

MIMIC, SIBTEST, Lojistik Regresyon ve Mantel-Haenszel Yöntemleriyle Gerçekleştirilen DMF ve Yanlılık Çalışması*

A DIF and Bias Study by using MIMIC, SIBTEST, Logistic Regression and Mantel-Haenszel Methods

Çiğdem AKIN ARIKAN**, Seçil UĞURLU***, Burcu ATAR****

ÖZ: Bu çalışmanın amacı değişen madde fonksiyonu (DMF) belirleme yöntemlerinden MIMIC, SIBTEST, Lojistik Regresyon ve Mantel-Haenszel yöntemlerinin benzerlik ve farklılıklarını belirlemektir. Ayrıca uzman görüşleri ile yöntemlere göre elde edilen sonuçların tutarlılığı incelenmiştir. Araştırma yaklaşık 340000 öğrencinin yer aldığı veri setinden 300, 600 1000, 1200 ve 2000 kişilik alt örneklem seçilerek yürütülmüştür. Dört ayrı yöntem için, örneklemelere göre DMF içeren ortak maddelere baktığımızda; 300 kişilik örnekleme 2. madde; 600 kişilik örnekleme 13. madde; 1000 kişilik örnekleme ortak DMF içeren madde olmadığı, 1200 kişilik örnekleme 19. madde ve 2000 kişilik örnekleme 2, 3 ve 4. maddelerin DMF içerdiği görülmüştür. Bu bakımdan büyük örnekleme, yöntemlerin birbiriyle daha uyumlu sonuçlar verdiği söylenebilir. Uzman kanılarına göre yanlı olduğu belirlenen 2, 5, 6 ve 12. maddeler istatistiksel analizlerin sonuçlarıyla karşılaştırıldığında bu maddelerin farklı yöntem ve ya farklı örneklemelerde DMF gösterdiği görülmüştür. Ayrıca uzman kanılarıyla analiz sonuçlarının birbirleriyle tutarlı olduğu söylenebilir.

Anahtar sözcükler: MIMIC, SIBTEST, lojistik regresyon, Mantel-Haenszel, DMF

ABSTRACT: The aim of this study was to find out similarities and differences between methods for differential item functioning (DIF) such as MIMIC, SIBTEST, Logistic Regression and Mantel-Haenszel methods. Also the results obtained with expert opinions according to the methods were examined in terms of consistency. The study was carried out with subsamples of 300, 600, 1000, 1200 and 2000 examinees selected from the dataset of approximately 340.000 students. For the four methods, common items containing DIF were examined by sample groups. It was seen that item 2 does not contain common DMF in the sample of 300 persons, item 13 in the sample of 600, and no items contain common DIF in the sample of 1000 persons; whereas item 19 contains DIF in the group of 1200 and items 2, 3, and 4 contain DIF in the group of 2000 persons. In the light of this, it can be suggested that the methods for 2000 persons yielded more compliant results in the large sample. By comparing items 2, 5, 6 and 12 identified to be biased according to expert opinion with statistical analysis results, it was found out that those items showed DIF with different methods or samples. In addition, expert opinions seem to be consistent with results of the analysis.

Keywords: MIMIC, SIBTEST, logistic regression, Mantel-Haenszel, DIF

1. GİRİŞ

Ülkemizdeki geniş ölçekli sınavlardan biri olan Seviye Belirleme Sınavı (SBS) ortaöğretim kurumlarına geçiş amacıyla 2007-2008 öğretim yılından 2012-2013 öğretim yılına kadar 6., 7. ve 8. sınıf öğrencilerine kademeli olarak uygulanmıştır. SBS; fen ve teknoloji, matematik, Türkçe, sosyal bilgiler ve İngilizce alt testlerinden oluşmaktadır. SBS gibi bireyler

* Bu çalışma, 9-13 Haziran 2014 tarihlerinde gerçekleştirilen IV. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresinde sunulmuştur.

** Arş. Gör., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: cakinarikan@hacettepe.edu.tr

*** Arş. Gör., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: secilarlan@hacettepe.edu.tr

**** Doç.Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: burcua@hacettepe.edu.tr

hakkında önemli kararlar alınmasına yardımcı olan sınavlar için hazırlanan maddelerin ölçülen özellik bakımından bireyler arasındaki farklılıkları ayırt edebilmesi ve aynı yetenek düzeyindeki bireyler için yanlılık göstermemesi gerekmektedir. Test sonuçlarının geçerli bir şekilde yorumlanmasını ve kullanılmasını tehdit eden etkenlerden biri olan yanlılık (Zieky, 2006); test maddesinin bazı özelliklerinden veya test koşullarından dolayı, bir gruptaki bireylerin maddeyi doğru cevaplama olasılığının diğer gruptaki bireylere göre daha fazla olmasıdır (Zumbo, 1999). Madde yanlılığı çalışmaları madde fonksiyonunun aynı yetenek düzeyinde ve farklı gruplarda yer alan bireyler için farklılaşıp farklılaşmadığını belirleyen istatistiksel bir süreçle başlar ve bu süreç iki aşamadan oluşur.

Madde yanlılığı çalışmalarında ilk aşama değişen madde fonksiyonu (DMF) gösteren maddelerin belirlenmesidir. DMF, aynı yetenek düzeyine sahip farklı gruptaki bireylerin maddeyi doğru cevaplama olasılıklarının farklılaşmasıdır. Doğan, Guerrero ve Tatsuoka (2005) göre ise DMF, aynı test puanına, aynı yetenek düzeyine sahip bir gruptaki bireylerin sistematik olarak belirli bir test maddesini çözme olasılığı bakımından farklılık göstermesidir. DMF belirleme yöntemlerinde, genel olarak testi alan grup referans ve odak grup olmak üzere ikiye ayrılır. Referans grup, odak gruba göre avantajlı olan gruptur. Referans ve odak grup üzerinde DMF belirleme yöntemleri uygulanarak, grup farklılıkları belirlenir. DMF; tek biçimli (uniform) ve tek biçimli olmayan (non-uniform) olmak üzere iki şekilde incelenebilir. Tek biçimli DMF, yetenek düzeyi ile grup üyeliği arasında herhangi bir etkileşim olmadığı zaman ortaya çıkar. Diğer bir ifadeyle, bütün yetenek düzeylerinde, maddenin doğru cevaplanma olasılığının bir grubun lehine işlenmesidir. Tek biçimli olmayan DMF ise; aynı yetenek düzeyindeki iki grubun maddeyi doğru cevaplama olasılığının bütün yetenek düzeylerinde aynı olmaması yani maddenin farklı yetenek düzeylerinde farklı grubun lehine işlenmesidir (Gierl, Rogers ve Klinger, 1999).

Madde yanlılığı çalışmalarında ikinci aşamada ise testin yapısı veya kapsamı dikkate alınarak yapılan incelemeler sonunda DMF gösteren bir maddenin yanlı olup olmadığına karar verilir (Dorans ve Holland, 1993). Ancak maddenin doğru cevaplanma olasılığındaki farklılık yanlılığın kanıtı olmayabilir. Bazı durumlarda farklı alt gruplardan gelen katılımcıların yetenekleri gerçekte birbirinden farklı olabilir. Bu durumlarda performanslar arasında gözlenen farklılığa madde etkisi denir (Clauser ve Mazor, 1998). Madde etkisi, maddeler tarafından ölçülen yetenek bakımından gruplar arasında gerçek bir farklılık olması nedeniyle DMF'nin ortaya çıktığı durumları anlatmak için kullanılır (Zumbo, 2007). DMF, bahsedildiği üzere gruplar arasındaki gerçek farklılıktan değil de madde özellikleri veya test koşullarından kaynaklanıyorsa bu durum yanlılık olarak ifade edilir. İdeal olarak, bir maddeyi doğru cevaplama olasılığı sadece bireyin yeteneğine bağlıdır. Eğer bu olasılık ırk, köken, cinsiyet ve/veya sosyoekonomik düzey gibi faktörlere bağlı ise DMF mevcut olabilir. DMF madde yanlılığını belirlemede gerekli ancak yeterli değildir. Eğer bir madde DMF içermiyorsa, yanlı da değildir. Bununla birlikte; eğer bir madde DMF içeriyorsa, yanlı olup olmadığı (içerik analizi, ampirik değerlendirme) araştırılmalıdır (Zumbo, 1999)

DMF belirleme yöntemleri hem iki kategorili hem de çok kategorili puanlanan veriler için geliştirilmiştir. İkili kategorili puanlanan veriler için, Mantel-Haenszel, lojistik regresyon, Lord'un Ki-Kare testi, SIBTEST, olabilirlik oranı testi, MIMIC gibi DMF yöntemleri kullanılmaktadır. Bu çalışmada da, iki kategorili verilerde DMF sonuçlarının karşılaştırabilmek amacıyla SIBTEST, Mantel -Haenszel (MH), MIMIC ve Lojistik regresyon (LR) yöntemleri birlikte ele alınmıştır.

Mantel-Haenszel (MH) Yöntemi: Parametrik olmayan bir yöntem olan Mantel-Haenszel istatistiği (Gierl, Rogers ve Klinger,1999) DMF belirlemede yaygın kullanılan yöntemlerden birisidir ve ilk olarak Holland ve Thayer (1988) tarafından geliştirildiğinden bu yana pek çok

çalışmaya konu olmuştur (Clauser ve Mazor, 1998). MH istatistiği serbestlik derecesi bir olan ki- kare dağılımı gösterir ve iki gruptaki aynı yetenek düzeyine sahip bireyler ile madde arasında herhangi bir ilişki olmadığı şeklinde ifade edilen yokluk hipotezini test etmek için kullanılır (Gierl, Jodoin, Ackerman, 2000). Bu yöntem, yeteneği karşılaştırılan grubun üyelerinin madde üzerindeki başarı olasılıklarının belirlenmesinde uygulanır. Bireyler toplam puanları baz alınarak yeteneklerine göre seviyelere ayrılır ve her bir yetenek düzeyi için aşağıdaki gibi 2X2'lik bir tablo oluşturulur.

Tablo 1: Madde cevaplarının Dağılımı

Grup	1 (Doğru)	0 (Yanlış)	Toplam
Referans	A_j	B_j	N_{rj}
Odak	C_j	D_j	N_{oj}
Toplam	M_{1j}	M_{0j}	T_j

Tablo 1'deki değerler kullanılarak aşağıdaki formülle α_{MH} indeksi elde edilir. Elde edilen indeks olabilirlik oranı üzerine kuruludur (Clauser ve Mazor, 1998).

$$\alpha_{MH} = \frac{\sum A_j D_j / T_j}{\sum B_j C_j / T_j} \quad (1)$$

$$\Delta MH = -(2.35) \ln(\alpha_{MH}) \quad (2)$$

Denklem (1)'de elde edilen olabilirlik oranının doğal logaritmasının -2,35 katı ΔMH istatistiğini verir. MH değerinin 1 olması her iki grup için doğru yanıt olasılıklarının eşit olduğunu gösterir. Eğer referans grubun doğru yanıt olma olasılığı büyük ise $MH > 1$ ve eğer odak grubun doğru yanıt olma olasılığı büyük ise $MH < 1$ olacaktır. Eğitimsel Test Hizmetleri (Educational Testing Service-ETS), DMF etki büyüklüğüne ait bazı değerler önermiş ve bu değerler Roussos and Stout (1996) tarafından modifiye edilmiştir. $|\Delta MH| < 1$ ise ihmal edilir düzeyde; $1 \leq |\Delta MH| < 1,5$ ise orta düzeyde, $|\Delta MH| \geq 1,5$ yani sıfırdan büyük ölçüde farklılık gösteriyor ise yüksek düzeyde DMF olduğunu gösterir. Bu dereceler ise; ihmal edilir düzey A, orta düzey B ve yüksek düzey C olarak adlandırılır (Gierl, Rogers ve Klinger, 1999). Bu yöntemle sadece tek biçimli DMF belirlenebilir. Ayrıca Clauser ve Mazor (1998), bu tekniğin küçük örneklerde de oldukça etkili olduğunu belirtmişlerdir.

SIBTEST Yöntemi: Eşzamanlı Yanlılık Testi (SIBTEST), Shelay ve Stout (1993) tarafından DMF belirlemek için önerilen istatistiksel bir metottur (Zheng, Gierl ve Cui, 2007). SIBTEST diğer bir DMF belirleme yöntemi olan standartlaştırma yöntemine benzer olmakla birlikte önemli bir farklılık içermektedir. Bu farklılık da anlamlılık testi olup, odak ve referans gruptakiler için doğru cevaplar sayısındaki ağırlıklandırılmış farkın standart hataya oranına dayalıdır. Ayrıca eşleştirme kriteri olarak gözlenen puandan çok örtük puanı kullanır (Clauser ve Mazor, 1998). SIBTEST, çok boyutlu veriler için geliştirilmiş ise de, tek boyutlu verilerde de kullanılabilir. DMF belirlemenin yanında, DMF miktarını da β etki büyüklüğünü kullanarak belirler. DMF etki büyüklüğünü yorumlamak için kullanılan ölçütler Roussos ve Stout, (1996) tarafından önerilmiştir. Bu ölçütler;

A_β İhmal edilebilir düzeyde DMF, $|\beta| < 0,059$;

B_β Orta düzeyde DMF, $0,059 \leq |\beta| < 0,088$;

C_β Yüksek düzeyde DMF, $|\beta| \geq 0,088$ 'dir.

Elde edilen pozitif β değeri referans grup lehine, negatif β değeri ise odak grup lehine maddenin DMF gösterdiğini belirtir.

MIMIC Yöntemi: DMF açısından tartışılan bir diğer yaklaşım ise Doğrulayıcı Faktör Analizi (DFA) modellerinin madde yanıt verilerine uygulanması ile ilgilidir. Camilli ve Shepard (1994), grup farklılıklarının ikincil bir faktör üzerinden karşılaştırılmasına izin verdiği için, DFA'nın DMF'nin belirlenmesinde kullanılabileceğini belirtmiştir. DFA kullanarak DMF belirlenmesine ilişkin bazı yaklaşımlar vardır. Bunlara modifikasyon indeksleri, çoklu grup DFA ve MIMIC (Çoklu Göstergeler Çoklu Nedenler) yöntemi örnek olarak verilebilir. Bu çalışmada MIMIC yöntemi kullanılacağından bu yöntem hakkında bilgi verilmiştir.

MIMIC yöntemi pratikte maddelerde tek biçimli DMF'in bulunup bulunmadığının kontrol edilmesinde kullanılmaktadır. DMF'in yönünün ve büyüklüğünün belirlenmesinde, bu yöntem ile elde edilen model parametrelerinin tepki kuramındaki madde güçlük indislerine nasıl dönüştürülebileceğini göstermişlerdir. DMF bağlamındaki MIMIC modeli aşağıdaki şekildedir (Finch, 2005):

$$y^*_i = \lambda_i \eta + \beta_i z_k + \varepsilon_i \quad (3)$$

y^*_i , gizil cevap değişkeni, η = gizil özellik, λ_i = i değişkeni için faktör yükü, ε_i = tesadüfi hata, z_k = etkisiz değişken olup grup üyeliği anlamındadır. β_i = yanıtla birlikte grup değişkenine ilişkin eğriyi gösterir.

MIMIC yöntemi kullanılarak DMF'nin değerlendirilmesi, bir grupta değişkeninin hem doğrudan hem dolaylı etkilerinin tahmin edilmesine dayalıdır. Dolaylı etki, gizil değişken ortalamasının gruptan gruba değişiklik gösterip göstermediğini belirtir, dolayısıyla gizil özellik üzerindeki grup farklılıklarını açıklar. Doğrudan etki ise madde yanıtı olasılıklarının gruplar arasında değişiklik gösterip göstermediğini belirtir. Bu formül gruplar için gizil özellik ortalamasındaki değişiklikler kontrol edildikten sonra tek biçimli DMF'nin kontrol edilmesinde kullanılır (Finch, 2005; Finch ve French, 2007). Ayrıca bu yaklaşımda kategorik ve sürekli değişkenlerle ikiden fazla grup için DMF belirlenebilir ve tek bir süresiz grupta değişkeni ile DMF analizi yapmaya izin veren geleneksel DMF yöntemlerine göre daha esneklerdir (Wang, Shih ve Yang, 2009). DFA da gözlenen değişkenler kategorik olduğundan gözlenen değişkenlerin normalliği sayılına dayanan maksimum likelihood (olasılık) yönteminin kullanılması hatalı sabit bulunmasına neden olur. Bu nedenle WLS ya da RWLS kestirimi kullanılması tavsiye edilir (Finch, 2005; Finch ve French, 2007).

Lojistik Regresyon Yöntemi: DMF'yi belirlemek amacıyla lojistik regresyon yöntemi tek biçimli olmayan DMF'yi tanımlaması açısından MH istatistiğine göre daha üstündür.

Lojistik regresyon yöntemi ile DMF araştırılırken üç alt model kullanılır:

$$\begin{aligned} z &= \beta_0 + \beta_1 X \\ z &= \beta_0 + \beta_1 + \beta_2 G \\ z &= \beta_0 + \beta_1 X + \beta_2 G + \beta_3 GX \end{aligned} \quad (4)$$

Burada X, G ve GX değişkenleri sırasıyla; test puanlarını, grubu ve grup ve test puanları etkileşimini ifade eder. Modelde X değişkeninin anlamlı olması modelin geçerliğini, G değişkeninin anlamlılığı tek biçimli DMF'yi, GX değişkeninin modeldeki anlamlılığı ise tek biçimli olmayan DMF'yi gösterir (Gierl, Jodoin ve Ackerman, 2000; akt. Bekçi, 2007).

Lojistik regresyon yönteminin üç aşamasında da etki büyüklüğü kestiricisi elde edilir. Etki büyüklüğü R^2 değeri ile hesaplanır. İki modelin R^2 değerlerinin farkı DMF miktarının yorumlanmasında kullanılan etki büyüklüğünü verir. Etki büyüklüğü miktarının, ΔR^2 , yorumlanması için Zumbo ve Thomas (1996) ile Jodoin ve Gierl (2001) Tablo 2’de verilen iki ayrı sınıflamayı önermişlerdir.

Tablo 2: ΔR^2 Değerlerinin Yorumlanması İçin Önerilen Sınıflama Kategorileri

Zumbo ve Thomas (1996)	Jodoin ve Gierl (2001)	Düzye
$\Delta R^2 < 0.13$	$\Delta R^2 < 0,035$	A (Yok veya ihmal edilebilir düzeyde)
$0.13 \leq \Delta R^2 < 0.26$	$0.035 \leq \Delta R^2 < 0.070$	B (Orta düzeyde)
$\Delta R^2 \geq 0.26$	$\Delta R^2 \geq 0.070$	C (Yüksek Düzeyde)

Tablo 2’de, Jodoin ve Gierl tarafından belirlenen ölçütlerin Zumbo ve Thomas’ın belirlediklerine göre daha katı olduğu görülmektedir. Bu çalışmada Jodoin ve Gierl (2001) tarafından önerilen ölçütler esas alınmıştır.

Ülkemizde, SBS de yer alan maddelerin DMF içerip içermediğinin belirlenmesine yönelik çalışmalar bulunmaktadır (Karakaya, 2012; Karakaya ve Kutlu, 2012). Literatürde DMF tespitinde MIMIC modelinin etkililiğinin değerlendirilmesi konusunda simülasyon çalışmaları bulunmaktadır (Finch, 2005; Stark, Chernyshenko ve Drasgow, 2006; Finch ve French, 2007; Shih ve Wang, 2009; Woods, 2009; Finch ve French, 2011; Kim ve Yoon, 2011). Gerçek veri setlerinde MIMIC modelinin kullanıldığı az sayıda çalışmaya rastlanılmıştır (MacIntosh ve Hashim, 2003; Woods, Oltmanns ve Turkheimer, 2009). Alan yazında MIMIC, SIBTEST, LR ve M-H yöntemlerinin bir arada kullanıldığı herhangi bir çalışmaya rastlanmamıştır. Literatür incelendiğinde DMF analizi sonucunda DMF içerdiği belirlenen maddelerle ilgili uzman kanılarına başvurulduğu görülmüştür (Yurdugül ve Aşkar, 2004; Balkan Kalaycıoğlu ve Kelecioğlu, 2011; Çepni, 2011; Karakaya, 2012; Kelecioğlu, B. Karabay ve E. Karabay 2014). Bu çalışmada ise testte yer alan tüm maddelerin farklı işleyip işlemediği uzman kanılarıyla incelenmiştir. Araştırmada farklı örneklem ve daha önce bir arada kullanılmamış farklı tekniklerin kullanılması, ayrıca uzmanlara soruların hepsinin verilerek cinsiyet açısından yanlı olup olmadıklarının sorulması bakımından bu araştırma sonuçlarının yeni çalışmalara katkı sağlayacağı düşünülmektedir.

Problem durumu

Bu çalışmayla, 2009 SBS 8. sınıf fen ve teknoloji alt testinde yer alan maddelerin farklı örneklemelerde SIBTEST, MH, MIMIC ve LR yöntemleriyle yapılan analizlerinde cinsiyete göre farklı fonksiyonlaşma gösterip göstermediği ve elde edilen sonuçların uyumlu olup olmadığı sorularına yanıt aranmıştır. Çalışmada SIBTEST, MH, MIMIC ve LR DMF belirleme yöntemleri kullanılarak elde edilen sonuçlar karşılaştırılmıştır. MIMIC yöntemi ile SIBTEST, MH ve LR yöntemleri arasında yapılacak karşılaştırmaların alan yazına sağlayacağı katkılar çalışmanın özgün yanını oluşturmaktadır. Örneklem büyüklükleri değiştikçe yöntemlerin 1. tip hata oranları da değişebileceğinden çalışmada kullanılan dört tekniğin ortak olarak belirlediği DMF’li maddelere ulaşılarak araştırmalarda birden çok tekniğin kullanılmasının daha doğru olacağı sonucuna varılabilir. Bu amaçla aşağıdaki sorulara cevap aranmıştır:

Odak ve referans grup için farklı örneklem büyüklükleri ve farklı oranlar söz konusu olduğunda, cinsiyet grupları bakımından SIBTEST, MH, MIMIC ve LR yöntemlerine göre ortak DMF'li maddeler ve bunların sayısı nedir?

Uzman kanıları sonucunda yanlı olduğu belirlenen maddeler ile istatistiksel analizler sonucunda DMF gösteren maddeler birbirleriyle tutarlı mıdır?

2. YÖNTEM

2.1. Araştırmanın Türü

Bu çalışmada 2009 8. sınıf Seviye Belirleme Sınavı fen ve teknoloji alt testlerinde yer alan 20 maddenin cinsiyet değişkenine göre DMF gösterip göstermediği analiz edilmiştir. Analiz sonucunda tüm maddeler için uzman kanısına başvurularak yanlılık araştırması yapılmıştır. Bu çalışmada DMF belirleme yöntemi olarak MIMIC, SIBTEST, LR ve MH yöntemleri kullanılmıştır. Bu çalışma var olan durumu belirlediğinden dolayı betimsel bir çalışmadır.

2.2 Araştırmanın Evreni ve Örneklemi

Araştırmanın evrenini, MEB tarafından 2009 yılında uygulanan SBS'ye giren yaklaşık 340.000 8. sınıf öğrencisi oluşturmaktadır. Araştırmanın örneklemi yansız olarak, öğrencilerin cinsiyetleri dikkate alınarak seçilmiştir.

Yaklaşık 340.000 öğrencinin yer aldığı veri setinden, $N_1=300$ kişilik (referans grup:150, odak grup:150), $N_2=1000$ kişilik (referans grup:500, odak grup:500), $N_3=2000$ kişilik (referans grup:1000, odak grup:1000), $N_4=600$ kişilik (referans grup:500, odak grup:100), $N_5=1200$ kişilik (referans grup:1000, odak grup:200) 5 alt örneklem seçilmiştir. Örneklem sayılarının seçiminde Finch (2005) 'in çalışmasında kullandığı simülasyon koşulları dikkate alınmıştır. Verilere ilişkin betimsel istatistikler için SPSS 21 programından yararlanılmıştır ve betimsel istatistikler Tablo 3'te verilmiştir.

Tablo 3: Örneklemelere ilişkin cinsiyet değişkenine göre betimsel istatistikler

İstatistikler	N_1		N_2		N_3		N_4		N_5	
	Kadın	Erkek	Kadın	Erkek	Kadın	Erkek	Kadın	Erkek	Kadın	Erkek
Öğrenci Sayısı	150	150	500	500	1000	1000	100	500	200	1000
Ortalama	8,28	7,92	8,23	7,91	8,91	7,89	8,6	7,41	8,55	7,62
Standart Sapma	5,33	5,31	5,62	4,83	5,4	5,05	5,56	4,84	5,53	4,86
Çarpıklık	0,66	0,87	0,41	0,86	0,47	0,92	0,25	0,96	0,58	0,93
Basıklık	-0,87	-0,29	-0,92	0,69	-0,98	-0,27	-0,98	0,06	-0,97	-0,23

Tablo 3'te yer alan betimsel istatistikler göz önüne alındığında; erkekler ve kadınlara ait betimsel değerlerin birbirine yakın olduğu görülmektedir. Kadınların ortalamasının erkeklerin ortalamasından biraz fazla olduğu, çarpıklık ve basıklık değerlerine bakıldığında normal dağılım (-1,1) aralığında olduğu görülmektedir.

2.3.Araştırma Verileri ve Analizi

Araştırma verileri, MEB'nin uyguladığı Seviye Belirleme Sınavının 2009 8. Sınıf A kitapçığı fen ve teknoloji alt testine aittir. Fen ve teknoloji alt testinde 20 madde yer almaktadır.

Her bir maddeyi doğru cevaplayan öğrenciler 1 puan, yanlış veya boş bırakan öğrenciler 0 puan almışlardır. Verilerin DMF analizlerinde; MH analizi için EZDIF (Waller, 2005), SIBTEST analizi için DIFPACK (Shealy ve Stout, 1993), LR analizi için SAS ve MIMIC analizi için M-Plus 6.12 (Muthen ve Muthen, 1998) programından yararlanılmıştır. Çalışmada erkekler referans, kadınlar odak grup olarak ele alınmıştır. Testte yer alan 20 maddenin cinsiyet açısından yanlış olup olmadığına ilişkin uzman kanılarına başvurulmuştur. Çalışmaya 6 uzman katılmıştır. Bu uzmanlar fen bilgisi öğretmenliği mezunu olup ölçme ve değerlendirme alanında doktora öğrencisidir. Uzmanlara testte yer alan 20 maddenin DMF gösterip göstermediğine ilişkin herhangi bir bilgi verilmemiştir. Bu maddelerin cinsiyet açısından yanlışlık gösterip göstermediği sorulmuş ve yanlış olarak değerlendirmişlerse nedenine ilişkin görüşleri alınmıştır. Uzmanlara hazırlanan formlar verilmiş ve yüz yüze görüşmeler de yapılmıştır.

3. BULGULAR

3.1. DMF Analizlerine İlişkin Bulgular:

2009 SBS 8.sınıf fen ve teknoloji alt testinde yer alan 20 maddenin cinsiyet değişkenine göre değişen madde fonksiyonu gösterip göstermediği SIBTEST, MH, LR ve MIMIC yöntemlerine göre incelenmiş ve bu yöntemlerin uyumları karşılaştırılmıştır. SIBTEST, LR ve MH analizinde tüm örneklerde ihmal edilebilir düzeyde DMF veren 16. madde sabitlenerek MIMIC analizi yapılmıştır. Fen ve teknoloji alt testinde yer alan 20 maddenin cinsiyet değişkenine göre yapılan SIBTEST, MH, LR ve MIMIC analiz sonuçları Tablo 4, 5, 6, 7 ve 8'de verilmiştir.

Tablo 4: N₁ için DMF analiz sonuçları

Madde No	SIBTEST		M.H.		MIMIC		LR	
	Beta Estimate	P	ΔMH	p	Beta Estimate	p	χ ²	ΔR ²
1	-0,058	0,21	1,197	0,266	0,164	0,195	0,7213	0,0021
2	0,095	0,036*	-1,825	0,033*	0,192	0,013*	4,9745*	0,0395
3	0,036	0,494	-0,572	0,495	0,285	0,005*	1,1462	0,0003
4	0,021	0,666	-0,965	0,17	0,163	0,129	2,8065	0,0074
5	0,037	0,466	0,452	0,705	0,125	0,106	0,3047	0,0005
6	0,026	0,563	-0,737	0,33	-0,057	0,575	1,6096	0,0043
7	-0,038	0,451	-0,3	0,786	0,23	0,068	0,1561	0,0004
8	-0,044	0,316	-0,68	0,403	0,216	0,021*	1,2638	0,0031
9	-0,062	0,251	-1,336	0,774	0,043	0,635	0,1749	0,0004
10	0,038	0,908	0,399	0,686	0,121	0,208	0,1292	0,0003
11	-0,047	0,346	0,524	0,58	0,378	0,000*	0,5693	0,0011
12	-0,062	0,338	0,797	0,268	-0,188	0,079	1,4877	0,0035
13	0,065	0,849	-0,092	0,988	0,386	0,000*	0,0837	0,0002
14	-0,047	0,633	0,298	0,765	0,007	0,939	0,2394	0,0006
15	-0,05	0,746	0,387	0,656	0,018	0,818	0,3752	0,0009
16	-0,009	0,105	0,139	0,949	-	-	0,0197	0

17	0,051	0,59	-1,157	0,579	0,161	0,06	0,5232	0,001
18	-0,016	0,21	0,1	0,99	-0,014	0,893	0,0196	0,0001
19	-0,086	0,026*	1,36	0,047*	0,108	0,292	1,8855	0,0055
20	-0,025	0,494	0,303	0,79	0,092	0,324	0,1648	0,0004

Not: Referans grup: Erkek (n=150) Odak grup: Kadın (n=150) ; $p < .05$, $df=1$, $\chi^2=3,84$.

300 kişilik örneklem için yapılan DMF analizinde; SIBTEST ve MH yöntemlerine göre 2. ve 19. maddeler DMF içermektedir. SIBTEST yöntemine göre her iki madde yüksek düzeyde (C); M-H yöntemine göre ise 2. madde yüksek düzeyde (C), 19. madde orta düzeyde (B) DMF göstermiştir. Her iki modelde de 2. madde erkekler lehine, 19. madde ise kadınlar lehine işlemiştir. MIMIC yöntemine göre ise; 2., 3., 8., 11. ve 13. maddeler DMF göstermektedir. DMF gösteren maddelerin tamamı erkekler lehine işlemiştir. Lojistik regresyon yöntemine göre ise sadece 2. madde orta düzeyde (B) erkekler lehine DMF göstermiştir.

Tablo 5: N₄ için DMF analiz sonuçları

Madde No	SIBTEST		M.H.		MIMIC		LR	
	Beta Estimate	P	Δ -MH	P	Beta Estimate	p	χ^2	ΔR^2
1	-0,083	0,059	1,142	0,115	-0,179	0,21	1,0099	0,0016
2	0,012	0,997	0,036	0,92	0,02	0,863	0,0002	0
3	0,063	0,187	-0,804	0,312	0,187	0,135	2,0206	0,0027
4	0,043	0,454	-0,357	0,663	-0,115	0,332	0,1167	0,0001
5	-0,002	0,955	0,515	0,641	0,001	0,996	0,3274	0,0003
6	0,089	0,059	-1,647	0,029*	0,207	0,116	6,495*	0,0091
7	0,075	0,141	-1,278	0,103	0,106	0,383	4,020*	0,0045
8	-0,096	0,017*	-1,182	0,102	0,178	0,176	4,129*	0,052
9	-0,05	0,347	0,87	0,269	-0,153	0,221	1,5579	0,0017
10	0,014	0,801	-0,823	0,020*	0,308	0,022*	2,0048	0,0028
11	0,049	0,214	-0,276	0,837	0,034	0,776	0,2985	0,0003
12	-0,041	0,509	0,572	0,426	-0,083	0,54	0,7558	0,0009
13	-0,114	0,031*	1,502	0,017*	-0,24	0,042*	4,731*	0,073
14	-0,021	0,697	0,259	0,822	-0,01	0,932	0,0436	0
15	0,041	0,472	-0,417	0,636	-0,034	0,755	0,5278	0,0006
16	-0,016	0,761	-0,244	0,83	-	-	0,1854	0,0002
17	0,058	0,131	0,905	0,364	0,121	0,348	1,4299	0,0013
18	-0,108	0,046*	1159	0,091	-0,318	0,018*	3,6139	0,0044
19	-0,08	0,155	1039	0,101	-0,276	0,028*	1,9408	0,0029
20	-0,06	0,305	0,629	0,369	-0,184	0,145	0,6251	0,0007

Not: Referans grup: Erkek (n=500) Odak grup: Kadın (n=100) ; $p < .05$, $df=1$, $\chi^2=3,84$.

Tablo 5 incelendiğinde, SIBTEST yöntemine göre; 8., 13. ve 18. maddeler yüksek düzeyde (C) kadınlar lehine DMF içermektedir. M-H yöntemine göre; 6., 10. ve 13. maddeler

DMF vermektedir. 6. madde yüksek düzeyde (C) kadınlar lehine, 10. madde ihmal edilebilir düzeyde (A) kadınlar lehine ve 13. madde yüksek düzeyde (C) erkekler lehine işlemektedir. MIMIC yöntemine göre ise; 10., 13., 18. ve 19. maddeler ($p < .05$) DMF vermektedir. Bu maddelerden 10. madde erkekler lehine; 13., 18. ve 19. maddeler ise kadınlar lehine DMF göstermektedir. LR yöntemine göre 16 madde de DMF olmadığı, 6., 7., 8. ve 13. maddelerin ise önemsiz düzeyde (A) DMF verdiği görülmüştür. Bu maddelerden 6., 7. ve 8. maddeler erkekler, 13. madde ise kadınlar lehine işlemiştir.

Tablo 6: N₂ için DMF analiz sonuçları

Madde No	SIBTEST		M.H.		MIMIC		LR	
	Beta Estimate	P	Δ -MH	P	Beta Estimate	p	χ^2	ΔR^2
1	-0,122	0,000*	1,667	0,000*	-0,392	0,000*	16,125*	0,014
2	0,02	0,418	-0,19	0,705	0,074	0,355	0,4818	0,0003
3	0,033	0,224	-0,457	0,271	0,098	0,249	1,5417	0,0011
4	0,036	0,197	-0,176	0,664	0,03	0,711	0,172	0,0002
5	0,032	0,191	-0,365	0,43	0,071	0,376	1,1281	0,0006
6	0,006	0,83	-0,132	0,777	-0,032	0,691	0,273	0,0003
7	0,073	0,005*	-0,901	0,030*	0,141	0,078	5,804*	0,0036
8	0,038	0,172	-0,1405	0,314	0,171	0,043*	1,7204	0,0013
9	0,032	0,222	-0,454	0,282	0,085	0,289	1,3801	0,0009
10	-0,019	0,5	0,154	0,746	-0,064	0,443	0,0835	0
11	0,042	0,084	-0,541	0,237	0,112	0,154	1,9316	0,0011
12	-0,034	0,23	0,767	0,038*	-0,119	0,168	4,949*	0,0365
13	-0,02	0,481	0,008	0,954	0,026	0,743	0,002	0
14	-0,016	0,578	0,132	0,78	0,034	0,675	0,0545	0
15	0,035	0,19	-0,507	0,213	0,087	0,274	2,1247	0,0013
16	-0,011	0,685	0,206	0,657	-	-	0,1209	0,0001
17	0,008	0,758	-0,082	0,926	0,067	0,391	0,075	0
18	-0,031	0,275	0,548	0,149	-0,036	0,675	2,4682	0,0018
19	0,011	0,724	-0,361	0,338	0,084	0,301	1,4351	0,0013
20	-0,053	0,041*	0,804	0,042*	-0,114	0,163	3,8185	0,0024

Not: Referans grup: Erkek (n=500) Odak grup: Kadın (n=500) ; $p < .05$, $df=1$, $\chi^2=3.84$

Tablo 6'da 1000 kişilik örneklem için yapılan analizler incelendiğinde, SIBTEST yöntemine göre; 1. madde yüksek düzeyde (C), 20. madde ihmal edilebilir düzeyde (A) ve 7. madde orta düzeyde (B) DMF içermektedir. Bu maddelerden 1. ve 20. maddelerin kadınlar lehine, 7. maddenin ise erkekler lehine DMF içerdiği görülmüştür. M-H yöntemine göre; 1. madde yüksek düzeyde (C), 7., 12. ve 20. maddeler ihmal edilebilir düzeyde (A) DMF içermektedir. Bu maddelerden 1., 12. ve 20. maddeler kadınlar lehine; 7. madde ise erkekler lehine işlemiştir. MIMIC yöntemine göre ise; 1. ve 8. maddeler DMF vermektedir. DMF gösteren maddelerden 1. madde kadınlar lehine. 8. madde ise erkekler lehine işlemiştir. LR yöntemine göre ise 1. madde önemsiz düzeyde (A) kadınlar lehine, 7. madde önemsiz düzeyde (A) erkekler lehine ve 12. madde orta düzeyde (B) kadınlar lehine DMF vermiştir.

Tablo 7:N₅ için DMF analiz sonuçları

Madde No	SIBTEST		M.H.		MIMIC		LR	
	Beta Estimate	P	Δ -MH	P	Beta Estimate	p	χ^2	ΔR^2
1	-0,011	0,751	0,377	0,471	-0,034	0,743	0,3607	0,0003
2	0,017	0,582	-0,082	0,952	0,086	0,426	0,0041	0
3	0,031	0,27	-0,254	0,691	0,071	0,49	0,2855	0,0001
4	-0,012	0,755	0,402	0,383	0,007	0,95	1,9318	0,0013
5	0,009	0,749	0,039	0,974	0,092	0,349	0,0622	0
6	-0,013	0,71	0,2	0,719	-0,016	0,873	0,0802	0
7	0,065	0,046*	-0,781	0,129	0,191	0,063	2,4797	0,0014
8	0,001	0,982	-0,083	0,919	0,123	0,235	0,2049	0
9	0,026	0,444	-0,407	0,457	0,168	0,117	0,6274	0,0003
10	0,032	0,359	-0,405	0,446	0,136	0,194	0,831	0,0006
11	0,042	0,167	-0,725	0,247	0,187	0,072	1,1894	0,0006
12	-0,036	0,343	0,523	0,271	-0,029	0,786	2,1364	0,0013
13	0,05	0,102	-0,659	0,192	0,314	0,003*	2,9113	0,0022
14	-0,017	0,646	0,234	0,673	0,031	0,744	0,2357	0
15	-0,072	0,048*	0,911	0,049*	-0,097	0,319	5,113*	0,0528
16	-0,008	0,83	0,179	0,775	-	-	0,1472	0
17	0,051	0,06	-0,866	0,162	0,107	0,288	2,0715	0,001
18	-0,049	0,17	0,674	0,161	-0,065	0,543	3,2153	0,002
19	0,083	0,014*	1,004	0,025*	0,322	0,001*	6,077*	0,0044
20	-0,035	0,302	0,813	0,09	-0,018	0,865	2,8002	0,0017

Not: Referans grup: Erkek (n=1000) Odak grup: Kadın (n=200) ; $p < .05$, $df=1$, $\chi^2=3.84$.

Tablo 7’de görüldüğü gibi, SIBTEST yöntemine göre; 7., 15. ve 19. maddeler orta düzeyde (B) DMF vermektedir. DMF veren maddelerden 7. ve 19. maddeler erkekler lehine; 15. madde ise kadınlar lehine işlemektedir. M–H yöntemine göre; 15. madde ihmal edilebilir düzeyde (A) ve 19. madde orta düzeyde (B) kadınlar lehine DMF içermektedir. MIMIC yöntemine göre ise; 13. ve 19. maddeler DMF vermektedir. Bu maddelerden her ikisi de erkekler lehine DMF göstermektedirler. LR yöntemine göre 15. madde orta düzeyde (B) kadınlar lehine ve 19. madde ihmal edilir düzeyde (A) erkekler lehine DMF içermektedir.

Tablo 8: N₃ için DMF analiz sonuçları

Madde No	SIBTEST		M.H.		MIMIC		LR	
	Beta Estimate	P	Δ-MH	P	Beta Estimate	P	χ ²	ΔR ²
1	-0,018	0,324	0,127	0,686	0,025	0,686	0,0483	0,0001
2	-0,04	0,029*	0,655	0,020*	-0,133	0,032*	3,912*	0,0012
3	-0,067	0,013*	0,684	0,012*	-0,14	0,024*	3,851*	0,044
4	-0,09	0,000*	1,39	0,000*	-0,352	0,000*	31,223*	0,0128
5	0,028	0,124	-0,453	0,119	0,025	0,68	4,14*	0,0013
6	0,013	0,5	-0,196	0,488	0,052	0,376	2,091	0,0008
7	-0,016	0,382	0,301	0,301	-0,081	0,185	0,4787	0,0001
8	0,041	0,031*	-0,61	0,029*	0,106	0,087	8,056*	0,003
9	0,003	0,859	0,018	0,995	-0,057	0,356	0,0762	0,0001
10	0,03	0,132	-0,392	0,146	0,068	0,262	4,569*	0,0018
11	0,029	0,086	-0,413	0,184	0,009	0,881	3,7549	0,0011
12	-0,059	0,003*	0,823	0,001*	-0,198	0,001*	10,565*	0,0039
13	0,014	0,449	-0,361	0,197	0,108	0,076	2,4325	0,0011
14	0,043	0,027*	-0,582	0,032*	0,105	0,072	6,655*	0,0022
15	0,014	0,444	-0,227	0,44	-0,003	0,955	1,4529	0,0004
16	-0,003	0,869	0,067	0,853	-	-	0,0823	0,0001
17	0,016	0,324	-0,262	0,441	-0,023	0,706	1,5285	0,0004
18	-0,062	0,001*	0,813	0,002*	-0,143	0,021*	9,631*	0,0034
19	0,032	0,121	-0,469	0,064	0,056	0,336	5,302*	0,0423
20	0,023	0,035*	-0,347	0,2	0,05	0,409	3,3081	0,0011

Not: Referans grup: Erkek (n=1000) Odak grup: Kadın (n=1000) ; p < .05, df=1, χ² = 3.84.

Tablo 8’de 2000 örneklem için yapılan analizlere bakıldığında, SIBTEST yöntemine göre; 8. madde DMF içermektedir. Bu maddelerden 2., 8., 14. ve 20. maddeler ihmal edilir düzeyde (A); 3., 4., 12. ve 18. maddeler orta düzeyde (B) DMF içerdiği görülmektedir. DMF veren maddelerden 8., 14. ve 20. maddeler erkekler lehine, 2., 3. ve 4. maddeler kadınlar lehine işlemektedir. M-H yöntemine göre; 2., 3., 8., 12., 14., 18. ve 20. maddeler ihmal edilir düzeyde, 4. madde ise orta düzeyde (B) (kadınlar lehine) DMF göstermiştir. MIMIC yöntemine göre ise; 2., 3., 4., 12. ve 18. maddeler (p < .05) DMF vermektedir. Bu maddelerden hepsi kadınlar lehine DMF göstermektedir. LR’a göre; 2, 3, 4, 5, 8, 10, 14, 18 ve 19. maddeler önemsiz düzeyde (A) DMF göstermiştir. Bu maddelerden 2., 3., 4. ve 18. maddeler kadınlar lehine, 5., 8., 10., 14. ve 19. maddeler erkekler lehine işlemiştir.

3.2. Madde Yanlılığına İlişkin Bulgular

20 maddenin cinsiyet açısından yanlı olup olmadığı sorulan 6 uzmandan biri testte cinsiyet açısından yanlı madde olmadığını belirtmiştir. Diğer 5 uzmandan en az biri aşağıda belirtilen 4 maddenin yanlı olduğunu belirtmişlerdir (Belirtilen bu maddelere EK-1’de yer verilmiştir):

2. madde hakkındaki beş uzman da maddenin erkekler lehine yanlı olduğunu, şıklarda geçen traktör, tren, tekerlek gibi kelimelerin erkeklerin daha fazla ilgisini çekebileceğini ve kadınların bu konuda daha az bilgiye sahip olabileceğini,

5. madde için üç uzman maddenin yanlı olmadığını belirtirken, diğer iki uzman maddenin erkekler lehine yanlılık gösterdiğini ifade etmişlerdir. Elektrik, mekanik gibi fizik konularının erkeklerin günlük hayatta da ilgilerini çektiğini, hem de soruda şekil kullanıldığından, bu sorunun erkekler lehine yanlılık gösterdiğini,

6. madde için beş uzman maddenin yanlı olmadığını belirtirken, sadece bir uzman erkeklerin elektrik, mekanik gibi fizik konuları erkeklerin günlük hayatta da ilgilerini çektiğini, bu sorunun erkeklerin lehine işlemesine sebep alabileceğini,

12. madde için üç uzman maddenin yanlı olmadığını belirtirken, iki uzman maddenin kadınlar lehine yanlı olduğunu, şıklarda yer alan ifadelerin kız öğrencilerin günlük hayatta karşılaşma olasılıklarının yüksek olmasından dolayı, erkeklere göre maddeyi daha yüksek oranda doğru cevaplayacaklarını ifade etmişlerdir.

4. TARTIŞMA, SONUÇ ve ÖNERİLER

Bu araştırmada 2009 8. sınıf SBS fen ve teknoloji alt testinde yer alan maddelerin cinsiyet değişkenine göre DMF gösterip göstermediği çeşitli yöntemlerle incelenmiştir. Yapılan analizler sonucunda önemsiz düzeyde DMF gösteren maddeler göz ardı edilerek orta ve yüksek düzeyde DMF gösteren maddeler yorumlanmıştır.

SIBTEST yöntemine göre; 300 kişiden oluşan örneklem için, 2. ve 19. maddeler, 600 kişiden oluşan örneklem için 8., 13. ve 18. maddeler; 1000 kişiden oluşan örneklem için 1. ve 7. maddeler; 1200 kişiden oluşan örneklem için 7., 15. ve 19. maddeler ve 2000 kişilik örneklem için 3., 4., 12. ve 18. maddeler DMF içermektedir. SIBTEST analizlerinde örneklem sayısı arttıkça DMF gösteren madde sayısının arttığı söylenebilir. Gierl, Jodoin ve Ackerman (2000) yaptıkları çalışmanın sonucu olarak örneklem büyüklüğü arttıkça SIBTEST yönteminin gücünün arttığından bahsetmişlerdir. Örneklem büyüklüğü arttıkça SIBTEST yönteminin gücünün artması, DMF içeren maddelerden SIBTEST yöntemiyle tespit edilen madde sayısının arttığı anlamına gelmektedir. Bu çalışmada da örneklem büyüklüğü arttıkça SIBTEST yönteminin belirlediği DMF'li madde sayısı artış göstermiştir. Bu artış SIBTEST yönteminin DMF belirleme gücünün artması şeklinde yorumlanabilir. Bu bakımdan iki çalışmanın uyumundan söz edilebilir. Araştırmada SIBTEST yöntemine göre belirlenen 11 madde orta veya yüksek düzeyde DMF vermiştir. Ancak farklı örneklem büyüklüklerinde DMF içeren maddelerin farklılaşması ise dikkat çekici bir sonuçtur.

Mantel-Haenszel yöntemine göre; 300 kişiden oluşan örneklem için, 2. ve 19. maddeler, 600 kişiden oluşan örneklem için 6. ve 13. maddeler; 1000 kişiden oluşan örneklem için 1. madde; 1200 kişiden oluşan örneklem için 19. madde ve 2000 kişilik örneklem için 4. madde DMF içermektedir. SIBTEST ve MH yöntemlerine göre DMF içeren maddelere baktığımızda genel olarak iki yöntemde de DMF içeren ortak maddeler gözlenmektedir. Atalay, Gök, Kelecioğlu ve Arsan (2012) SIBTEST, MTK-OO, MH ve LR yöntemlerini karşılaştırdıkları bir simülasyon çalışmasında, MH yönteminin diğer yöntemlerle uyum yüzdesinin genel olarak düşük olduğundan bahsetmişlerdir. Bu bakımdan söz konusu iki çalışmanın uyumlu olmadığından bahsedilebilir.

MIMIC yöntemine göre; 300 kişiden oluşan örneklem için, 2., 3., 8., 11. ve 13. maddeler, 600 kişiden oluşan örneklem için 10., 13., 18. ve 19. maddeler; 1000 kişiden oluşan örneklem için 1. ve 8. maddeler; 1200 kişiden oluşan örneklem için 13. ve 19. maddeler ve 2000 kişilik

örneklem için 2., 3., 4., 12. ve 18. maddeler DMF içermektedir. Finch (2005) yaptığı araştırmada MIMIC DFA modelini geleneksel DMF belirleme yöntemlerinden Mantel-Haenszel, SIBTEST ve MTK olabilirlik oranı ile karşılaştırmıştır. Bahsedilen araştırmada örneklem büyüklüğü arttıkça MIMIC yönteminin gücünün arttığı belirtilmiştir. Bu çalışmada 2000 kişilik büyük örneklemde, belirlenen DMF'li maddeler açısından MIMIC yönteminin diğer yöntemlerle daha uyumlu hale geldiği söylenebilir. Bu bakımdan iki çalışma birbirini destekler niteliktedir.

LR yöntemine göre ise; 300 kişiden oluşan örneklem için, 2. madde; 1000 kişiden oluşan örneklem için 12. madde; 1200 kişiden oluşan örneklem için 15. maddeler DMF içermektedir. Bekçi (2007) "Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı'nın Değişen Madde Fonksiyonlarının Cinsiyete ve Okul Türüne Göre İncelenmesi" isimli araştırmasında LR ve MH yöntemlerinin alt testler ve grup değişkenine göre farklı sonuçlar verebildiğini gözlemlemiştir. Bu bakımdan MH ve LR yöntemleri arasında tam bir uyumdan söz edilememiştir. Bu çalışmada da tam bir uyumdan bahsedilemez. Böylece iki çalışmanın bulgularının benzerlik gösterdiği söylenebilir. Benzer şekilde Doğan ve Öğretmen (2008) yaptıkları araştırmanın sonucunda MH ve LR yöntemlerinin tam bir uyum göstermediğinden bahsetmiştir. Doğan ve Öğretmen'in sonuçları bu araştırmanın sonuçları ile benzerlik göstermektedir. Finch ve French (2007); LR, SIBTEST, MTK-OO ve DFA yöntemleriyle yürüttükleri çalışmanın sonucu olarak örneklem büyüklüğü arttıkça LR, SIBTEST ve MTK-OO yöntemlerinin güçlerinde artış gözlemlediklerini ancak DFA yönteminin gücünün bu değişkenden etkilenmediğinden bahsetmişlerdir. Bu çalışmada örneklem büyüklüğü arttıkça SIBTEST yöntemiyle belirlenen DMF içeren madde sayısının artarken, LR yöntemi için böyle bir uyum söz konusu değildir. MIMIC DFA yöntemi ise örneklem büyüklüklerine göre belirlenen DMF içeren madde sayısı açısından değişkenlik göstermiştir. Bu bakımdan iki çalışmanın sadece SIBTEST yöntemi açısından uyumlu olduğundan bahsedilebilir.

Dört ayrı yöntem için örneklemelere göre DMF içeren ortak maddelere baktığımızda sadece 300 kişilik örneklemde 2. maddenin ortak şekilde DMF içerdiği görülmüştür. Ayrıca en çok DMF içeren madde sayısı SIBTEST ve MIMIC yöntemleriyle belirlenmiştir.

Uzman kanılarına göre yanlı olduğu belirlenen 2., 5., 6. ve 12. maddeler DMF analizlerinin sonuçlarıyla karşılaştırıldığında bu maddelerin farklı yöntem ve ya farklı örneklemelerde DMF gösterdiği görülmüştür. Ancak bu maddelerden 2., 6. ve 12. maddeler farklı yöntemlerde orta ve yüksek düzeyde DMF gösterirken, 5. madde önemsiz düzeyde DMF göstermiştir. Böylece uzman kanılarıyla analiz sonuçlarının birbirleriyle tutarlı olduğu söylenebilir.

Bu çalışmanın sonucu olarak, çok sayıda bireyin bulunduğu verilerde DMF analizi yaparken verinin tamamını kullanmak yerine çalışmanın referans ve odak grupları için belli büyüklükte örneklem oluşturmak daha uygun olacaktır. Ayrıca elde edilen sonuçlar bir arada değerlendirildiğinde birden fazla DMF belirleme tekniğinin bir arada kullanılması daha doğru sonuçlar elde etmek açısından gerekli görülmüştür. Atalay, Gök, Kelecioğlu ve Arsan (2012) benzer şekilde çalışmalarında DMF belirlemede birden fazla yöntemin bir arada kullanılmasını önermişlerdir.

Bu çalışmada grup değişkeni olarak cinsiyet değişkeni ele alınmıştır. Bireylerin performanslarını etkileyecek okul türü, bölge, sosyoekonomik statü vb. değişkenlerle de çeşitli araştırmalar yürütülebilir.

5. KAYNAKLAR

- Atalay, K., Gök, B., Kelecioğlu, H. & Arsan, N. (2012). Değişen madde fonksiyonunun belirlenmesinde kullanılan farklı yöntemlerin karşılaştırılması: Bir simülasyon çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 43, 270- 281.
- Bekçi, B. (2007). *Orta öğretim kurumları öğrenci seçme ve yerleştirme sınavının değişen madde fonksiyonlarının cinsiyete ve okul türüne göre incelenmesi* (Yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.
- Calvert, T. (2001). *Exploring differential item functioning (DIF) with the Rasch Model: A cross-country comparison of gender differences on eight grade science items*. Seattle,WA: AERA.
- Camili, G., & Shephard, L. A. (1994). *Methods for identifying biased test items*. London: Sage.
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17(1), 31- 44.
- Doğan, N., & Öğretmen, T. (2008). Değişen madde fonksiyonunu belirlemede Mantel-Haenszel, Ki-Kare ve Lojistik Regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim*, 33(148),100-112.
- Dogan, E., Guerrero, A., & Tatsuoka, K. (2005, April). *Using DIF to investigate strengths and weaknesses in mathematics achievement profiles of 10 different countries*. Paper presented at the Annual Meeting of the National Council on Measurement in Education (NCME), Montreal, Canada. Abstract retrieved from http://cms.tc.columbia.edu/i/a/1688_ncme-05-enis.pdf
- Dorans, N. J., & Holland, P. W. (1993). *DIF detection and description: Mantel-Haenszel and standardization* (ETS-RR-92-10). Retrieved from <http://files.eric.ed.gov/fulltext/ED387526.pdf>
- Finch, W. H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel- Haenszel, SIBTEST and the IRT likelihood ratio. *Applied Psychological Measurement*, 29, 278-295.
- Finch, W. H., & French, F. B. (2007). Detection of crossing differential item functioning: A comparison of four methods. *Educational and Psychological Measurement*, 67(4), 565-582.
- Finch, W. H., & French, B. F. (2011). Estimation of MIMIC model parameters with multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(2), 229-252.
- Gierl, M. J., Jodoin, M., & Ackerman, T. (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, New Orleans, Louisiana, USA.
- Gierl, M. J., Rogers, W. T., & Klinger, D. (1999). *Consistency between statistical procedures and content reviews for identifying translation DIF*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montréal, Quebec, Canada.
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the Logistic Regression procedure for DIF detection. *Applied Measurement in Education*, 14, 329-349.
- MacIntosh, R., & Hashim, S. (2003). Variance estimation for converting MIMIC model parameters to IRT parameters in DIF analysis. *Applied Psychological Measurement*, 27(5), 372-379.
- Roussos, L. A. & Stout, W. F. (1996). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel type I error performance. *Journal of Educational Measurement*, 33, 215-230.
- Shih, C.-L., & Wang, W.-C. (2009). Differential item functioning detection using the multiple indicators, multiple causes method with a pure short anchor. *Applied Psychological Measurement*, 33, 184-199.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*, 91(6), 1292-1306.
- Wang, W.-C., Shih, C.-L., & Yang, C.-C. (2009). The MIMIC method with scale purification for detecting differential item functioning. *Educational and Psychological Measurement*, 69, 713-731.
- Woods, C. M. (2009). Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research*, 44, 1-27.
- Woods, C. M., Oltmanns, T. F., & Turkheimer, E. (2009). Illustration of MIMIC-model DIF testing with the schedule for nonadaptive and adaptive personality. *J Psychopathol Behav Assess*, 31(4), 320-330.

- Zheng, Y., Gierl, M. J., & Cui, Y. (2007). *Using real data to compare DMF detection and effect size measures among Mantel-Haenszel, SIBTEST, and Logistic Regression procedures*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, USA.
- Zieky, M. (2006). Fairness reviews in assessment. In S. M. Downing, & T. M. Haladyna (Eds.), *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- 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 DMF Analysis: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4(2), 223-233.

Extended Abstract

Items that are prepared for tests that help make important decisions about individuals, should have the ability to distinguish the differences between individuals and must make more accurate measurements. Items providing advantage or disadvantage to any subgroup, in other words items that have bias, threaten valid interpretation and usage of test results. *Bias* is defined as the probability of answering an item correctly being higher for individuals in a group with respect to individuals in another group (Zumbo, 1999). Items with DIF decreases the validity of test scores since it unnecessarily changes the variance of test scores (Zieky, 2006). DIF analyses seem to be useful for determining weak and strong sides of abilities of a subgroup (Calvert, 2001). DIF appears in two forms: *uniform* and *non-uniform*. Uniform DIF emerges when there is no interaction between the ability level and group membership; that is, the difference in probability of answering the item correctly is constant for all ability levels. On the other hand, non-uniform DIF is the probability of answering the item correctly of two groups at the same ability level being not the same for all ability levels (Gierl, Rogers and Klinger, 1999).

In this study, it is investigated whether the items in 8th grade science and technology subtest of 2009 placement test (SBS) show DIF or not, with respect to sex, in analyses conducted using SIBTEST, MH, MIMIC and LR methods with different samples, whether the obtained results are compatible with each other and whether the items that are determined, in the light of expert opinions, to have bias and the items that show DIF as the result of statistical analyses are consistent with each other or not. In the aspect of the usage of distinct samples and different techniques that have never been used together before in the research, besides the questioning of all items by experts if they are biased in terms of sex or not, it is believed that the results of this research will make a contribution to new studies. The universe of the research consists of about 340.000 8th grade students that attended to the SBS placement test held in 2009 by the Ministry of National Education (MEB). Samples of the research are chosen randomly, considering sex of the students. From the data set of about 340.000 students, 5 sub sample spaces are chosen; $N_1 = 300$ students (reference group:150, focal group:150), $N_2 = 1000$ students (reference group:500, focal group:500), $N_3 = 2000$ students (reference group:1000, focal group:1000), $N_4 = 600$ students (reference group:300, focal group:300), and $N_5 = 1200$ students (reference group:600, focal group:600). Simulation conditions in the study of Finch (2005) are considered in selection of sample numbers. In DIF analyses of the data, the following computer programs are used; EZDIF (Waller, 2005) for MH analysis, DIFPACK (Shealy and Scout, 1993) for SIBTEST analysis, SAS for LR analysis, and M-Plus 6.12 (Muthen and Muthen, 1998) for MIMIC analysis. In the study, male students are considered as reference, and female students are considered as focal group. Furthermore, experts were consulted for their opinions about the 20 items in the test if they are biased in terms of sex or not. Six experts attended to the research. These experts have graduated from science education and are PhD students in measurement and evaluation area. They were not preinformed whether the 20 items in the test show DIF or not. They were asked if these items are biased in terms of sex or not, and also asked for their opinions related to the reason if they evaluated the items as biased.

According to the results, when we look at the items which are commonly determined to have DIF by the four methods; it is seen that 2nd item in 300-student sample space, 13th item in 600-student sample space, 19th item in 1200-student sample space, and 2nd, 3rd and 4th items in 2000-student sample space have DIF, while there is no item with DIF in 1000-student sample space. In this respect, it can be said that

the methods yield in more compatible results in the largest sample space (2000 students). There is an increase in the number of items having DIF that are determined by SIBTEST method as the number of samples increases. 11 of 18 items, which show DIF according to SIBTEST method, have a medium or high DIF level. Differentiation of items with DIF, as the number of samples increases, is an interesting result. It can be said that items that have DIF according to SIBTEST and MH methods show parallelism. In the largest sample space having 2000 students, MIMIC method can be said to be more compatible with other methods in determining items that have DIF. It is seen that the maximum number of items which are determined by LR method to have DIF is reached in the 2000-student sample space. In addition, the change in the number of individuals in reference or focal group also affects the number of items having DIF.

Among 6 experts, whom were asked if any of the 20 items is biased with respect to sex or not, one of them stated that there is no biased item. Other 5 experts stated that some items are biased in the test. When the items (2, 5, 6 and 12th items) that are determined to be biased by the expert opinions are compared with the results of statistical analyses, it is seen that those items show DIF using different methods or in different sample spaces. Thus, it can be said that expert opinions and results of analyses are consistent with each other. Even though the level of compatibility between the methods is better in large sample spaces, no limit is determined for MIMIC method yet. So, while making a DIF analysis with data containing many individuals, it is better to form sample spaces with particular sizes for reference and focal groups instead of using all the data. Besides, when the obtained results are evaluated all together, usage of more than one DIF determination methods is considered as necessary in terms of obtaining more accurate results.

EK-1

Not: Sayfa sayısı nedeniyle sadece uzmanlar tarafından yanlı olduğu belirtilen maddeler verilmiştir.

Sayın Katılımcı,

Aşağıda, 2009 SBS fen ve teknoloji alt testine ait sorular ile ilgili görüşünüze başvurulacaktır. Bu sorulara yönelik cinsiyet açısından öğrencilerin bu sorulara yönelik performansları **DMF** analizleri ile incelenmektedir.

Maddelerin hangi özelliklerinin DMF'ye yol açtığının belirlenmesi için uzman görüşlerine ihtiyaç duyulmaktadır. Sizden istenen aşağıda verilen maddelerin yanlılık gösterip göstermediğini ve yanlı ise neden kaynaklanmış olabileceğine dair uzman görüşünüzü belirtmenizdir.

Vereceğiniz cevaplar sadece “uzman görüşü” kapsamında, belirtilen çalışmada kullanılacaktır.

2. Aşağıdakilerden hangisi basıncı arttırmak için yapılmıştır?

- A) Rayların şekil bozukluğuna uğramaması için trenlerde tekerlek sayısının artırılması
- B) Meyvenin daha rahat kesilmesi için bıçağın keskinleştirilmesi
- C) Karda daha rahat yürümek için kar ayakba-bısı kullanılması
- D) Traktörlerin toprağa saplanmaması için geniş tekerlekli yapılması

Yukarıdaki maddenin cinsiyet açısından (erkek veya kadın öğrencilerin) yanlılık gösterip göstermediği, yanlı ise neden kaynaklanmış olabileceğini belirtiniz.

Erkek () Kadın ()

Kesinlikle Katılıyorum ()

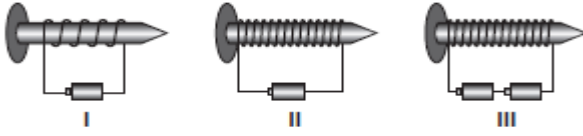
Katılıyorum ()

Katılmıyorum ()

Cevabınızın gerekçesini belirtiniz.

.....

- 5. 1. hipotez: Bobindeki sarım sayısı arttıkça elektromıknatısın çekim gücü artar.**
2. hipotez: Üzerinden geçen akım şiddeti arttıkça elektromıknatısın çekim gücü artar.
 Bir öğrenci yukarıdaki hipotezleri için özdeş çivi, tel ve pillerle I, II ve III elektromıknatıslarını yapıyor.



Daha sonra her bir elektromıknatısı özdeş iğnelere yaklaştırarak kaçar tane iğne çektiklerini kaydediyor.

Buna göre, öğrenci 1. ve 2. hipotezlerini test etmek için hangi elektromıknatısları ile elde ettiği verileri birlikte değerlendirmelidir?

- | <u>1. hipotez</u> | <u>2. hipotez</u> |
|-------------------|-------------------|
| A) I - II | II - III |
| B) II - III | I - II |
| C) I - III | II - III |
| D) I - II | I - III |

Yukarıdaki maddenin cinsiyet açısından (erkek veya kadın öğrencilerin) yanlılık gösterip göstermediği, yanlı ise neden kaynaklanmış olabileceğini belirtiniz.

Erkek () Kadın ()

Kesinlikle Katılıyorum ()

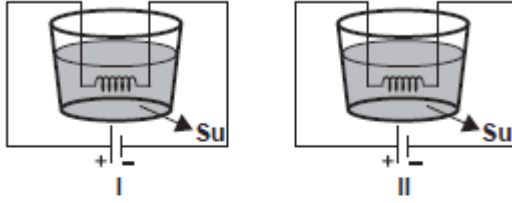
Katılıyorum ()

Katılmıyorum ()

Cevabınızın gerekçesini belirtiniz.

.....

6. Fatih, birbiriyle her bakımdan özdeş aşağıdaki iki devreyi kuruyor.



Fatih, bu iki devreyi kullanarak üzerinden akım geçen bir telde açığa çıkan ısı miktarı ile ilgili olarak aşağıdaki sonuçlardan hangisine ulaşabilir?

- A) Telin cinsine bağlıdır.
- B) Telin direncine bağlıdır.
- C) Telin üzerinden geçen akımın şiddetine bağlıdır.
- D) Telin üzerinden geçen akımın geçiş süresine bağlıdır.

Yukarıdaki maddenin cinsiyet açısından (erkek veya kadın öğrencilerin) yanlılık gösterip göstermediği, yanlı ise neden kaynaklanmış olabileceğini belirtiniz.

Erkek () Kadın ()

Kesinlikle Katılıyorum ()

Katılıyorum ()

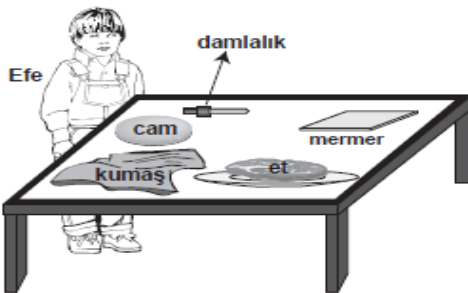
Katılmıyorum ()

Cevabınızın gerekçesini belirtiniz.

.....

.....

12.



Efe, masadaki maddelerin üzerine asit damlattıktan sonra etin, mermerin ve kumaşın tahriş olduğunu, camın ise tahriş olmadığını gözlemliyor. Efe'nin etkinlik sonucu edindiği bilgiye göre, aşağıdakilerden hangisini yapması uygun olmaz?

- A) Sirkeyi cam şişede saklaması
- B) Mermer tezgahın üzerinde limon kesmesi
- C) Laboratuvarda çalışırken koruyucu kıyafet kullanması
- D) Tuz ruhu ile banyoyu temizlerken koruyucu eldiven kullanması

Yukarıdaki maddenin cinsiyet açısından erkek veya kadın öğrencilerin lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.

Erkek () Kadın ()

Kesinlikle Katılıyorum ()

Katılıyorum ()

Katılmıyorum ()

Cevabınızın gerekçesini belirtiniz.

.....
.....

Kaynakça Bilgisi

Akın Arıkan, Ç.; Uğurlu, S. & Atar, B. (2016). MIMIC, SIBTEST, lojistik regresyon ve Mantel-Haenszel yöntemleriyle gerçekleştirilen dmf ve yanlılık çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi [Hacettepe University Journal of Education]*, 31(1), 34-52.

Citation Information

Akın Arıkan, Ç.; Uğurlu, S. & Atar, B. (2016). A DIF and bias study by using MIMIC, SIBTEST, logistic regression and Mantel-Haenszel methods [in Turkish]. *Hacettepe University Journal of Education [Hacettepe Üniversitesi Eğitim Fakültesi Dergisi]*, 31(1), 34-52.