



DEĞİŞEN MADDE FONKSİYONUNUN BELİRLEMESİNDE KULLANILAN FARKLI YÖNTEMLERİN KARŞILAŞTIRILMASI: BİR SİMÜLASYON ÇALIŞMASI

COMPARING DIFFERENT DIFFERENTIAL ITEM FUNCTIONING METHODS: A SIMULATION STUDY

Kübra ATALAY* Bilge GÖK** Hülya KELECİOĞLU*** Nihan ARSAN****

ÖZET: Bu araştırmanın amacı değişen madde fonksiyonunun (DMF) belirlenmesinde kullanılan gözlenen puan yöntemlerinden Mantel-Haenszel ve lojistik regresyon yöntemleri ile örtük puan yöntemlerinden MTK-OO ve SIBTEST yöntemlerinin karşılaştırılmasıdır. Yöntemlerin karşılaştırılması simülasyon çalışmasıyla yapılmıştır. Simülasyon koşulları örneklem büyüklüğü, yetenek dağılımı ve testteki DMF'li madde oranıdır. Araştırmada kullanılan örtük puan yöntemlerinin gözlenen puan yöntemlerine göre DMF'li maddeleri belirlemede daha duyarlı ve etkili olduğu görülmüştür. Araştırmada, LR'nin diğer yöntemlere göre DMF belirleme oranının daha düşük olduğu, MTK-OO'nun ise daha yüksek olduğu sonucuna varılmıştır. Tek biçimli DMF'yi belirlemede MH, SIBTEST ve MTK-OO yöntemleri; tek biçimli olmayan DMF'yi belirlemede ise LR, SIBTEST ve MTK-OO yöntemleri birbiriyle uyumlu sonuçlar vermiştir.

Anahtar kelimeler: Değişen madde fonksiyonu, Mantel-Haenszel, lojistik regresyon, SIBTEST, olabilirlik oranı.

ABSTRACT: The primary purpose of this research is to compare and evaluate the effectiveness of observed score methods -Mantel-Haenszel, logistic regression- and latent score methods -IRT-LR, SIBTEST- which used to determine DIF under variety conditions. These methods were compared by simulation study. Sample sizes, ability distribution, proportion of items with DIF were considered for data simulation conditions. Results of this research revealed that latent score methods were more sensitive and effective in determining items with DIF rather than observed score methods. Latent score methods were more liberal and observed score methods were more conservative in identifying items with DIF. As a result, MH, SIBTEST and IRT-LR methods present consistent result in determining uniform DIF in all conditions. Furthermore, consistent results were found in identifying non-uniform DIF with LR, SIBTEST and IRT-LR methods.

Keywords: Differential item functioning, Mantel-Haenszel, logistic regression, SIBTEST, likelihood ratio.

1. GİRİŞ

Yanlılık, bir testin psikometrik özellikleri incelenirken ele alınan önemli konulardan biridir. Madde yanlılığı, testin koşullarından ya da maddenin bazı özelliklerinden dolayı testi alan aynı yetenek düzeyindeki alt gruplarda maddenin doğru cevaplandırılma olasılığının farklı olmasıdır (Zumbo, 1999). Aynı yetenek düzeyine sahip olan bireyler için yanlı bir maddenin doğru cevaplandırılma olasılığı, bireylerin testle ölçülmeyen bazı özelliklerine göre değişebilmektedir. Yanlı bir madde, testle ölçülen özellik dışındaki bazı özelliklere sahip olan bireylere avantaj sağlarken, diğerlerine de dezavantaj sağlamaktadır. Bu durum testin geçerliğini büyük ölçüde zedelemekte ve testle verilen kararların hatalı olmasına yol açmaktadır.

Madde yanlılığını belirleme çalışması, istatistiksel bir süreçle başlar. Bir maddenin aynı yetenek düzeyinde ve farklı gruplardaki doğru cevaplandırılma olasılığı incelenerek alt gruplardaki fonksiyonlarının karşılaştırıldığı değişen madde fonksiyonu (DMF) çalışmaları, madde yanlılığı belirleme sürecinin ilk aşamasıdır. Bu aşamada DMF belirlenen maddelerin yanlı olup olmadığına karar vermek için farklılığın kaynağı araştırılır. DMF gösteren bir maddenin yanlı olup olmadığına, alt gruplar, testin yapısı veya kapsamı dikkate alınarak yapılan inceleme sonunda karar verilir (Dorans ve Holland, 1993).

DMF tek biçimli ve tek biçimli olmayan fonksiyonlar olarak ortaya çıkabilir. Tek biçimli DMF yetenek düzeyi ile grup üyeliği arasında madde performansında etkileşim olmadığı durumlarda ortaya çıkar. Madde performansı doğru cevap olasılığı kullanılarak tanımlandığında, tek biçimli DMF olasılık oranının yetenek dağılımı boyunca sabit olduğu durumlarda ortaya çıkmakla birlikte, tek biçimli olmayan

* Arş. Gör., Hacettepe Üniversitesi, katalay@hacettepe.edu.tr.

** Arş. Gör., Hacettepe Üniversitesi, bilgeb@hacettepe.edu.tr.

*** Doç. Dr., Hacettepe Üniversitesi, hulyaebb@hacettepe.edu.tr.

**** Arş. Gör., Hacettepe Üniversitesi, nihanarsan@hacettepe.edu.tr.

DMF ise olasılık oranının yetenek dağılımı boyunca değiştiği durumlarda ortaya çıkar. Tek biçimli olmayan DMF'nin varlığı yetenek düzeyi ile grup üyeliği arasında madde performansında etkileşim olduğu durumlarda oluşur (Camili ve Shephard, 1994, Penfield, 2003a, Güler ve Penfield, 2009). Bir başka deyişle, aynı yetenek düzeyindeki iki grubun bir maddeyi doğru cevaplandırma olasılığındaki fark sabit ise bu durum tek biçimli DMF, grupların doğru cevaplandırma olasılıkları arasındaki fark sabit değilse, tek biçimli olmayan DMF görülür.

DMF'yi değerlendirmek amacı ile kullanılan yöntemler araştırmacılar tarafından farklı başlıklar altında sınıflandırmaktadır. Bu sınıflamalardan biri; yöntemlerin KTK'ya ve MTK'ya dayalı olmalarına göre yapılmıştır. KTK yöntemleri arasında Mantel-Haenszel, dönüştürülmüş madde indeksi, lojistik regresyon, standartlaştırma yöntemi, ANOVA ve ki-karedir. MTK'ya dayanan yöntemler Lord'un ki-karesi, Raju'nun alan ölçüleri ve MTK-olabilirlik oranıdır (Camili ve Shephard, 1994). DMF belirleme yöntemlerinin bir diğer sınıflaması ise, gözlenen puan ve örtük puanlar yaklaşımıdır. Standartlaştırma yöntemi, Mantel-Haenszel ve lojistik regresyon yaygın olarak kullanılan gözlenen puan yöntemleridir. Olabilirlik oranı testleri, ki kare, maddenin ve testlerin farklı fonksiyonlaşması (DFIT), SIBTEST ve test karakteristik eğrilerinin alanlarının karşılaştırmaları da yaygın olarak kullanılan örtük puan modellerindedir (Potenza ve Dorans, 1995). Bu çalışmada gözlenen puan yöntemlerinden Mantel-Haenszel ve lojistik regresyon, örtük puan yöntemlerinden ise SIBTEST ve MTK-OO yöntemleri kullanılmıştır.

Mantel-Haenszel (MH): En yaygın kullanılan DMF belirleme yöntemlerinden biridir. Tek biçimli DMF'yi belirlemek amacı ile geliştirilen ve ki-kare istatistiğine dayanan bu yöntem, olasılık oranlarını elde etmek için kullanılmaktadır (Agesti, 1984). MH analizinde 2 (alt gruplar) \times 2 (puanlama kategorisi) \times k (eşleştirme ölçütündeki kategori ya da düzey) boyutlarından oluşan üç boyutlu olasılık tablosundaki değişkenlerin bağımsızlığı test edilir (Dorans ve Holland, 1993). Sadece tek biçimli DMF analizinin yapılabildiği MH yönteminin ilk aşamasında olasılık oranı hesaplanır. Bir madde için olasılık oranı (1) no'lu formülle elde edilir.

$$\hat{\alpha}_{MH_i} = \frac{\sum_{j=1}^k \frac{n_{1rij}n_{0fij}}{n_{ij}}}{\sum_{j=1}^k \frac{n_{0rij}n_{1fij}}{n_{ij}}} \quad (1)$$

Burada j, eşleştirme ölçütündeki j. puan düzeyini göstermektedir (j=1, ..., k). Formül (1)'de j. puan kategorisinde ve referans grupta olanlardan maddeyi doğru cevaplayanların sayısı n_{1rij} , yanlış cevaplayanların sayısı n_{0rij} ; j. puan kategorisinde ve odak grupta olanlardan maddeyi doğru cevaplayanların sayısı n_{1fij} , yanlış cevaplayanların sayısı n_{0fij} ile gösterilmektedir. n_{ij} ise toplam cevaplayıcı sayısıdır. Her j kategorisi için elde edilen sonuçlar toplanarak maddenin olasılık oranı elde edilir. Olasılık oranı, bir gruptaki bireylerin bir maddeye doğru cevap verme olasılığının diğer gruptaki bireylerden yüksek olup olmadığını belirler. Olasılık oranı (0, ∞) arasında değerler alabilir. $\hat{\alpha}_{MH_i}$ değerinin 1 olması maddede DMF olmadığı; 1'den küçük olması odak grup lehine DMF olduğu; 1'den büyük olması da referans grup lehine DMF olduğu şeklinde yorumlanır. Olasılık oranının daha kolay yorumlanabilmesi için logaritmik bir dönüşüm ile ΔMH_i elde edilir.

$$\Delta MH_i = -2,35 \ln(\hat{\alpha}_{MH_i}) \quad (2)$$

DMF'nin düzeyi ΔMH_i değerine göre yorumlanır. $|\Delta MH_i| < 1$ ise, maddenin ihmal edilebilir düzeyde (A düzeyi); $1 \leq |\Delta MH_i| < 1,5$ ise orta düzeyde (B düzeyi); $|\Delta MH_i| \geq 1,5$ ise önemli düzeyde (C düzeyi) DMF gösterdiği kabul edilmektedir (Zieky, 1993).

Lojistik Regresyon (LR): Swaminathan ve Rogers (1990), lojistik regresyonu DMF belirleme sürecine uygulamışlardır. Bu yöntem tek biçimli ve tek biçimli olmayan DMF'yi belirleyebilmektedir (Swaminathan ve Rogers, 1990).

$$p(u_i = 1 | \theta, g) = \frac{e^{\beta_0 + \beta_j \theta + \beta_{jg} + \beta_{j(g)}}}{1 + e^{\beta_0 + \beta_j \theta + \beta_{jg} + \beta_{j(g)}}} \quad (3)$$

Formül (3)'de yer alan p_i , i maddesini doğru cevaplandırma olasılığı; β_0 , kesme noktası; β_j , eğim; θ , toplam puanla tanımlanan yetenek düzeyi ve g grup üyeliği indisidir. Model denkleminde, grup üyeliği (g)

teriminin anlamlı olması tek biçimli DMF'yi, grup üyeliği ve yetenek terimlerinin etkileşiminin ($gx\theta$) anlamlı olması da tek biçimli olmayan DMF'yi gösterir (Swaminathan ve Rogers, 1990).

LR ile DMF belirleme sürecinde standartlaştırılmış regresyon katsayıları (R^2) etki büyüklüğü olarak yorumlanır ve DMF'nin düzeyini verir (Gierl, Jodoin ve Ackerman, 2000). LR sürecinde DMF üç düzeyde belirlenir: $R^2 < 0,035$ ise maddede DMF yoktur ya da ihmal edilebilir düzeydedir (A düzeyi), $0,035 \leq R^2 < 0,07$ ise maddede orta düzeyde (B düzeyi), $R^2 \geq 0,070$ ise maddede önemli düzeyde (C düzeyi) DMF vardır (Jodoin ve Gierl, 2001).

Olabilirlik Oranı (LikelihoodRatio-LR): Olabilirlik oranı ile DMF belirleme yöntemi, madde tepki kuramına dayalı olarak oluşturulan modellerin karşılaştırıldığı bir yöntemdir (Thissen, Steinberg ve Wainer, 1993). Olabilirlik oranı ile DMF belirlemede odak ve referans grupların madde parametreleri arasında manidar fark olmadığı şeklinde ifade edilen yokluk hipotezi test edilir. Yokluk hipotezini test etmek için sınırlandırılmış ve genişletilmiş model oluşturularak bu modellerin olabilirlik oranları karşılaştırılır. Sınırlandırılmış modelde, madde parametrelerinin odak ve referans grup için eşit olduğu varsayılır. Genişletilmiş model ise test edilen i. maddenin parametrelerinin odak ve referans grup için farklı olduğu, diğer maddeler için eşit olduğu varsayımına dayalı olarak kurulur. Analiz edilen her bir madde için bir tane sınırlandırılmış model oluşturulur ve bu model genişletilmiş model ile oranlanır.

Olabilirlik Oranı (LR) = Olabilirlik*(Sınırlandırılmış Model) / Olabilirlik*(Genişletilmiş Model)

İki modelin olabilirliklerinin logaritması alınarak G^2 değeri elde edilir (Thissen, 2001). Bu işlem tüm maddeler için tekrarlanır.

$$G^2 = -2LL_s - (-2LL_g) \quad (4)$$

G^2 istatistiği tüm parametrelerin iki grup arasında eşit olup olmadığına dair eş zamanlı bir hipotez testi imkânı sağlar. G^2 değeri, serbestlik derecesi MTK modelindeki parametre sayısı olan χ^2 dağılımının kritik değeri ile kıyaslanarak bu hipotez testi gerçekleştirilir. Anlamlı olan χ^2 değeri DMF'nin varlığını gösterir (Thissen, 2001).

SIBTEST: SIBTEST (Simultaneous Item Bias Test) parametrik olmayan örtük değişken modelidir (Potenza ve Dorans, 1995). İki kategorili verilerde SIBTEST sadece tek biçimli DMF'yi belirlerken, tek biçimli olmayan DMF'yi belirlemek üzere Li ve Stout (1993) tarafından çaprazlanmış-SIBTEST yöntemi geliştirilmiştir. SIBTEST yöntemi, DMF belirlemede I. Tip hatayı kontrol etmede kullanışlı olan regresyon temelli düzeltmeyi kullanır (Clauser ve Mazor, 1998; Cheng, 2005). SIBTEST madde tepki kuramına dayalı olmasına rağmen madde kalibrasyonu gerektirmeyen ve eşleştirme değişkeninin örtük puan olduğu bir yöntemdir. SIBTEST, grup farklılıklarını gerçek puana göre eşleştirilmiş referans ve odak grupların i. maddeyi doğru cevaplandırma olasılıklarını karşılaştırarak hesaplar.

$$B(T) = P_R(T) - P_F(T) \quad (5)$$

İncelenen madde için, T gerçek puanı göstermek üzere, B(T) gruplar arasındaki farklılığı, $P_R(T)$ referans grubun doğru cevaplandırma olasılığını, $P_F(T)$ odak grubun doğru cevaplandırma olasılığını göstermektedir. Toplam DMF, referans ve odak grupların toplamına ait T gerçek puan dağılımından elde edilen B(T)'nin beklenen değeri olan β ile tanımlanır. SIBTEST, $E[B(T)] \equiv \beta = 0$ yokluk hipotezini (DMF yoktur) test eder (Russos ve Stout, 1996).

SIBTEST'te β etki büyüklüğü olarak yorumlanır ve bu değer DMF'nin düzeyini ifade eder. $\beta < |0.059|$ ise, maddede ihmal edilebilir düzeyde (A düzeyi), $\beta |0.059-0.088|$ arasında ise orta düzeyde (B düzeyi), $\beta \geq |0.088|$ ise önemli düzeyde (C düzeyi) DMF olduğu kabul edilmektedir (Rousses ve Stout, 1996).

Tüm DMF yöntemlerinin kendilerine özgü güçlü ve zayıf yönleri vardır. Ancak örtük puan yöntemlerine son yıllarda daha sık başvurulmaktadır. Bunun nedeni örtük puan yöntemlerinin dayandığı MTK'nın, testleri ölçeklemede ve büyük ölçekli sınavların değerlendirilmesinde kullanılmasıdır. Uygun bir yöntem kullanılarak DMF'nin belirlenmesi yansız ve güvenilir bir test oluşturmada önemli bir role sahiptir. Bu nedenle büyük ölçekli sınavlarda yanlı maddelerden kaçınmak ve sınavlardan elde edilen test puanlarına dayalı olarak doğru kararlar almak için hangi koşullarda hangi yöntemin ya da yöntemlerin kullanılmasının daha doğru sonuçlar vereceği önemlidir. Gözlenen puan yöntemlerinden olan LR ve MH araştırmalarda yaygın olarak kullanılmasına rağmen, olabilirlik oranı ve SIBTEST bu yöntemler kadar araştırılmasa da son yıllarda araştırmacıların ilgi odağı olmuştur. Bu araştırmada, çeşitli koşullar altında

bu dört yöntemden hangisinin DMF'yi daha iyi belirlediğini ortaya çıkarmak ve bu yöntemlerin performanslarını karşılaştırmak amaçlanmıştır.

Simülasyon çalışmaları, test uygulamalarında ortaya çıkabilecek olası koşulları ele alır. Bu çalışmanın, araştırmacılara DMF belirleme yöntemlerinin belirli koşullar altında nasıl işlediği hakkında bilgi vereceği düşünülmektedir. Ülkemizde DMF ile ilgili araştırmalar yapılmakla birlikte simülasyon verisi üzerinde yapılan çalışmaların sayısı da oldukça azdır. Bu açıdan da çalışmanın bu konuda araştırma yapacak kişilere örnek olacağı düşünülmektedir.

Gözlenen puan yöntemlerinden olan LR ve MH araştırmalarda yaygın olarak kullanılmaktadır. MTK-OO ve SIBTEST ise değişmezlik özelliği olduğundan ve diğer yöntemlerle karşılaştırıldığında gerçek grup farklılığını hatasız ve yansız olarak kestirmede analitik bir çerçeveye sunduğundan (Shepard, Camilli ve Williams, 1984) son yıllarda araştırmacıların ilgi odağı olmuştur.

Bu çalışmada Mantel-Haenszel, lojistik regresyon, MTK-Olabilirlik Oranı ve SIBTEST yöntemlerinin örneklem büyüklüğü, yetenek dağılımı, DMF gösteren madde oranına göre DMF'li maddeleri doğru bir şekilde belirleyip belirlemediği incelenmiştir. Bu çalışmanın amacı, belirtilen koşullar altında bu dört yöntemden hangisinin DMF'yi daha iyi belirlediğini ortaya çıkarmak ve bu yöntemlerin performanslarını karşılaştırmaktır.

2. YÖNTEM

Bu araştırma, çeşitli koşullar altında DMF belirleme yöntemlerinin performanslarını karşılaştıran bir simülasyon çalışmasıdır. Yapılan araştırmalar DMF belirleme yöntemlerinin örneklem büyüklüğü, odak ve referans grupların yetenek dağılımı, testin uzunluğu, DMF içeren maddelerin oranı ve büyüklüğü gibi pek çok değişkenden etkilendiğini ortaya koymaktadır (Clauser, Mazor ve Hambleton, 1993; Narayanan ve Swaminathan, 1996; Finch ve French, 2007). Araştırmada kullanılan koşulların, yukarıda da ifade edildiği gibi hem yöntemler üzerindeki etkileri göz önüne alınmış, hem de Türkiye'deki gerçek veri uygulamalarına benzer olmasına dikkat edilmiştir.

2.1 Simülasyon Koşulları

Bu araştırmada yöntemlere göre DMF üzerinde etkisi incelenecek değişkenler örneklem büyüklüğü, odak ve referans gruplarının yetenek düzeyi ve testteki DMF'li maddelerin oranıdır. DMF belirleme yöntemlerini etkileyen bir diğer faktör ise testin uzunluğudur. Ancak, çalışmada test uzunluğu 60 madde ile tüm koşullarda sabit tutulmuştur.

Örneklem büyüklüğü: Yapılan araştırmalarda örneklem büyüklüğü arttıkça DMF belirleme yöntemlerinin gücünün de arttığı belirtilmektedir (Rogers ve Swaminathan, 1993; Narayanan ve Swaminathan, 1993, Finch ve French, 2007). DMF'nin etkisini araştıran simülasyon çalışmaları her grup için 200–250; toplamda ise 600 bireyin olduğu veri setlerinin yeterli olduğunu ortaya koymuştur (Zieky, 1993, Narayanan ve Swaminathan, 1994). Bu çalışmada küçük örneklem büyüklüğü için odak ve referans grupları 400'er kişi, büyük örneklem için referans ve odak grupları 1500'er kişi olarak belirlenmiştir.

Yetenek dağılımı: Araştırmada ele alınan ikinci faktör ise, referans ve odak grupları arasındaki yetenek dağılımı farklılığıdır. Yapılan araştırmalar, referans ve odak grupları arasındaki yetenek dağılımlarındaki farklılığın DMF'yi etkilediğini ortaya koymuştur (Narayanan ve Swaminathan, 1994; Jodoin ve Gierl, 2001). Bu çalışmada yetenek dağılımı için ilk durumda, standart normal dağılım gösteren ve yetenek dağılımları eşit olan iki grup alınırken ($N(0,1)$ ve $N(0,1)$), ikinci durumda ise, normal dağılım gösteren eşit olmayan gruplar alınmıştır ($N(0,1)$ ve $N(0.5,1)$).

DMF'li madde oranı: DMF'li maddelerin sayıca fazla olması hem tüm testin geçerliğini düşürmekte, hem de DMF belirleme yöntemlerinin gücünü azaltmaktadır (Jodoin ve Gierl, 2001). Araştırmada DMF'li maddelerin oranı %5 ve %10 olmak üzere iki düzey olarak alınmıştır.

2.2 Verilerin Türetilmesi

Bu çalışmada veriler WINGEN3 yazılımı kullanılarak üç parametrelili lojistik modele göre türetilmiştir (Han, 2007). DMF'li madde parametreleri Tablo 1'de yer almaktadır. DMF'li madde oranının %5 olduğu koşulda Tablo 1'deki ilk üç madde kullanılırken; %10 olduğu koşulda altı maddenin tümü kullanılmıştır.

Tablo 1. DMF’li Maddelerin Türetilmesinde Kullanılan Madde Parametreleri*

Madde no	Referans Grup		Odak Grup		DMF Türü
	a	b	a	b	
1	0.25	0	1	0	Tek biçimli olmayan
2	1	-0.25	1	0.50	Tek biçimli
3	1	-0.30	1	0.45	Tek biçimli
4	0	0	0.75	0	Tek biçimli olmayan
5	1	-0.40	1	0.35	Tek biçimli
6	1	-0.50	1	0.25	Tek biçimli

*Tüm maddeler için $c=.20$ ’dir.

MH, SIBTEST ve LR yöntemleriyle yapılan analizlerde ve büyük ölçekli sınavlarda genellikle orta düzeyde DMF belirlendiğinden (Rogers ve Swaminathan, 1993; Narayanan ve Swaminathan, 1994) bu çalışmada da DMF’li maddelerin tümü orta düzeydedir (B düzeyi). Bu nedenle DMF’li maddeler için odak ve referans gruplar arasındaki parametre farkı 0.75 alınmıştır. Türkiye’de gerçek veri üzerinde yapılan araştırmalarda, tek biçimli olmayan DMF, tek biçimli DMF’den daha az görülmüştür (Yurdugül, 2003; Bekci, 2007; Bakan Kalaycıoğlu, 2009). Bu nedenle bu araştırmada tek biçimli DMF’nin tek biçimli olmayan DMF’ye oranı 2:1 olarak belirlenmiştir. Tek biçimli DMF için, ayırt edicilik parametresi, a , iki grup için aynı, madde güçlük parametresi, b , iki grup için farklı türetilmiştir. Tek biçimli olmayan DMF için ise a parametresi iki grup için farklı, b parametresi iki grup için aynı türetilmiştir. Tüm maddeler için c parametresi 0.20 alınmıştır.

DMF’li olmayan maddelerin parametreleri gerçek veriye dayalı olarak türetilmiştir. Bu çalışmada kullanılan madde parametreleri “Graduate Management Admissions Testi”ne aittir (Clauser, Mazor ve Hambleton, 1993) ve Tablo 2’de gösterilmiştir. DMF’li madde oranının %5 olduğu koşulda Tablo 2’de verilen ilk 54 madde (7–60. maddeler) kullanılırken; %10 olduğu koşulda tüm maddeler kullanılmıştır.

Tablo 2. DMF İçermeyen Maddelerin Parametreleri*

Madde no	a	b	Madde no	a	B	Madde no	a	b	Madde no	a	b
7	0.29	-2.95	22	0.95	0.81	37	0.51	-0.09	52	0.85	-0.37
8	0.41	-2.93	23	1.01	0.81	38	0.55	1.26	53	0.87	-0.75
9	0.94	-1.21	24	0.98	1.67	39	0.73	0.61	54	0.78	-0.05
10	0.88	-0.24	25	0.92	0.42	40	0.88	0.95	55	0.45	-1.49
11	0.42	-1.15	26	0.65	1.68	41	0.55	1.09	56	0.61	-0.53
12	0.74	0.60	27	0.56	-2.70	42	1.40	1.64	57	0.98	0.31
13	0.35	-0.35	28	0.56	-1.77	43	1.35	0.82	58	0.50	0.80
14	0.44	-0.30	29	0.29	-1.39	44	0.75	1.28	59	0.29	-1.00
15	0.55	-1.06	30	0.75	-1.01	45	0.92	1.13	60	0.70	1.05
16	0.82	1.02	31	0.93	-0.23	46	0.73	1.18	61	1.02	0.64
17	0.52	-1.96	32	0.35	-1.12	47	0.75	-1.97	62	1.16	1.11
18	1.02	1.28	33	0.31	-1.37	48	0.73	-1.60	63	0.48	2.12
19	0.65	0.49	34	0.31	-0.86	49	0.64	-1.55			
20	0.82	0.61	35	0.39	-1.17	50	0.81	-0.62			
21	1.04	2.11	36	1.05	0.10	51	0.39	0.04			

*Tüm maddeler için $c=.20$ ’dir.

DMF analizleri iki farklı örneklem büyüklüğü, iki farklı yetenek dağılımı ve iki farklı DMF’li madde oranına göre türetilen sekiz veri seti üzerinde yapılmıştır. Araştırmada MH ve LR analizleri için EZDIF yazılımı (Waller, 1998), SIBTEST ve çaprazlanmış SIBTEST analizleri için SIBTEST yazılımı (Stout, 2005), MTK-OO analizleri için IRTL RDIF yazılımı (Thissen, 2001) kullanılmıştır. LR yönteminde etki büyüklükleri SPSS 11.5’de yazılan syntax ile hesaplanmıştır.

3. BULGULAR

Bu araştırmada sekiz farklı simülasyon koşuluna göre DMF belirlemede kullanılan yöntemlerin etkililiğini belirlemek amaçlanmıştır. Bu amaçla ilk olarak referans ve odak grup için 400'er kişiden oluşan küçük örneklemeler üzerinde MH, LR, SIBTEST ve MTK-OO yöntemlerinden elde edilen sonuçlar Tablo 3'te verilmiştir.

Tablo 3'te tek biçimli DMF gösteren maddeler için MH yönteminin birinci ve üçüncü koşullarda tüm maddeleri B ve C düzeyinde; ikinci ve dördüncü koşullarda dört maddeden ikisini B ve C düzeyinde belirleyebildiği görülmektedir. MH yöntemi ile hiç bir koşulda tek biçimli olmayan maddelerde DMF belirlenmemiş; tüm koşullarda DMF olmayan maddelerde 2-5 arasında değişen sayıdaki maddede A düzeyinde DMF belirlenmiştir. LR ile tek biçimli DMF gösteren maddelerde birinci ve üçüncü koşullardaki iki maddeden birinde A ve B düzeylerinde DMF görülmüş; ikinci ve dördüncü koşulda dört maddeden hiç birinde DMF görülmemiştir.

LR yöntemi ile tek biçimli olmayan maddelerin tümünde DMF görülmüş; ancak LR yöntemi toplam altı maddeden beşini A düzeyinde, birini B düzeyinde belirleyebilmiştir. LR yöntemi ile DMF olmayan maddelerde 2-4 arasında değişen sayıda A düzeyinde DMF belirlenmiştir. SIBTEST yöntemi ile tek biçimli DMF olan maddelerde birinci ve üçüncü koşullardaki iki maddede C düzeyinde; ikinci koşuldaki dört maddenin birinde A, ikisinde C düzeyinde; dördüncü koşuldaki dört maddenin de birinde B, üçünde C düzeyinde DMF belirlenmiştir. SIBTEST ile tek biçimli olmayan DMF gösteren tüm maddeler belirlenmiş; tüm koşullarda B ve C düzeyinde DMF görülmüştür. DMF olmayan maddelerde ise ilk üç koşulda ikisi A düzeyinde, biri B düzeyinde ve üçü C düzeyinde olmak üzere toplam altı maddede SIBTEST yöntemi ile DMF belirlenmiştir. MTK-OO yöntemi ile tek biçimli DMF olan maddelerde birinci ve üçüncü koşullardaki iki maddede B düzeyinde; ikinci koşuldaki dört maddenin birinde A, ikisinde B düzeyinde; dördüncü koşuldaki dört maddenin de ikisinde A, ikisinde B düzeyinde DMF belirlenmiştir. DMF olmayan maddelerde ise 16 maddede A düzeyinde, üç maddede B düzeyinde MTK-OO yöntemi ile DMF olduğu görülmüştür.

Tablo 3. Küçük Örneklemeler İçin DMF'li Olarak Belirlenen Madde Sayıları

Koşullar	Yetenek dağılımı	DMF oranı	DMF türü (madde sayısı)	DMF'li olarak belirlenen madde sayısı												
				MH			LR			SIBTEST			MTK-OO			
				A	B	C	A	B	C	A	B	C	A	B	C	
1.Koşul	OG N(0,1); RG N(0,1)	%5	TB (2)	-	1	1	1	-	-	-	-	-	2	-	2	-
			TBO (1)	-	-	-	1	-	-	-	-	1	-	1	-	
			DMF yok (57)	2	-	-	4	-	-	-	-	1	3	1	-	
2. Koşul	OG N(0,1); RG N(0,1)	%10	TB (4)	-	-	2	-	-	-	1	-	2	1	2	-	
			TBO (2)	-	-	-	2	-	-	-	1	1	1	1		
			DMF yok (54)	3	1	-	2	-	-	1	1	1	4	1		
3.Koşul	OG N(0,1); RG N(0.5,1)	%5	TB (2)	-	1	1	-	1	-	-	-	2	-	2	-	
			TBO (1)	-	-	-	-	1	-	-	-	1	-	1	-	
			DMF yok (57)	5	2	-	2	-	-	1	-	1	5	1	-	
4.Koşul	OG N(0,1); RG N(0.5,1)	%10	TB (4)	-	1	1	-	-	-	-	1	3	2	2	-	
			TBO (2)	-	-	-	2	-	-	-	2	-	-	2	-	
			DMF yok (54)	2	-	-	4	-	-	-	-	-	4	-	-	

OG: odak grup (N=400); RG: referans grup (N=400); TB: tek biçimli; TBO: tek biçimli olmayan

MH, LR, SIBTEST ve MTK-OO yöntemleri DMF içermeyen maddeler açısından incelendiğinde, tüm koşullarda farklı sayıda maddede ihmal edilebilir düzeyde ve önemli düzeyde DMF belirlenmiştir. DMF içermeyen maddeler içinden en az sayıda DMF belirleyen yöntem SIBTEST iken onu sırayla LR, MH ve MTK-OO yöntemleri izlemektedir.

Tablo 4'te odak ve referans grubun 1500'er kişiden oluştuğu büyük örneklemeler için son dört koşula ait sonuçlar verilmiştir.

Tablo 4. Büyük Örneklem İçin DMF’li Olarak Belirlenen Madde Sayıları

Koşullar	Yetenek dağılımı	DMF oranı	DMF türü (madde sayısı)	DMF’li olarak belirlenen madde sayısı											
				MH			LR			SIBTEST			MTK-OO		
				A	B	C	A	B	C	A	B	C	A	B	C
5.Koşul	OG N(0,1); RG N(0,1)	%5	TB (2)	-	1	1	-	-	-	-	-	2	-	-	2
			TBO (1)	-	-	-	1	-	-	-	-	1	-	1	-
			DMF yok (57)	7	-	-	2	-	-	1	-	-	8	-	-
6.Koşul	OG N(0,1); RG N(0,1)	%10	TB (4)	1	3	-	-	-	-	-	1	3	-	2	2
			TBO (2)	-	-	-	1	1	-	1	-	1	-	1	1
			DMF yok (54)	3	-	-	3	-	-	-	-	-	7	-	-
7.Koşul	OG N(0,1); RG N(0.5,1)	%5	TB (2)	1	1	-	-	-	-	-	2	-	-	2	
			TBO (1)	-	-	-	1	-	-	1	-	-	-	1	-
			DMF yok (57)	8	-	-	2	-	-	1	-	-	5	-	-
8.Koşul	OG N(0,1); RG N(0.5,1)	%10	TB (4)	2	1	1	1	-	-	-	1	3	-	2	2
			TBO (2)	-	-	-	2	-	-	-	1	1	-	2	-
			DMF yok (54)	10	-	-	1	-	-	3	-	-	7	-	-

OG: odak grup (N=1500); RG: referans grup (N=1500); TB: tek biçimli; TBO: tek biçimli olmayan

Tablo 4’te tek biçimli DMF gösteren maddeler için MH yönteminin tüm maddelerde A, B ve C düzeylerinde DMF belirlediği; tek biçimli olmayan maddelerde DMF belirlemediği; tüm koşullarda DMF olmayan maddelerde 3–10 arasında değişen sayıdaki maddede A düzeyinde DMF belirlediği görülmektedir. LR yöntemi ile tek biçimli DMF olan maddelerden sadece sekizinci koşulda bir maddede A düzeyinde DMF belirlerken, tek biçimli olmayan DMF gösteren maddelerin tümünde biri B düzeyinde diğerleri A düzeyinde DMF belirlenmiştir. SIBTEST yöntemi, Tablo 4’teki tüm koşullarda tek biçimli DMF gösteren maddelerin tümünde B ve C düzeylerinde; tek biçimli olmayan DMF gösteren maddelerin tümünde A, B ve C düzeylerinde DMF bulmuştur. DMF olmayan maddeler için 6. koşulda SIBTEST tarafından da DMF bulunmamış; diğer koşullarda ise 1 ve 3 arasında değişen sayıda A düzeyinde DMF bulunmuştur. MTK-OO yöntemi ile tek biçimli ve tek biçimli olmayan DMF gösteren maddelerin tümünde B ve C düzeylerinde DMF bulunmuştur. Tablo 4’te yer alan koşullarda DMF olmayan maddelerde 5–8 arasında değişen sayıdaki maddede MTK-OO yöntemi ile A düzeyinde DMF belirlenmiştir.

MH, LR, SIBTEST ve MTK-OO yöntemleri DMF içermeyen maddeler açısından incelendiğinde, tüm koşullarda farklı sayıda maddede ihmal edilebilir düzeyde DMF belirlenmiştir. DMF içermeyen maddeler içinden en az sayıda DMF belirleyen yöntem SIBTEST iken onu sırayla LR, MTK-OO ve MH yöntemleri izlemektedir.

Tablo 5 araştırmada ele alınan koşullar doğrultusunda yöntemler arasındaki uyum yüzdelerini göstermektedir.

Tablo 5. Simülasyon Çalışmasında Kullanılan Yöntemler Arasındaki Uyum Yüzdeleri

Koşullar	YÖNTEMLER											
	MH-LR		MH-SIBTEST		MH-MTKOO		LR-SIBTEST		LR-MTKOO		SIBTEST-MTKOO	
	TB	TBO	TB	TBO	TB	TBO	TB	TBO	TB	TBO	TB	TBO
1 Koşul	%50	%0	%100	%0	%100	%0	%50	%100	%50	%100	%100	%100
2. Koşul	%0	%0	%66	%0	%66	%0	%0	%100	%0	%100	%100	%100
3. Koşul	%50	%0	%100	%0	%100	%0	%50	%100	%50	%100	%100	%100
4. Koşul	%0	%0	%50	%0	%50	%0	%0	%100	%0	%100	%100	%100
5. Koşul	%0	%0	%100	%0	%100	%0	%0	%100	%0	%100	%100	%100
6. Koşul	%0	%0	%100	%0	%100	%0	%0	%100	%0	%100	%100	%100
7. Koşul	%0	%0	%100	%0	%100	%0	%0	%100	%0	%100	%100	%100
8. Koşul	%25	%0	%100	%0	%100	%0	%25	%100	%25	%100	%100	%100

TB: tek biçimli; TBO: tek biçimli olmayan

Tablo 5’de, SIBTEST ve MTK-OO yöntemlerinin tüm koşullarda tek biçimli ve tek biçimli olmayan DMF’yi doğru olarak belirlediği ve aralarındaki uyumun mükemmel olduğu (%100) görülmektedir. MH yöntemi ise, tek biçimli DMF’yi belirlemede SIBTEST ve MTK-OO yöntemleri kadar tutarlı sonuçlar vermemiştir. 2. ve 4. koşullarda, MH yönteminin tek biçimli maddelerin bazılarını doğru olarak belirleyemediği ve 3., 7. ve 8. koşullarda da bazı maddeleri ihmal edilebilir düzeyde belirlediği görüldüğünden bu yöntemin diğer yöntemlerle uyum yüzdesinin genel olarak düşük olduğu görülmüştür. LR yöntemi ise, tek biçimli DMF’yi belirlemede diğer yöntemlere göre oldukça zayıf kalmıştır. Bu yöntem 1., 3. ve 8. koşullarda birer maddeyi doğru olarak belirlemekle birlikte, bu maddelerde ihmal edilebilir düzeyde DMF bulunmuştur. Bu nedenle LR yönteminin tek biçimli DMF’yi belirleme açısından diğer yöntemlerle uyum yüzdesi düşüktür (%50 ve %25). Araştırmada ele alınan yöntemler tek biçimli olmayan DMF açısından incelendiğinde, LR, SIBTEST ve MTK-OO yöntemlerinin uyumunun genel olarak mükemmel olduğu (%100) ancak LR yönteminin bu maddelerde çoğunlukla ihmal edilebilir düzeyde DMF belirlediği görülmüştür. MH yöntemi tek biçimli olmayan DMF’yi belirleyemediği için tüm koşullarda diğer yöntemlerle uyum gösterememiştir. Araştırmada elde edilen bu bulgu, verilerin örtük özelliği dikkate alan bir yaklaşımla (3 parametrelili lojistik modelle) türetilmesi ve dolayısıyla bu şekilde türetilen maddelerin örtük özelliğe dayanan DMF belirleme yaklaşımlarında (SIBTEST VE MTK-OO) uyumlu sonuç vermesiyle açıklanabilir. Bu durumun sonucu olarak da, maddelerin bu şekilde türetilmesinin örtük özelliğe dayalı yöntemler lehine yanlı sonuçlar doğurabileceği de düşünülebilir.

4. TARTIŞMA ve SONUÇ

DMF’yi belirlemede etkili tek bir yöntem olmadığından yöntemlerin etkililiğini karşılaştırmak amacıyla çok sayıda çalışma yapılmıştır. Yapılan araştırmalar, farklı DMF belirleme yöntemlerinin genellikle farklı sonuçlar ortaya çıkardığını (Gao ve Wang, 2005) ve DMF belirleme yöntemleri arasındaki uyumun düşük ya da orta düzeyde olduğunu ortaya koymaktadır (Engelhard, Hansche ve Rutledge, 1990; Hambleton ve Jones, 1994; Raju ve Quartetti, 1995; Akt. Fidalgo, Ferreres ve Muniz, 2004). Bu nedenle, DMF belirlemede birden fazla yöntemin bir arada kullanılması önerilmektedir.

Bu araştırmada kullanılan örtük özelliğe dayalı yöntemlerin (SIBTEST ve MTK-OO) gözlenen puana dayalı yöntemlerden (MH ve LR yöntemleri) DMF’li maddeleri belirlemede daha duyarlı ve etkili olduğu görülmüştür. Verilerin örtük özelliğe dayalı bir yöntem temel alınarak türetilmesinin bu durumun ortaya çıkmasında etkili olabileceği de düşünülmektedir. Bu yöntemler birbirleriyle karşılaştırıldığında ise, DMF’li maddeleri belirlemede MTK-OO yönteminin SIBTEST yönteminden; SIBTEST yönteminin LR ve MH yöntemlerinden ve MH yönteminin ise LR yönteminden daha duyarlı olduğu sonucuna varılmıştır. Araştırmadan elde edilen bu bulgular, Greer (2004)’in yaptığı çalışmayla tutarlılık göstermektedir.

Tek biçimli ve tek biçimli olmayan DMF gösteren maddeleri belirleme açısından yöntemler incelendiğinde, gözlenen puana dayalı olanlardan MH’nin tek biçimli DMF’yi belirlemede; LR’nin ise tek biçimli olmayan DMF’yi belirlemede daha etkili olduğu bulunmuştur. LR yöntemi tek biçimli olmayan DMF’yi belirlemek için; MH yöntemi ise tek biçimli DMF’yi belirlemek için tasarlandığından farklı türdeki DMF’yi belirlemede etkili olamamaktadır (Swaminathan ve Rogers, 1990). Araştırmadan elde edilen bu sonuç, Hidalgo ve Lopez-Pina (2004)’nin elde ettiği sonuçla benzerlik göstermektedir. Rogers ve Swaminathan (1993) ile Swaminathan ve Rogers (1990) yaptıkları araştırmada, tek biçimli olmayan DMF’yi belirlemede LR’nin MH’den daha güçlü olduğunu, tek biçimli DMF’yi belirlemede ise MH ile yaklaşık aynı güçte olduğunu bulmuşlardır. Aynı şekilde Narayanan ve Swaminathan (1996) ise tek biçimli olmayan DMF’yi belirlemede LR ve SIBTEST’in MH’den daha güçlü olduğunu bulmuşlardır.

Bu araştırmada LR ile doğru tespit edilen madde sayısının MH ile doğru tespit edilenden daha az olması, Gomez-Benito ve Navas-Ara (2000)’nin sonuçları ile de tutarlılık göstermektedir. LR için belirlenen kategorilendirme sisteminin DMF’in gerçek miktarını belirlemede yeterince hassas olmamasının ve LR’deki etki büyüklüğünü belirlemede MH’den farklı bir kategorilendirme sisteminin kullanılmasının, LR ile daha az sayıda DMF’li madde belirlemeye yol açtığı belirtilmektedir (Pitoniak, Cook, Cline ve Cahalan-Laitusis, 2006).

Örtük puana dayalı yöntemler tek biçimli ve tek biçimli olmayan DMF’yi belirleme açısından karşılaştırıldığında ise, SIBTEST ve MTK-OO yöntemlerinin genel olarak eşit güçte olduğu bulunmuştur.

SIBTEST'in, MTK-OO ve diğer yöntemlerle karşılaştırıldığı araştırmalarda da en güçlü ve etkili yöntemin SIBTEST olduğu sonucuna ulaşılmıştır (Cheng, 2005; French ve Finch, 2007).

Araştırmada MH ve LR yöntemlerinin SIBTEST ve MTK-OO yöntemlerine göre orta düzeyde DMF'yi tanımlamada yetersiz olduğu bulunmuştur. Bu sonuç Wiberg (2009)'in yaptığı araştırmada MH yönteminin orta düzeyde DMF'yi tanımlamada yetersiz olduğu bulgusu ile paralellik göstermektedir. Ancak Narayanan ve Swaminathan (1996) LR ve SIBTEST yöntemlerinin tek biçimli olmayan DMF belirlemede aynı güce sahip olduklarını bulmuşlardır. Ayrıca MTK-OO yönteminin tüm koşullar için DMF'li maddeleri doğru bir şekilde belirlemede oldukça güçlü olduğu sonucu Atar (2007)'in çalışmasıyla paralellik göstermektedir.

DMF göstermeyen maddelerde tüm koşullarda tüm yöntemler tarafından farklı sayıda maddede DMF gözlenmiştir. Bu durum, I. tip hatayı ifade eder. Büyük örneklerde yöntemlerin daha fazla I. tip hataya düştüğü, ancak bu maddelerin hepsinin ihmal edilebilir düzeyde olduğu görülmüştür. Küçük örneklerde I. tip hata daha az olmasına rağmen, bazı maddelerde önemli düzeyde DMF belirlenmiştir. SIBTEST ile tüm koşullarda diğer yöntemlere göre daha az I. tip hata yapılmıştır. Bu durum, SIBTEST'in regresyon temelli düzeltmeyi kullanarak I. Tip hatayı kontrol etmesi ile açıklanabilir. Atar (2007) örneklem büyüklüğü ve DMF etki büyüklüğü arttıkça LR ve MTK-OO yöntemlerinin gücünün ve I. tip hata oranlarının arttığı sonucuna varmıştır.

Araştırmada, LR'nin diğer yöntemlere göre DMF belirleme oranının daha düşük olduğu, MTK-OO'nun ise daha yüksek olduğu sonucuna varılmıştır. Araştırmada elde edilen bu sonuç, Berberoğlu ve Bakan-Kalaycıoğlu (2010) tarafından yapılan çalışmada elde edilen sonuçla tutarlılık göstermektedir. Araştırmanın bu sonucu, yöntemlerin etki büyüklüklerinin sınıflandırılmasındaki farklılıklardan kaynaklanabilir. Hidalgo ve Lopez-Pina (2004) LR'nin DMF'li maddeleri belirlemede çok tutucu bir teknik olduğunu ifade etmişlerdir.

MH ve LR yöntemleri arasında DMF'li maddeleri belirleme bakımından 1. ve 3. koşullarda orta düzeyde; 8. koşulda düşük düzeyde bir uyum olduğu; diğer koşullarda uyum olmadığı görülmüştür. Pitoniak, Cook, Cline ve Cahalan-Laitusis (2006)'in araştırmalarında da MH ve LR arasında düşük düzeyde uyum görülmüştür. MH ile SIBTEST ve MTK-OO arasında 2. ve 4. koşullarda orta derecede; diğer koşullarda tam uyum gözlenmiştir. Fidalgo, Ferrerez ve Muniz (2004) yaptıkları araştırmada MH ile SIBTEST arasında yüksek düzeyde uyum bulmakla birlikte SIBTEST'in DMF'li maddeleri tespit etme oranının daha yüksek olduğu sonucuna varmışlardır. Narayanan ve Swaminathan (1994) ise yaptıkları çalışmada tek biçimli DMF belirlemede eşit yetenek dağılımında MH ve SIBTEST'in aynı güce olduğunu, eşit olmayan yetenek dağılımlarında SIBTEST'in MH'den daha güçlü olduğunu bulmuşlardır.

LR ile SIBTEST ve MTK-OO arasında tek biçimli DMF belirlemede 1. ve 3. koşullarda orta, 8. koşulda düşük uyum görülürken, diğer koşullarda hiç uyum gözlenmemiştir. LR ile SIBTEST ve MTK-OO arasında tek biçimli olmayan DMF belirlemede ise tam uyum görülmüştür. Bu durum, tek biçimli olmayan DMF'yi belirlemede LR'nin diğer iki yöntem kadar etkili olduğunu göstermekle birlikte; LR'nin genellikle A düzeyinde, diğer iki yöntemin ise B ve C düzeylerinde DMF belirledikleri de dikkate alınmalıdır. Örtük puana dayalı yöntemler olan SIBTEST ve MTK-OO'nun tek biçimli ve tek biçimli olmayan DMF belirlemede tam bir uyum gösterdikleri ortaya çıkmıştır.

4.1 Öneriler

Bu araştırmada, çeşitli koşullarda MH, LR, SIBTEST ve MTK-OO yöntemlerinden hangisinin DMF'yi daha iyi belirlediğini ortaya çıkarmak ve bu yöntemlerin performanslarını karşılaştırmak amaçlanmıştır. Odak ve referans grubun 400'er kişi alındığı küçük örneklerde MH, SIBTEST ve MTK-OO yöntemlerinin kullanılması önerilebilir. Ancak bu yöntemlerden SIBTEST'in DMF'li olmayan maddelerde DMF belirleme miktarının diğerlerine göre daha az olduğu dikkate alınarak bu yöntemin kullanılması tercih edilebilir. Büyük örneklerde ise SIBTEST dışındaki yöntemlerde DMF olmayan maddelerde DMF belirleme miktarında dikkate değer bir artış görülmüştür. Bu nedenle büyük örneklerde de I. tip hataya kontrol etmek için SIBTEST yönteminin kullanılması önerilebilir. Sonuç olarak, gerçek uygulamalarda araştırmada ele alınan koşullar dikkate alındığında, hem tek biçimli hem de tek biçimli olmayan DMF'yi belirlemede örtük puana dayalı yöntemlerin kullanılması önerilebilir.

Bu araştırmanın tek bir veri seti üzerinde gerçekleştirilmesi bir sınırlılık olarak görülebilir fakat bu konuda yapılan araştırmalar incelendiğinde (Thissen, Steinberg ve Wainer, 1988; Akt, Wang ve Yeh,

2003; Yıldırım, 2008; Padilla, Hidalgo, Benitez ve Gomez-Benito, 2011) bu tür araştırmaların da olduğu görülmektedir. Bununla birlikte tek bir veri seti üzerinde araştırma yapmanın birden fazla veri seti üzeri üzerinde yapmaktan daha az güvenilir sonuçlar vereceğini de unutulmamalıdır. Bundan sonraki araştırmalarda, birden fazla veri seti kullanılarak yöntemlerin I. tip hata oranları ve güçleri hesaplanabilir. Ayrıca, diğer DMF belirleme yöntemleri kullanılarak bu yöntemlerin I. tip hata oranları ve güçleri de araştırılabilir.

KAYNAKLAR

- Agresti, A. (1984). *Analysis of ordinal categorical data*. Newyork: John Wiley & Sons.
- Atar, B. (2007). *Differential item functioning analyses for mixed response data using Irt likelihood ratiotest, logistic regression and gllamm procedures*. Unpublished doctoral thesis, The Florida State University, Florida State.
- Bakan Kalaycıoğlu, D. (2009). *Öğrenci Seçme Sınavı'nın madde yanlılığı açısından incelenmesi*. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Bekci, B. (2007). *Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı'nın değişen madde fonksiyonlarının cinsiyete ve okul türüne göre incelenmesi*. Yüksek lisans tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Berberoğlu, G., ve Kalaycıoğlu Bakan, D. (2010, in Press). Differential item functioning analysis of the science and mathematics items in the university entrance examinations in Turkey. *Journal of Psychoeducational Assessment*.
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4thed., Vol. 4, pp. 221-256). Westport: American Council on Education & Praeger Publishers.
- Camili, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. London: Sage Publications.
- Chang, H., Mazzeo, J., & Roussos, L. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *Journal of Educational Measurement*, 33(3), 333-353.
- Cheng (2005). *A study on differential item functioning of the basic mathematical competence test for junior high schools in Taiwan*. Unpublished doctoral thesis, Michigan State University.
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice*, 17, 31-44.
- Clauser, E. B., Mazor, K., & Hambleton, K.R. (1993). The effects of purification of the matching criterion on the identification of DIF using the Mantel-Haenszel procedure. *Applied Measurement in Education*, 6(4), 269-279.
- Dorans, N. J., & Holland, P.W. (1993). DIF detection and description: Mantel-Haenszel and Standardization. *Differential item functioning*. In P.W. Holland & H.Wainer (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hidalgo, A. M., Ferreres, D., & Muniz, J. (2004). Liberal and conservative differential item functioning detection using Mantel-Haenszel and SIBTEST: Implications for Type I and Type II error rates. *Journal of Experimental Education*, 75(1), 23-29.
- Finch, W. H., & French, B.F. (2007). Detection of crossing differential item functioning: Comparison of four methods. *Educational and Psychological Measurement*, 67, 4, 565-582.
- Gao, L., & Wang, C. (2005). *Using five procedures to detect DIF with passage-based testlets*. A paper prepared for the Poster Presentation at the Graduate Student Poster Session at the Annual Meeting of the National Council of Measurement in Education, Montreal, Quebec.
- 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.
- Gomez-Benito, J., & Navas-Ara, M. J. (2000). A comparison of Chi-square, RFA and IRT based procedures in the detection of DIF. *Quality and Quantity*, 34, 17-31.
- Güler, N., & Penfield, R. (2009). A comparison of the logistic regression and contingency table methods for the simultaneous detection of uniform a nonuniform differential item functioning. *Journal of Educational Measurement*, 46, 3.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications
- Han, K. T. (2007). *WinGen2: Windows software that generates IRT parameters and item responses [computer program]*. Amherst, MA: Center for Educational Assessment, University of Massachusetts Amherst. 10 Aralık 2010 tarihinde <http://people.umass.edu/kha/wingen/> adresinden indirilmiştir.
- Hidalgo, M. D., & Lopez-Pina, J. A. (2004). Differential item functioning detection an effect size: A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement*, 64(6), 903-915.
- 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.
- Li, H.-H., & Stout, W. (1996). A new procedure for detection of crossing DIF. *Psychometrika*, 61, 647-677.
- Miller, T. R., & Spray, J. A. (1993). Logistic discriminant function analysis for DIF identification of polytomously scored items. *Journal of Educational Measurement*, 30(2), 107-122.
- Narayanan, P., & Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential. *Applied Psychological Measurement*, 18 (4), 315-328.
- Narayanan, P., & Swaminathan, H. (1996). Identification of items that non-uniform DIF. *Applied Psychological Measurement*. 20(3), 257-274.
- Padilla, J. L., Hidalgo, M. D., Benítez, I. & Gómez-Benito, J. (2011). Comparison of three software programs for evaluating DIF by means of the Mantel-Haenszel procedure: EASY-DIF, DIFAS and EZDIF. *International journal of methodology and*

- experimental psychology, preprints. <http://www.uv.es/psicologica/preprints/Padilla.pdf> adresinden 16.11.2011 tarihinde alınmıştır.
- Penfield, R. D. (2003a). Application of the Breslow-Day test of trend in odds heterogeneity to the detection of nonuniform DIF. *ALberta Journal of Educational Research*, 49, 231-243
- Pitoniak, M., Cook, L., Cline, F., & Cahalan-Laitusis, C. (2006). *Using differential item functioning to investigate the impact of accommodations on the scores of students with disabilities on english-language arts assessments*. Educational Testing Service, NCME Presentation.
- Potenza, T. M., & Dorans, J. N. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. *Applied Psychological Measurement*, 19(1), 23-37.
- Roussos, L.L., & 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(2), 215-230.
- Rogers, H. J., & Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement*, 17, 105- 116.
- Shepard, L. A., Camilli, G., & Williams, D. M. (1984). Accounting for statistical artifacts in item bias research. *Journal of Educational Statistics*, 9, 93-128.
- Shealy, R. T., & Stout, W. F. (1993). An item response theory model for test bias and differential test functioning. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp.197-239). Hillsdale, NJ: Erlbaum.
- Stout, W. (2005). *SIBTEST Software* (1.7 sürümü) [Bilgisayar Yazılımı]. Institute for Measurement: Assessment System Corporations.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361- 370.
- Thissen, D. (2001). *IRTLRDIF* (2.0b Sürümü) [Computer software]. Chapel Hill: L. L. Thurstone Psychometric Laboratory, University of North Carolina.
- Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group-mean differences: The concept of item bias. *Psychological Bulletin*, 99, 118-128.
- Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H.Wainer&H. Braun (Eds.), *Test validity* (pp. 147-169). Hillsdale, NJ: Lawrence Erlbaum.
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using parameters of itemresponsemodels. In P.W. Holland & H.Wainer (Eds.), *Differential item functioning* (pp. 67-113). Hillsdale.
- 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.
- Wiberg, M. (2009). Differential item functioning in mastery tests: A comparison of three methods using real data. *International Journal of Testing*, 9, 41-59.
- Yıldırım, S. (2008). Farklı işleyen maddelerin belirlenmesinde sınırlandırılmış faktör çözümlenmesinin olabilirlik-oranı ve mantel-haenszel yöntemleriyle karşılaştırılması. *H.Ü. Eğitim Fakültesi Dergisi*, 34, 297-307.
- Yurdugül, H. (2003). *Orta öğretim kurumları öğrenci seçme ve yerleştirme sınavının madde yanlılığı açısından incelenmesi*, Doktora tezi, Hacettepe Üniversitesi, Ankara.
- Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. Holland& H. Wainer (Eds.), *Differential item functioning*, (pp. 337-347). Hillsdale, NJ: Erlbaum
- 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.

Extended Abstract

Bias is an important subject while investigating the validity of test. Test and item bias reduces the validity of test and causes error in decisions. Item bias occurs when examinees of one group are less likely to answer an item correctly (or endorse an item) than examinees of another group because of some characteristic of the test item or testing situation that is not relevant to the test purpose (Zumbo, 1999). This situation provides advantage to a group of individual with specific characteristics. Especially item bias is become a crucial problem at large scale tests. Determination of item bias starts with statistical process. Differential item functioning (DIF) -which investigate the success probabilities of examinees from different groups on item after matching on the underlying ability- is the first step on the determination of item bias. DIF is required, but not sufficient, for item bias. Thus, after the determination of DIF, sources of difference, sub groups, test structure or content were explored to decide if item is biased or not (Dorans & Holland, 1993). Two types of DIF can occur: uniform and non-uniform. Uniform DIF is defined as a difference between the groups in the probability of a correct response to an item at all ability levels. Nonuniform DIF refers to the case where an item discriminates across the levels of ability differently for the groups (Camilli & Shepard, 1994). Four DIF detection methods were compared in this investigation: simultaneous item bias test (SIBTEST), IRT likelihood ratio test (IRTLR), logistic

regression, and the Mantel-Haenszel technique. The first two are IRT-based, parametric techniques; logistic regression and the Mantel-Haenszel method are CTT-based procedures.

The primary purpose of this research is to compare and evaluate the effectiveness of observed score methods -Mantel-Haenszel, logistic regression- and latent score methods -IRT-LR, SIBTEST- which used to determine DIF under variety conditions. These methods were compared by simulation study. Sample sizes, mean ability distribution, amount of items with DIF were considered for data simulation conditions.

Power of DIF detection can increase as sample size increases (Rogers & Swaminathan, 1993; Narayanan & Swaminathan, 1993, Finch & French, 2007). Therefore, two sample sizes were included to examine conditions: (a) 800, with 400 in each group; (b) 3000, with 1500 in focal and 1500 in reference group. Ability differences can influence DIF detection (Jodoin & Gierl, 2001; Narayanan & Swaminathan, 1994). Thus, two conditions were simulated. In one condition, the reference and focal groups' population means and standard deviations were equal and show a standard normal distribution ($N(0, 1)$). As a second, group ability differences were simulated where the mean of the reference group was 0.0 and the focal group was 0.5, standard deviation was 1 for both of them. Numerous DIF items were decrease the validity of test (Jodoin & Gierl, 2001). In addition to this, power of DIF detection methods decline. The percentage of DIF items was set at 5% and 10%.

Item response data were simulated by WINGEN3 with a 3 PL model (Han, 2007). During this process, certain parameters were needed for both DIF and non DIF items. For acquiring realistic results, non DIF parameters were received from real data set. Uniform DIF was generated same for both group in item discrimination parameter (a) and item difficulty parameter (b) was set different for two groups. In non-uniform DIF, item difficulty parameter was set same and item discrimination parameter (a) was set different for two groups. Since it was not included in the focus of the investigation, the c-parameter was fixed to 0.20 so as to reduce the risk of estimation inaccuracy. Differences between item parameters were set 0.75 for generating moderate DIF. In simulation studies, DIF items with large impact magnitude were not generated because moderate DIF were detected by MH, SIBTEST and LR methods and usually moderate DIF occurs in large scale tests (Narayanan & Swaminathan, 1994; Rogers & Swaminathan, 1993). The ratio of uniform DIF to non-uniform DIF was set 2:1.

Eight data sets were used to DIF analyses (two different combinations of sample size, two different combinations of mean ability differences and two different combinations of percentage of DIF items). EZDIF for MH and LR analyses, DIFPACK for SIBTEST and crossing SIBTEST analyses and IRTLRDIF for IRTLR analyses were used in this research. In addition to EZDIF software, SPSS 11.5 was used to calculate impact magnitudes in LR method.

As a result, MH, SIBTEST and IRT-LR methods present consistent result in determining uniform DIF in all conditions. Furthermore, consistent results were found in identifying non-uniform DIF with LR, SIBTEST and IRT-LR methods. Also, results of this research revealed that latent score methods were more sensitive and effective in determining items with DIF rather than observed score methods. Consequently, latent score methods were more liberal and observed score methods were more conservative in identifying items with DIF. The reason of this condition can be the criteria's that used to classification of impact magnitudes. In this respect, LR method is more conservative and IRT-LR is more liberal than other methods in the detection of DIF.

In high stake test conditions, for identifying DIF items more reliably, using combination of multiple methods may consider. In this research, simulation study was executed by single data set which is seen as a limitation. Further research will use several data sets and calculate the Type I error rates and powers of methods. Although, other DIF detection methods' power will explore in future researches.