR ve MH tekniklerini karşılaştırmışlar; tek biçimli olmayan DMF'yi belirlemede LR tekniğinin daha güçlü; tek biçimli DMF'yi belirlemede de iki tekniğin benzer sonuçlar verdiğini bulmuşlardır. Bu sonuçlar eldeki çalışma ile kısmen uyumsuzdur. Diğer yandan, DMF belirlemede kullanılan SIBTEST (Simultaneous Item Bias Test) LR ve MH tekniklerinin gözlem sayısı 250, 500 ve 1000 olan farklı örneklem büyüklüklerinde gücünün (power analysis) araştırıldığı bir çalışmada (Gierl, Jodoin & Ackerman, 2000), LR ve MH tekniklerinin birbirine yakın sonuçlar verdiği sonucuna ulaşılmıştır. Söz konusu çalışma ile bu çalışmada elde edilen sonuçların uyumlu olduğu söylenemez. Gomez-Benito ve Navas-Ara (2000), örtük özellikler kuramına dayanan alan indeksleri Ki-kare temelli MH, lojit model ve lojistik regresyon; son olarak sınırlandırılmış faktör analizi teknikleriyle elde edilen DIF değerlerinin karşılaştırmasını amaçladığı çalışmada M-H tekniğinde daha fazla maddenin DMF verdiğini bulmuştur. Benito ve Navas-Ara'nın bulduğu sonuçla bu çalışmadaki sonuçların uyumlu olduğu söylenebilir. Bu çalışmada da LR ile elde edilen DMF'li madde sayısı MH ile elde edilen DMF'li madde sayısından az bulunmuştur. Näsström (2003) ise yaptığı çalışmada MH ve LR teknikleri ile birbirine benzer sonuçlar elde ettiğini bildirmektedir. Yurdugül (2003) 2001 yılı Ortaöğretim Kurumları Seçme ve Yerleştirme Sınavı'nın alt ölçeklerinin madde yanlılığı açısından incelenmesiyle ilgili çalışmada, LR ve MH tekniklerinin Türkçe, Fen Bilimleri, Sosyal Bilimler alt testi için tamamen benzer sonuçlar verdiğini; Matematik alt testi için sonuçların bir iki madde için değişebildiğini belirtmiştir. Bekci (2007) tarafından yapılan "Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı'nın değişen madde fonksiyonlarının cinsiyet ve okul türüne göre incelenmesi" başlıklı çalışmada, LR ve MH tekniklerinin tam bir uyum göstermedikleri, alt testlere ve grup değişkenine göre farklı sonuçlar üretebildikleri gözlenmiştir. Bekci'nin bulguları ile bu araştırma bulgularının benzerlik gösterdiği söylenebilir.

Tablo 4'e göre teknikler için elde edilen Ki-kare değerlerinin birbirine büyüklük sırası bakımından belirli ölçülerde paralel olduğu söylenebilir. 9, 10, 12, 13, 15, 16, 19, 20, 22, 23 ve 25. maddelerin MH ve işaretli (Camilli) Ki-kare tekniklerinin her ikisinde de DMF verdiği görülmektedir. LR ve işaretli (Shepard) Ki-kare tekniklerinde ise hiçbir DMF'li madde belirlenmesi yapılmamasına rağmen, diğer tekniklerde DMF veren maddelerin bu tekniklerle hesaplanan ve DMF değerlendirmesinde kullanılan Ki-kare değerlerinin diğer maddelere göre görece bir yükseklığe sahip olduğu söylenebilir.

Ancak DMF veren madde sayılarına ilişkin sonuçlar, teknikler arasında fark olduğuna ilişkin yorumlar yapılmasına neden olsa da her bir teknikte maddeler için elde edilen ve DMF belirlemede kullanılan Ki-kare değerleri arasındaki ilişkiler ($P < 0,01$ düzeyinde) manidar bulunmuştur. Her bir teknikte DMF belirlemek için Ki-kare değerinin hesaplanması teknikleri bir ölçüde karşılaştırılabilir kılınmaktadır. En azından Ki-kare değerlerinin büyüklük sıraları karşılaştırılabilir. Bunun için kullanılan sıra farkları korelasyonları incelendiğinde, Ki-kare değerleri arasındaki sıra farkları korelasyonlarının manidar olduğu, dolayısıyla teknikler arasında bir paralellik ya da ilişki olduğu söylenebilir.

Tam bu noktada, her teknikte Ki-kare değerinin elde edilişi ve DMF veren maddelerin belirlenmesinde kullanılan ölçütlerin farklılaştığını vurgulamak yararlı olabilir. DMF'li maddeleri belirlemede kullanılan ölçütlerin farklılığı, tekniklere göre elde edilen DMF değerlerinin paralellüğünün aksine, tekniklere göre DMF olduğu öne sürülen maddelerin sayısını ve DMF düzeyine ilişkin yorumları değiştirmektedir. Özellikle DMF için Ki-kare değeri üreten ve kullanan bu tekniklerin hepsinde kullanılacak ölçütler geliştirme, karşılaştırmaların daha sağlam temellere dayanmasını getirebilir.

Elde edilen bulgulara dayanarak hangi tekniğin pratikte daha avantajlı olduğu sorusuna birkaç yönden bakarak cevap verilebilir ve bu bakış açılarına göre avantajlı kabul edilecek teknikler değişmektedir. Tekniklerden elde edilen korelasyonlara dayanarak Ki-kare temelli bu tekniklerin hepsinin de kullanılabileceği, aralarında önemli farklılıklar olmadığı, birbirine benzer sonuçlar verdiği söylenebilir. Diğer yandan her bir teknikte DMF'li maddeler ve sayısı dikkate alındığında, duyarlı sonuçlar veren ve daha önceki çalışmalarla (Bekci 2007, Yurdugül, 2003) tutarlı sonuçlar veren lojistik regresyon tekniğinin kullanılması önerilebilir. Diğer yandan ortak olan DMF'li madde sayısı dikkate alındığında, MH ve işaretli Ki-kare tekniğinin birbirinin yerine kullanılabilmesi de mümkün görünmektedir.

Özet olarak, MH, Ki-kare ve lojistik regresyon tekniklerinin Ki-kare değerlerinin büyüklüğü bakımından benzer; DMF verdiği belirlenen madde sayısı bakımından farklı sonuçlar üretmesi bu araştırmanın en önemli sonucudur.

Kaynakça

- Adams, R. J., ve Rowe, K. J. (1988). Item bias. In J.P.Keeves (ed.) *Educational research, methodology, and measurement: An international handbook*. Oxford: Pergamon Press.
- Camilli, G. (1979). *A critique of the chi-square method for assessing item bias*. Unpublished paper Laboratory of Educational Research, University of Colorado.
- Camilli, G., Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks: Sage Publications.
- Crocker, L., ve Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando: Rinehart and Winston, Inc.
- Devine, P. J., ve Raju N. S. (1982). Extent of overlap among four item bias methods. *Educational and Psychological Measurement*, 42, 1049-1066.

- Doğan, N., ve Öğretmen, T. (2006). Madde yanlılığını belirleme teknikleri arasında bir karşılaştırma. *Eğitim Araştırmaları*, 23, 94–105
- Dorans, N. J., ve Kulick, E. M. (1983). *Assessing unexpected differential item performance of female candidates on SAT and TSWE forms administered in December 1977*(ETS Research Report RR–83–9), Princeton: New Jersey.
- Dorans, N. J., ve Holland, P. W. (1993). DIF detection and description: Mantel Haenszel and standardization. In P. W. Holland, ve H. Wainer, (Eds.), *Differential Item Functioning* (pp. 35–66), New Jersey: USA.
- Gierl, M., Khaliq, S. N., Boughthon, K. (1999). *Gender differential item functioning in mathematics and science: Prevalence and policy implications*. Paper presented at the semposium entitled “Improving large – scale assessment in education” at the Annual Meeting of the Canadian Society for the Study of Education, Canada, June, 1999.
- Gierl, M., Jodoin, G. M. & Ackerman, T. A. (2000). Performance of Mantel-Haenszel, Simultaneous Item Bias Test, and Logistic Regression when the proportion of DIF items is Large. Paper Presented at the Annual Meeting of the American educational Research Association (AERA). New Orleans, Louisiana, USA. Erişim 17 Temmuz 2006, from the Centre for Research in Applied Measurement and Evaluation (CRAME) website: <http://www.education.ualberta.ca/educ/psych/crame/>
- Gomez-Benito, J. ve Navas-Ara, M. J. (2000). A comparison of Ki-kare, RFA and IRT based procedures in the detection of DIF. *Quality ve Quantity* 34: 17–31.
- Swaminathan, H., ve Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361-370.
- Hambleton, R K., Swaminathan, H. ve Rogers, H. J. (1991). *Fundamentals of item response theory*. London: Sage Publication.
- Hidalgo, M.D. ve Lopez-Pina, J.A. (2004). Differential item functioning detection and effect size: Comparison between logistic regression and mantel-haenszel procedures. *Educational and Psychological Measurement*, 64, 6, 903–915
- Hills, J. R. (1984). *Quantitative methods used in the study of item bias*. ERIC Document Reproduction Service No. ED 247 271.
- 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.
- Ironson, G. H., ve Craig, R. (1982). *Item bias techniques when amount of bias is varied and score differences groups are presented*. University of South Florida, Tampa. Department of Psychology. (ERIC Document Reproduction Service No. ED 227 146).
- Mellenberg, G. J. (1983). Conditional item bias methods. In S. H. Irvine and W. J. Barry (Eds), *Human assesment and cultural factors* (pp. 293–302). New York: Plenum Pres.
- Mellenberg, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research: Applications of Item Response Theory*.13, 123–144.
- Näsström, G. (2003). Differential item functioning for items in the swedish national test in mathematics, course B. *Nordic pre-conference to ICME10 at Växjö University*, May 9–11
- Osterlind, S. (1983). *Test item bias*. Newbury Park: Sage Publications.
- Öğretmen, T. (1995). *Differential item functioning analysis of the verbal ability section of the first stage of the university entrance examination in Turkey*. Yayınlanmamış yüksek lisans tezi, Orta Doğu Teknik Üniversitesi..
- Öğretmen, T., ve Doğan, N. (2004). OKÖSYS Matematik alt testine ait maddelerin yanlılık analizi. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*. 8, 61–76.
- Özdemir, D. (2003). Çoktan seçmeli testlerde iki kategorili ve önsel ağırlıklı puanlamanın diferansiyel madde fonksiyonuna etkisi ile ilgili bir araştırma. *Eğitim ve Bilim*. 25; 37–44

- 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., ve 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.
- Rudner, L., Getson, P. R. ve Knight, D. L. (1980). Biased item detection techniques. *Journal of Educational Statistics*. 5, 213–233.
- Scheuneman, J. (1979). A new method for assessing bias in test items. *Journal of Educational Measurement*, 16, 143–152.
- Seong, Tae-Je., ve Subkoviak, M. J. (1987). *A comparative study of recently proposed item bias detection methods*. Paper presented at the annual meeting of the American Educational Research Association, Toronto. (ERIC Document Reproduction Service No. ED 157 942).
- Shepard, L. A., Camilli, G., ve Williams, D. M. (1984). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement*. 22, 77–105.
- Swaminathan, H., ve Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361–370.
- Thomas, D. R., ve Zumbo, B. D. (1998). *Variable importance in logistic regression based on partitioning an R-squared measure*. Presented at the Psychometric Society Meetings, Urbana, IL.
- Tittle, C. K. (1988). Test Bias. In J.P. Keeves, (ed.). *Educational research, methodology, and measurement: An international handbook*. Oxford: Pergamon Press.
- Waller N. G. (1998). EZDIF: Detection of uniform and nonuniform differential item functioning with the Mantel-Haenszel and Logistic regression procedures. *Applied Psychological Measurement*, 22: 391
- Yenal, E. (1995). *Differential item functioning analysis of the quantitative ability section of the first stage of the university entrance examination in Turkey*. Yayınlanmamış yüksek lisans tezi, Orta Doğu Teknik Üniversitesi.
- Yurdugül, H. (2003). *Ortaöğretim kurumları seçme ve yerleştirme sınavının madde yanlılığı açısından incelenmesi*. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi.
- 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*. Canada: Ottawa, Directorate of Human Resources Research and Evaluation National Defense Headquarters: Author.
- Zumbo, B. D., ve Thomas, D. R. (1997) *A measure of effect size for a model-based approach for studying DIF*. Working Paper of the Edgeworth Laboratory for Quantitative Behavioral Science, University of Northern British Columbia: Prince George, B.C.

Makale Geliş: 26.06.2007
İnceleme Sevk: 28.06.2007
Düzeltilme: 09.09.2007
Kabul: 03.01.2008