

## Öğrenci Seçme Sınavı'nın Madde Yanlılığı Açısından İncelenmesi

### Item Bias Analysis of the University Entrance Examination

Dilara BAKAN KALAYCIOĞLU\*

Hülya KELECİOĞLU\*\*

ÖSYM

Hacettepe Üniversitesi

#### Öz

Bu çalışmada, 2005 yılı Öğrenci Seçme Sınavı'nda (ÖSS) yer alan maddelerin cinsiyete göre Değişen Madde Fonksiyonu (DMF) içerip içermediği incelenmiştir. DMF içerdiği belirlenen maddelerin yanlı olup olmadığına karar vermek için uzman görüşüne başvurulmuştur. DMF belirleme yöntemlerinden Mantel-Haenszel (MH) ve Lojistik Regresyon (LR) kullanılmıştır. Analizler 2005 yılında ÖSS'ye giren 599,330 lise son sınıf öğrencisinin her biri 45 maddeden oluşan Türkçe, sosyal bilimler, matematik ve fen bilimleri alt testlerine verdikleri cevaplar kullanılarak yapılmıştır. Araştırmanın sonucunda, cinsiyete göre Türkçe alt testini oluşturan maddelerin DMF içermediği sonucuna varılmıştır. Sosyal bilimler alt testinde yedi, matematik ve fen bilimleri alt testlerinde de üç maddenin DMF içerdiği ve bu maddelerden fen bilimleri testinde yer alan bir tanesinin yanlılık gösterdiği belirlenmiştir.

*Anahtar Sözcükler:* Madde yanlılığı, Değişen Madde Fonksiyonu, Mantel-Haenszel yöntemi, Lojistik Regresyon.

#### Abstract

This study investigated the gender related Differential Item Functioning (DIF) of the 2005 University Entrance Examination (UEE) in order to decide whether a DIF item is biased. In this study, measurement specialists' opinions were gathered and Mantel-Haenszel (MH) and Logistic Regression (LR) DIF detection methods were used. The analyses were based on the answers of 599.330 high school seniors to the Turkish, social sciences, mathematics and natural science subtests of the 2005 UEE. It was found that no items were flagged for gender DIF in the Turkish subtest. However, seven social sciences items and three mathematics and natural sciences items displayed DIF. Among these items, only one natural sciences item was identified as biased.

*Keywords:* Item bias, differential item functioning, Mantel-Haenszel method, logistic regression.

#### Summary

#### Purpose

The purpose of this research was two fold. The first purpose was to identify items that function differently between female and male students in the University Entrance Examination. In order to investigate gender DIF, MH and LR DIF detection methods were employed. Since DIF is a necessary but not sufficient condition for finding any item bias, a secondary purpose of the study was set which was to determine whether gender DIF items were biased.

#### Method

The sample of the study consisted of 599.330 high school senior students who took the UEE in 2005. Forty-six percent of these students were female (273.419) and fifty-four percent of were male (325,911).

\* Dr. Dilara BAKAN KALAYCIOĞLU, Öğrenci Seçme ve Yerleştirme Merkezi, dilara.bakan@osym.gov.tr

\*\* Doç. Dr. Hülya KELECİOĞLU, Hacettepe Üniversitesi, Eğitim Fakültesi, hulyaebb@hacettepe.edu.tr

For the DIF analysis, students' answers to the UEE items were used. There were Turkish, social sciences, mathematics and natural sciences subtests in the UEE. Each subtest consisted of 45 items. Turkish subtest assesses the proficiency of Turkish language, social sciences (history, geography and philosophy) subtest assesses using social science concepts and generalizations, mathematics subtest assesses using basic mathematical concepts and rules and natural sciences (physics, chemistry and biology) subtest assesses using natural science concepts and generalizations ([www.osym.gov.tr](http://www.osym.gov.tr)). The Cronbach's alpha reliability values of the Turkish, social sciences, mathematics and natural sciences subtest' were 0.93, 0.95, 0.96 and 0.97, respectively. All items were in multiple choice formats with five alternatives. The UEE data were provided by Student Selection and Placement Center, Research and Development Unit.

In order to identify the biased test items, a survey was developed for measurement specialists. The survey consisted of DIF items and possible sources of the item bias. Specialists were requested to classify all items favoring females and males. Then, the question of "What are the possible sources of the bias?" was asked to the specialists. MH and LR analyses were carried out to find out whether UEE items display DIF.

### *Results*

In the Turkish subtest, no item was identified for gender DIF. In the social sciences subtest, one history and six philosophy DIF items were observed. The history item was in favor of the male students and six philosophy items were in favor of female students. In the mathematics subtest, one algebraic calculation and two geometry DIF items were observed. The calculation item was in favor of the female students and the geometry items were in favor of the male students. In the natural sciences subtest, three physics DIF items were observed and all of these DIF items were in favor of the male students.

Among thirteen DIF items, only one physics item was determined as biased by the experts.

### *Conclusion*

Social sciences, mathematics and natural sciences subtests of the UEE displayed gender DIF items. Measurement spacialists indicated that some gender based differential item performance is expected, but this cannot always be interpreted as item bias. The item content (velocity) and item format (diagram with automobiles) of the physics item may be the causes of gender bias favoring the male students.

Results of this study yielded that female students performed relatively better in philosophy, and algebraic calculations once compared to the male students who performed relatively better in physics, problem solving and geometry.

### *Giriş*

Yükseköğrenime öğrenci seçimi tüm dünyada çeşitli eleme ve seçme yöntemleri ile yapılmaktadır. Amaç, yükseköğretimde başarılı olabilecek öğrencilerin kapasiteleri doğrultusunda üniversitelere yerleştirilerek kaynakların en etkili şekilde kullanılmasıdır. Ülkemizde bu amaçla her yıl Öğrenci Seçme Sınavı (ÖSS) yapılmaktadır. 2008 yılında Öğrenci Seçme ve Yerleştirme Merkezi'ne 120,798'i sınavsız geçiş olmak üzere toplam 1,730,876 aday başvurmuştur. Bu adaylardan, 198,509'u lisans, 180,221'i ön lisans ve 229,264'ü açık öğretim programlarına yerleştirilmiştir ([www.osym.gov.tr](http://www.osym.gov.tr)). Yüzdesel olarak ifade etmek gerekirse, öğrencilerin sadece %11'i örgün lisans eğitimine devam edebilmeye hak kazanmıştır; bu durum da ÖSS'nin önemini açıkça ortaya koymaktadır.

ÖSS gibi seçme ve yerleştirme amaçlı kullanılan testlerin amacı, ölçülmek istenen yetenek doğrultusunda bireyler arasındaki farklılıkları olabildiğince hassas ve doğru bir şekilde ölçebilmektir. Bu ölçme yapılırken kullanılan sınavın, sınava giren kitle içindeki hiçbir alt gruba avantaj sağlamaması, yanlı olmaması beklenir.

*Yanlılık ve Değişen Madde Fonksiyonu*

Yanlılık, farklı alt gruplardaki bireylerin test puanlarının buldukları gruba bağlı olarak sistematik hata içermesidir (Camilli ve Shepard, 1994; Zumbo 1999). Bu anlamda yanlılık bir geçerlik sorunudur. Testte yanlı maddelerin bulunması, testin geçerliğini ve buna bağlı olarak güvenilirliğini olumsuz yönde etkiler. Bir testin tarafsızlığı ve geçerliği için en büyük tehdit yanlılıktır (Kristanjansson, Aylesworth, McDowell ve Zumbo, 2005).

Yanlılık çalışmalarında ilk adım, alt grupların maddeleri yanıtlama yapıları arasında farklılık olup olmadığını yani maddelerin Değişen Madde Fonksiyonu (DMF) içerip içermediğinin belirlenmesine yönelik istatistiksel analizlerin uygulanmasıdır. Bir maddenin yanlı olduğunu söyleyebilmek için öncelikle o maddenin Değişen Madde Fonksiyonu içermesi gerekir. Ardından, maddenin DMF içermesinin olası sebepleri tespit edilerek, maddenin ölçülmek istenen yapıdan ilişkisiz olarak bazı alt gruplar için adil olmayan bir avantaj sağlayıp sağlamadığının uzman kişiler tarafından belirlenmesi gerekir (Camilli ve Shepard, 1994; Zumbo, 1999; Zumbo, 2007).

DMF, testle ölçülen özellik bakımından benzer olan ancak cinsiyet, sosyoekonomik düzey gibi değişkenler açısından birbirinden farklı alt gruplarda yer alan bireylerin, bir maddeyi doğru cevaplandırma olasılıklarının farklılaşması olarak tanımlanabilir (Hambleton, Swaminathan ve Rogers, 1991). DMF'nin belirlenmesi farklı alt gruplarda yer alan bireylerin testle ölçülen özellikler bakımından eşit olduğu ya da eşitlendiği varsayımı altında yapılır. Amaç, gerçek grup farklılıkları ile ölçmedeki yanlılık arasındaki farkı ayırt etmektir. Bunu yaparken de öncelikle farklı gruplarda aynı yetenek düzeyine (aynı toplam puana) sahip bireyler eşleştirilir ve böylece karşılaştırılabilen bireyler belirlenir. Ardından, bu bireylerin test maddelerinde gösterdikleri performansın aynı olup olmadığına bakılır.

DMF, tek biçimli ve tek biçimli olmayan olmak üzere iki grupta değerlendirilir. Bir grupta bir maddeye doğru cevap verme olasılığı, tüm yetenek düzeylerinde diğer gruptan daha yüksek ise buna tek biçimli DMF adı verilir. Tek biçimli olmayan DMF ise bir grupta bir maddeye doğru yanıt verme olasılığının, bazı yetenek düzeylerinde diğer gruptan düşük, bazı yetenek düzeylerinde ise diğer gruptan yüksek olması durumudur (Swaminathan ve Rogers, 1990).

DMF'yi belirlemek amacıyla klasik test kuramına, varyans analizine,  $\chi^2$  analizine ve madde tepki kuramına dayalı çok sayıda istatistiksel yöntem geliştirilmiştir. Bu yöntemlerden bazıları: Mantel-Hanszel, Lojistik Regresyon, SIBTEST, Lord's  $\chi^2$ , Raju'nun alan ölçüleri, olabirlik oran testi olarak sıralanabilir (Benito ve Ara, 2000; Camilli ve Shepard, 1994; Hambleton, Swaminathan ve Rogers, 1991, Oshima & Morris, 2008). Bu çalışmada DMF,  $\chi^2$  analizine dayalı yöntemlerden Mantel-Haenszel (MH) ve Lojistik Regresyon (LR) ile belirlenecektir.

Literatürde, özellikle sınavların cinsiyete göre yanlılık gösterip göstermediğini belirlemeye yönelik farklı yöntemler kullanılarak yürütülmüş çalışmalar bulunmaktadır. Amerika Birleşik Devletleri'nde üniversiteye giriş sınavı olarak kullanılan The American College Testing ve Scholastic Aptitude Test üzerinde yapılan çalışmalarda, algoritmik, hesaplamaya dayanan, karmaşık olmayan, değişken içeren, daha soyut konularla ilgili maddelerin kız öğrenciler lehine DMF içerdiği; geometri, aritmetik ve cebirsel muhakeme ve yüksek zihinsel faaliyetler gerektiren maddelerin ise erkek öğrenciler lehine DMF içerdiği belirlenmiştir (Doolittle ve Cleary 1987; Harris ve Carlton 1993). Azen, Bronner ve Gafni (2002) İsrail'de aynı amaçla yapılan Psychometric Entrance Testi'nin sayısal bölümünde erkek öğrencilerin avantajlı olduklarını ifade etmişlerdir.

ÖSS üzerinde yapılan farklı DMF çalışmalarında, kız öğrenciler lehine çalışan maddeler: sosyal bilimlerdeki sosyoloji ve psikoloji, fen bilimleri alt testinde biyoloji maddeleri, erkek öğrenciler lehine çalışan maddeler ise Türkçe alt testindeki paragraf, sosyal bilimlerdeki tarih ile fen bilimleri alt testindeki fizik maddeleri olarak belirlenmiştir. Bir başka çalışmada ise sosyal bilimlerdeki tarih ve coğrafya maddelerinin erkek öğrenciler lehine çalışırken, psikoloji ve sosyoloji maddelerinin kız öğrenciler lehine çalıştığı; fen bilimleri alt testinde ise en çok biyoloji maddelerinin kız öğrenciler lehine DMF içerdiği belirlenmiştir (Berberoğlu, 1996; Öğretmen, 1995; Yenal, 1995).

DMF'nin kaynağını tespit etmek amacıyla çeşitli ölçme araçları üzerinde yapılan çalışmalarda; madde formatı, içeriği, bilişsel seviyesi gibi bazı madde karakteristiklerinin öğrencilerin performansı üzerinde etkili olabildiği belirlenmiştir (Mendes-Barnett ve Ercikan, 2006; Yıldırım ve Berberoğlu, 2009; Zumbo ve Gelin, 2005). Educational Testing Service'in (ETS) 2009 yılında madde yazarları ve test geliştirme uzmanlarına yönelik yayımladığı tarafsızlık yönergesinde, maddelerde, bir cinsiyet grubunun daha aşına olduğu konulardan (spor bilgisi, alet kullanımı vb.), testin amacı dışında gereksiz derecede zor kelimelerden, söz dizimlerinden ve test ile ölçülmek istenen yapıyla ilgisiz harita, tablo ve grafik kullanımından kaçınılması gerektiği belirtilmiştir.

#### *Amaç*

Sonuçlarına dayanılarak yerleştirme, mezuniyet, geçme/kalma gibi çok önemli kararların verildiği sınavların hukuksal olarak savunulabilmesi, bu sınavlar üzerinde DMF çalışmalarının yapılmış olmasına bağlıdır (Haladyna, 1997). ÖSS'de kullanılan maddelerin ÖSS'nin eğitim sistemimizdeki yeri nedeni ile önceden denenmesi mümkün olmamaktadır. Buna bağlı olarak da maddelerin farklı gruplardan gelen öğrenciler için aynı şekilde çalışıp çalışmadığı yani DMF içerip içermediği ancak ÖSS'nin ardından yapılan analizlerle mümkün olmaktadır. Bu çalışma ile ÖSS'yi oluşturan maddelerde cinsiyete göre DMF'ye neden olabilecek özelliklerin belirlenmesi ile sonraki yıllarda DMF içermesi olası maddelerden kaçınılması ve ÖSS'nin niteliğinin artırılması amaçlanmıştır.

#### *Yöntem*

##### *Çalışma Grubu*

ÖSS'nin hedef kitlesinin lise son sınıf öğrencileri olması sebebiyle, çalışma 2005 ÖSS'ye giren tüm lise son sınıf öğrencileri (599,330) üzerinde yürütülmüştür. Bu öğrencilerin %46'sını kız (273,419), %54'ünü erkek (325,911) öğrenciler oluşturmaktadır.

##### *Veri Toplama Araçları*

Çalışmada DMF analizi için, kız ve erkek öğrencilerin ÖSS'deki her bir maddeye verdikleri cevaplar, yanlılık için ise cinsiyete göre DMF içeren maddelerle oluşturulan anket kullanılmıştır.

2005 ÖSS, tek basamaklı olup her biri 90 madde içeren, sözel ve sayısal olarak adlandırılan iki bölümden oluşmaktadır. Sözel bölüm Türkçe ve sosyal bilimler (tarih, coğrafya, felsefe) maddelerini, sayısal bölüm de matematik ve fen bilimleri (fizik, kimya biyoloji) maddelerini içermektedir. Çalışmada, bölümleri oluşturan konular göz önüne alınarak dört alt test oluşturulmuştur. 2005 ÖSS kılavuzunda; Türkçe alt testinin "Türkçeyi kullanma gücünü", sosyal bilimler alt testinin "Sosyal bilimlerdeki temel kavram ve ilkelerle düşünme gücünü", matematik alt testinin "Matematiksel ilişkilerden yararlanma gücünü" ve fen bilimleri alt testinin "Fen bilimlerindeki temel kavram ve ilkelerle düşünme gücünü" ölçmeyi amaçladığı belirtilmiştir (www.osym.gov.tr). Sözel bölüm madde numaraları Türkçe (1-45), tarih (46-64), coğrafya (65-80) ve felsefe (81-90); sayısal bölüm madde numaraları ise matematik (1-45), fizik (46-64), kimya (65-78) ve biyoloji (79-90) şeklindedir. Türkçe, sosyal bilimler, matematik ve fen bilimleri alt testlerinin güvenilirlik değerleri (Cronbach Alpha) sırası ile 0.93, 0.95, 0.96 ve 0.97 olarak hesaplanmıştır. Her bir alt test, beş seçenekli çoktan seçmeli maddelerden oluşmaktadır.

Çalışmada, DMF gösteren maddelerde yanlılık olup olmadığı konusunda uzman görüşlerini almak için bir anket geliştirilmiştir. Anket, literatür taraması ile DMF kaynaklarını belirlemeyi amaçlayan çalışmaların sonuçları göz önüne alınarak oluşturulmuştur. (ETS, 2009; Mendes-Barnett ve Ercikan, 2006; Yıldırım ve Berberoğlu, 2009; Zumbo ve Gelin, 2005). Ankette, analizler sonucunda DMF içerdiği belirlenen maddelere yer verilmiş ve uzmanlara öncelikle maddenin hangi grubun lehine çalışmış olabileceği sorusu yöneltilmiştir. Ardından maddelerin yanlı

olmasına sebep olabilecek bazı durumlar listelenmiş ve 12 ölçme-değerlendirme uzmanından, yanlılığın neden kaynaklandığını belirtmeleri istenmiştir. DMF çıkan maddelere göre o alanda lisans eğitimi almış ölçme-değerlendirme uzmanlarına başvurulmuştur. Uzmanlardan ikisi tarih, ikisi coğrafya, biri felsefe, ikisi Türkçe, dördü matematik ve biri fizik lisansına sahiptir. Anket örneği ekte verilmiştir.

Çalışma için gerekli olan ÖSS verileri ÖSYM Araştırma Geliştirme Birimi tarafından temin edilmiştir.

#### *Verilerin Analizi*

Verilerin tek boyutluluk varsayımını karşılayıp karşılamadığını incelemek üzere tüm testi oluşturan 180 maddeye temel bileşenler faktör analizi yöntemi uygulanmıştır. Varimax rotasyon yöntemi kullanılarak elde edilen döndürülmüş bileşenler matrisi incelendiğinde; fen bilimleri, sosyal bilimler, matematik ve Türkçe alt testini oluşturan maddelerin sırasıyla 1., 2., 3. ve 4. faktörlere yüklenmiş oldukları gözlenmiştir. Döndürülmüş bileşenler matrisine göre ait oldukları alt teste yüklenmeyen maddeler analiz dışı bırakılmışlardır. Türkçe alt testi 38, sosyal bilimler alt testi 40, matematik alt testi 41 ve fen bilimleri alt testi 45 madde üzerinden analiz edilmiştir. Bu maddelerden oluşan alt testlere MH ve LR yöntemleri ile DMF analizi yapılmıştır. Analiz sonucunda belirlenen ölçüte göre DMF içeren maddelerin yanlı olup olmadığının saptanması amacıyla uzman görüşü alınmıştır.

#### *Mantel-Haenszel Yöntemi*

Mantel-Haenszel yöntemi,  $\chi^2$  istatistiğine dayanan bir yöntem olup bu yöntemde odak ve referans gruplarındaki bireyler, gösterdikleri performansa göre denkleştirilir. Denkleştirme öğrencilerin toplam puan sayılarının eşit olmasına dayanılarak yapılır. Denkleştirilmiş iki grubun performansı daha sonra olasılık oranı değerleri dikkate alınarak karşılaştırılır. Olasılık oranı, bir gruptaki bireylerin bir maddeye doğru cevap verme olasılığının diğer gruptaki bireylerden yüksek olup olmadığını belirler. İstatistiksel olarak daha anlamlı sonuçlar elde etmek amacıyla olasılık oranı delta ölçeğine çevrilerek kullanılmaktadır. Bu istatistik Mantel-Haenszel delta farkı (D) olarak adlandırılmaktadır (Zieky, 1993). MH yöntemi ile sadece tek biçimli DMF tespit edilebilmektedir.

MH istatistiği sonucunda test maddeleri içerdikleri DMF oranında, üç kategoriye ayrılmıştır.

A kategorisi,  $|D| < 1$  DMF yoktur veya göz ardı edilebilecek düzeydedir,

B kategorisi,  $1 < |D| < 1.5$  orta düzeyde DMF vardır,

C kategorisi,  $|D| \geq 1.5$  yüksek düzeyde DMF vardır (Zieky, 1993).

#### *Lojistik Regresyon*

Lojistik Regresyon yöntemi sürekli olmayan bir bağımlı değişkeni sürekli bağımsız değişkenler yardımıyla tahmin etmeyi amaçlayan istatistiksel bir modeldir. LR yöntemi, DMF belirleme çalışmalarına ilk olarak Swaminathan ve Rogers (1990) tarafından uygulanmıştır. Bu yöntemle hem tek biçimli hem de tek biçimli olmayan DMF tespit edilebilmektedir. DMF belirlemeye yönelik çeşitli yöntemler arasında en etkili ve önerilen yöntemlerden biridir (Camilli ve Shepard, 1994).

DMF analizinde değişkenler, modele hiyerarşik olarak sırayla dahil edilirler. Model 1'de sınırlanacak değişken, yani toplam puan modelde yerini alır. Model 2'de, grup değişkeni (cinsiyet) eklenir ve üçüncü modelde ise önceki değişkenlere ilave olarak etkileşim değişkeni modele alınır. Zumbo (1999), Swaminathan ve Rogers'ın (1990), serbestlik derecesi 2 olan  $\chi^2$  testi üzerinde yoğunlaştıklarını ama üçüncü model ile birinci modelden elde edilen  $R^2$  değerleri farkının  $\chi^2$  testine karşılık geldiğini savunmuştur. Bu değer hem tek biçimli hem de tek biçimli



olmayan DMF'yi içermektedir. Bu çalışmada LR analizi sonucunda elde edilen Negelkerke  $R^2$  değer farkları göz önüne alınmıştır.

MH yönteminde olduğu gibi Lojistik Regresyonda da test maddeleri içerdikleri DMF oranında, kategorilere ayrılmıştır. Bunun için standartlaştırılmış regresyon katsayılarından yararlanılmakta ve bu katsayıların değerlendirilmesinde çeşitli ölçütler kullanılmaktadır (Jodoin ve Gierl, 2001; Zumbo ve Thomas, 1996). Pitoniak, Cook, Cline ve Cahalan-Laitusis (2006), yaptıkları çalışmada MH ve LR yöntemlerini karşılaştırmışlar, iki yöntemde DMF içerdiği belirlenen maddeler arasında tutarsızlıklar olduğunu ve ayrıca LR için belirlenen kategorilendirme sisteminin DMF'nin gerçek miktarını belirlemede yeterince hassas olmadığını ifade etmişlerdir.

Kullanılan farklı kategorilendirme sistemlerinden farklı sonuçlar elde edilmesi, yöntemlerin örneklem büyüklüğünden etkileniyor olması ve belirlenen kategorilendirme sistemlerinde görüş birliği sağlanamaması sebebiyle bu çalışmada daha hassas bir kategorilendirme sisteminin kullanılmasına karar verilmiştir (Hidalgo ve Lopez-Pina, 2004; Jodoin ve Gierl, 2001). Böylece DMF içerdiği halde belirlenen ölçütün altında değer aldığı için gözden kaçırılacak maddelerin de incelenmesi amaçlanmıştır. MH yöntemiyle C kategorisinde, LR yöntemiyle  $0.020 \leq R^2$  olan maddeler ile hem MH yöntemiyle B kategorisinde hem de LR yöntemiyle  $0.010 \leq R^2$  olan maddeler de DMF içeriyor olarak kabul edilecek ve bu maddeler için uzman kanısına başvurulacaktır.

Bu çalışmada, faktör analizi ve betimsel istatistiklerde SPSS, MH yöntemi ile DMF belirlenirken Ezdiff programı (Waller, 2005), LR yöntemi ile DMF belirlenirken Zumbo (1999)'nun yazdığı SPSS betik kullanılmıştır.

### Bulgular ve Yorum

Araştırmanın ilk aşamasında MH yöntemi ile DMF analizi yapılarak her bir alt test için elde edilen bulgular Tablo 1' de verilmiştir.

Tablo 1.

*MH Yöntemi ile DMF Analizi Sonuçları*

	Düzyey	Kız Öğrenciler Lehine Çalışan Maddeler	Erkek Öğrenciler Lehine Çalışan Maddeler
Türkçe	B	-	-
	C	-	-
Sosyal Bilimler	B	-	55, 70
	C	84, 86, 87, 88, 89, 90	61
Matematik	B	1, 2, 18	34, 39, 40
	C	-	-
Fen Bilimleri	B	-	50, 54, 59
	C	-	49, 63

Türkçe alt testinde 22 madde kız öğrenciler lehine, 16 madde de erkek öğrenciler lehine çalışmıştır. Türkçe alt testini oluşturan maddelerin tamamı A düzeyinde DMF içermektedir; yani DMF yok ya da göz ardı edilebilecek düzeydedir. Belirlenen ölçüte göre hiçbir Türkçe maddesi cinsiyete göre DMF içermemektedir.

Sosyal bilimler alt testinde 13 madde kız öğrenciler lehine, 27 madde de erkek öğrenciler lehine çalışmıştır. Bu maddelerden ikisi B, yedisi C düzeyinde DMF içermektedir. Felsefe maddeleri kız öğrenciler, tarih ve coğrafya maddeleri ise erkek öğrenciler lehine çalışmaktadır.

Matematik alt testinde 18 madde kız öğrenciler lehine, 23 madde de erkek öğrenciler lehine çalışmıştır. Bu maddelerden altısı B düzeyinde DMF içermektedir. B düzeyinde DMF içeren maddelerin üçü erkek öğrenciler lehine çalışan geometri maddeleridir. Kız öğrenciler lehine çalışan maddeler ise iki işlem maddesi ile bir sözel problem maddesidir. Kız öğrenciler

işlem maddelerinde, erkek öğrenciler ise özellikle geometri maddelerinde daha avantajlı görünmektedirler.

Fen bilimleri alt testinde 20 madde kız öğrenciler lehine, 25 madde de erkek öğrenciler lehine çalışmıştır. Fen bilimleri alt testindeki üç madde B düzeyinde, iki madde ise C düzeyinde DMF içermektedir. Genel olarak, kimya ve özellikle biyoloji maddeleri kız öğrenciler, fizik maddeleri ise erkek öğrenciler lehine çalışmıştır. MH yöntemine göre üçü B, ikisi C düzeyinde DMF içerdiği belirlenen fizik maddeleri erkek öğrenciler lehine çalışmıştır.

Araştırmanın ikinci aşamasında LR yöntemi ile DMF analizi yapılarak her bir alt test için elde edilen bulgular Tablo 2' de verilmiştir. Öncelikle tek biçimli olmayan DMF'yi belirlemek amacıyla ikinci model ile birinci model arasındaki Negelkerke R<sup>2</sup> farkları incelenmiş, ancak çoğu madde için modeller arasında fark olmadığı gözlenmiştir. Bu sebeple tabloda sadece, hem tek biçimli hem de tek biçimli olmayan DMF içeren, üçüncü model ile birinci modelden elde edilen Negelkerke R<sup>2</sup> değer farkları ( $\Delta R^2$ ) verilmiştir.

Tablo 2.

*LR Yöntemi ile DMF Analizi Sonuçları*

No	Türkçe $\Delta R^2$	No	Sosyal Bil. $\Delta R^2$	No	Matematik $\Delta R^2$	No	Fen Bilimleri $\Delta R^2$
1	.000	46	.001	1	.010	46	.002
2	-	47	.002	2	.004	47	.004
3	.007	48	.001	3	.006	48	.001
4	-	49	.004	4	.000	49	.007
5	.000	50	.006	5	.004	50	.011
6	.002	51	.000	6	.001	51	.001
7	-	52	.003	7	.001	52	.001
8	.001	53	.000	8	.000	53	.002
9	.002	54	.009	9	.004	54	.008
10	-	55	.004	10	.001	55	.004
11	.001	56	.002	11	-	56	.004
12	.001	57	.000	12	-	57	.000
13	.001	58	.007	13	.001	58	.002
14	.003	59	.001	14	.003	59	.004
15	.002	60	.000	15	.005	60	.003
16	.003	61	.027	16	.000	61	.002
17	.002	62	.001	17	.002	62	.002
18	.003	63	.000	18	.005	63	.007
19	.005	64	.000	19	.002	64	.006
20	-	65	.000	20	.001	65	.001
21	.004	66	-	21	.001	66	.001
22	-	67	.002	22	.001	67	.000
23	-	68	-	23	.002	68	.001
24	.001	69	.001	24	.004	69	.001
25	.001	70	.007	25	.004	70	.005
26	.001	71	-	26	.002	71	.001
27	.000	72	.002	27	.001	72	.002
28	.002	73	.000	28	.004	73	.001
29	.000	74	.001	29	.001	74	.001
30	.000	75	.001	30	.001	75	.003
31	.004	76	.000	31	.001	76	.002
32	.000	77	.002	32	.000	77	.002
33	.001	78	-	33	.004	78	.002
34	.001	79	.000	34	.001	79	.002
35	.004	80	.001	35	.001	80	.000
36	.001	81	.005	36	.001	81	.001
37	.002	82	.004	37	.002	82	.001
38	.001	83	.004	38	.011	83	.001
39	.004	84	.023	39	.010	84	.002
40	.001	85	-	40	.001	85	.001
41	.000	86	.019	41	.002	86	.002
42	.000	87	.018	42	-	87	.003
43	.001	88	.017	43	.002	88	.001
44	.002	89	.015	44	.002	89	.000
45	.000	90	.022	45	.002	90	.000

Tablo 2'den görüldüğü gibi Türkçe alt testi için elde edilen LR sonuçları da MH sonuçlarını destekler nitelikte olup  $\Delta R^2$  değeri hiçbir madde için belirlenen 0.010 değerinin üzerine değildir; yani Türkçe alt testinin tamamında DMF yoktur ya da göz ardı edilebilecek seviyededir. Diğer alt testler için elde edilen LR sonuçlarına göre, sosyal bilimler alt testinde yedi, matematik alt testinde üç ve fen bilimleri alt testinde bir maddenin  $\Delta R^2$  değeri 0.010'dan büyüktür.

MH ve LR analizi sonuçlarına dayanılarak DMF içerdiği belirlenen madde numaraları Tablo 3' te verilmiştir.

Tablo 3.

*MH ve LR Yöntemleriyle DMF İçerdiği Belirlenen Maddelerin Karşılaştırılması*

	MH		LR	
	B	C	$\leq 0.10$	$\leq 0.20$
Türkçe	-	-	-	-
Sosyal Bilimler	55, 70	61, 84, 86, 87, 88, 89, 90	86, 87, 88, 89	61, 84, 90
Matematik	1, 2, 18, 34, 39, 40	-	1, 39, 40	-
Fen Bilimleri	50, 54, 59	49, 63	50	-

Türkçe alt testindeki hiçbir madde DMF içermemektedir.

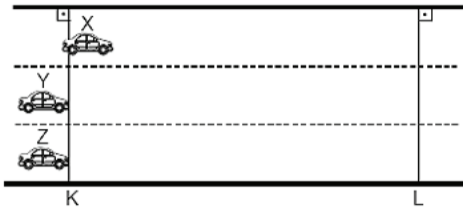
Sosyal bilimler alt testinde, belirlenen ölçüte göre DMF içeren maddeler 61 nolu tarih maddesi ile 84, 86, 87, 88, 89, 90 nolu felsefe maddeleridir. Tarih maddesi erkek öğrenciler lehine, felsefe maddeleri ise kız öğrenciler lehine DMF içermektedir.

Matematik alt testinde, belirlenen ölçüte göre DMF içeren maddeler 1 nolu işlem maddesi ile 39 ve 40 nolu geometri maddeleridir. İşlem maddesi kız öğrenciler, geometri maddeleri ise erkek öğrenciler lehine çalışmaktadır.

Fen bilimleri alt testinde belirlenen ölçüte göre DMF içerdiği belirlenen maddeler 49, 50 ve 63 nolu fizik maddeleridir. Her üç madde de erkek öğrenciler lehine çalışmaktadır.

Araştırmanın son aşamasında yukarıda belirlenen on üç madde için uygulanan ankette; uzmanlar, bazı konu alanlarındaki maddelerin öğrencilerin cinsiyetine göre farklı çalışmasının beklendiğini, ancak bu durumun yanlılık olarak değerlendirilemeyeceğini belirtmişlerdir. Anket sonucunda uzmanlar, sadece 49 nolu fizik maddesinin yanlı olarak kabul edilebileceğine dair ortak görüş bildirmişlerdir. Bu maddenin, otomobiller ve hız konusunun erkek öğrencilerin daha yakın olduğu konular olması sebebiyle erkek öğrenciler lehine çalışmasının beklendiğini ve yanlı olarak kabul edilebileceğini bildirmişlerdir. 49 nolu fizik maddesi ile kız ve erkek öğrencilerin bu maddeyi doğru cevaplama oranı grafiği aşağıda verilmiştir.

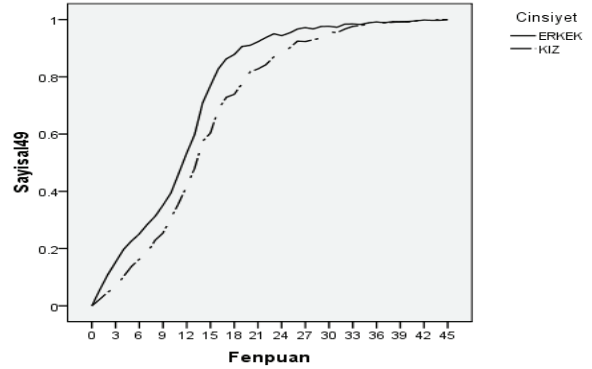
49.



Şekildeki doğrusal yolda X, Y, Z otomobilleri değişmeyen hızlarla KL yönünde gitmektedir. K çizgisinden önce X, sonra da Y ile Z aynı anda; L çizgisinden de önce Y, sonra da X ile Z aynı anda geçiyor.

X, Y, Z otomobillerinin hızlarının büyüklükleri sırasıyla  $V_x$ ,  $V_y$ ,  $V_z$  olduğuna göre, bunlar arasındaki ilişki nedir?

- A)  $V_x = V_y < V_z$                       B)  $V_y = V_z < V_x$   
 C)  $V_z < V_x = V_y$                       D)  $V_x < V_y < V_z$   
 E)  $V_x < V_y < V_z$





### Sonuç ve Tartışma

Skolastik yetenekteki farklılıkların ölçme aracından bağımsız olarak gerçek cinsiyet farklılığından kaynaklandığına dair tutarlı ve genel bir kanı bulunmaktadır. Kız ve erkek öğrencilerin yapısal farklılıklarına ve ilgilerine bağlı olarak farklı alanlarda daha başarılı oldukları bir gerçektir.

Cinsiyeti konu edinen DMF çalışmalarında genel olarak sözel derslerde kız öğrencilerin, sayısal derslerde ise erkek öğrencilerin daha başarılı olduğu sonucuna varılmıştır (Azen, Bronner, Gafni 2002; Coley 2001). Bu durumun öğrencilerin testlere göre avantajlı dezavantajlı olmalarına yol açabileceği belirtilmektedir. (Azen, Bronner ve Gafni, 2002). Bu çalışma da sosyal bilimlerden özellikle felsefe alt testinde kız öğrencilerin, fen bilimlerinde özellikle fizik alt testinde erkek öğrencilerin daha başarılı oldukları ortaya çıkmıştır. Ancak, matematik ve fen bilgisi testlerinde kızlar ve erkekler arasında yanlılığa rastlanmadığına ilişkin bulgular da yer almaktadır. (Gök, Kalecioğlu ve Doğan, 2010).

Cinsiyete göre alt testler göz önüne alındığında, Türkçe alt testini oluşturan maddelerin DMF içermediği sonucuna varılmıştır. Sosyal bilimler alt testinde bir tarih maddesinin erkek öğrenciler lehine altı felsefe maddesinin kız öğrenciler lehine DMF içerdiği gözlenmiştir. Bulunan sonuçlar konu alanları göz önüne alındığında, benzer nitelikteki çalışmalar tarafından da desteklenmektedir (Le 1999; Yurdugül ve Aşkar 2004; Zwick ve Ercikan 1989).

Uzmanlar tarafından bazı maddelerin öğrencilerin cinsiyetine göre farklı çalışmasının beklendiği, ama bu durumun yanlılık olarak değerlendirilemeyeceği belirtilmiştir. Bununla birlikte, kız öğrencilerin felsefede, matematik işlemlerinde ve biyolojide avantajlı oldukları, erkek öğrencilerin tarihte savaş ve siyaset içerikli konularda, fizikte özellikle hareket, hız ile üç boyutlu algı gerektiren aynalar konusunda ve matematikte problem çözme ve geometri konularında avantajlı oldukları gözlenmiştir.

Yanlı olarak belirlenen fizik maddesi hız konusundan olup, erkek öğrencilerin ilgi duydukları bilinen otomobillerle ilgili bir de diyagram içermektedir. Zenisky, Hambleton ve Robin (2003), görsel-uzamsal zekâ gerektiren (grafikler, tablolar, diyagramlar, haritalar içeren) maddelerin erkek öğrenciler lehine çalıştığını belirlemişlerdir.

### Öneriler

Ülkemizde yapılan büyük ölçekli sınavlar üzerinde, okul türü, coğrafi bölge, sosyoekonomik düzey gibi farklı değişkenleri göz önüne alan DMF çalışmaları yapılabilir. Farklı yıllara ait veriler üzerinde çalışılıp, DMF içerdiği belirlenen maddelerin kapsamları, madde formları ve ölçtükleri bilişsel alanlar bakımından karşılaştırılabilir. Böylece bir teste ait maddeler arasından DMF içerenlerini belirlemekten öteye gidilerek yıllar boyunca kullanılan testlerde DMF içeren ve içermeyen pek çok maddenin karşılaştırılması sonucunda, DMF'nin olası sebepleri istatistiksel olarak test edilerek örüntüler belirlenebilir. Gerekirse konu alanı uzmanları ve soru yazarları bu konuda eğitilebilir. Bu çalışmalar başta ÖSS olmak üzere ÖSYM tarafından hazırlanan sınavların geçerliğinin artırılmasına katkı getirecektir.

ÖSS'deki maddelerin DMF gösterip göstermediği madde tepki kuramına dayalı yöntemlerle ve SIBTEST ile incelenebilir. Ayrıca, yöntemlerden hangisinin DMF belirlemede daha etkili olduğunu ortaya çıkarmak için simülasyon çalışmaları yapılmalıdır.

### Ek 1-Anket

Yanlılık, bir maddenin farklı gruplardan gelen benzer yetenek düzeyindeki bireyler tarafından doğru yanıtlanma olasılığının farklılık göstermesidir. Yapılan analizlere sonucunda, farklı cinsiyetten gelen benzer yetenek seviyesindeki öğrencilerin aşağıdaki maddeleri doğru yanıtlama oranları arasında farklılıklar bulunmuştur. Bu durum maddenin yanlı olabileceğine

dair bir kanıttır. Aşağıda yanlılığın olası nedenleri listelenmiştir. Lütfen her madde için ilgili yerleri doldurunuz. Yanlılığın listelenen nedenler dışında bir nedenden kaynaklandığını düşünüyorsanız diğer kısmına ekleyiniz.

Sizce madde hangi cinsiyet grubunun lehine çalışmaktadır?

Kız  Erkek

Bunun nedeni aşağıdakilerden hangisi olabilir?

Maddede kullanılan sözcük ya da ifade nedeni ile

Evet  Hayır

Sözcüğü, ifadeyi yazınız:

Cinsiyete bağlı olarak bir grubun maddenin içeriğine aşına olması nedeni ile

Evet  Hayır

İçeriği yazınız:

Cinsiyet gruplarının daha yakın olduğu konuların soru kapsamında içerilmesi nedeni ile

Evet  Hayır

Konuyu yazınız:

Maddede kullanılan tarzın cinsiyet özelliklerine göre farklı olması nedeni ile (sözel anlatımın uzun olması, grafik içermesi vs. gibi)

Evet  Hayır

Soru tarzını yazınız:

Diğer:

#### Kaynakça

- Azen, R., Bronner, S., & Gafni, N. (2002). Examination of Gender Bias in University Admissions. *Applied Measurement in Education*, 15, 75-94.
- Balluerka, N., Gorostiaga, A., Gomez-Benito, J. & Hidalgo, M. D. (2010). Use of Multilevel Logistic Regression to Identify the Causes of Differential Item Functioning. *Psicothema*, 22, 4, 1018-1025.
- Benito J.G. & Ara M.J.N. (2000). A Comparison of X<sup>2</sup>, RFA and IRT Based Procedures in the Detection of DIF, *Quality & Quantity*, 34, 17-31.
- Berberoğlu, G. (1996). The University Entrance Examinations in Turkey. *Studies in Educational Evaluation*, 22, 4, 363-373.
- Camilli, G. & Shepard, L.A. (1994). *Methods for Identifying Biased Test Items*. Sage Publications, California.
- Coley, R. J. (2001). Differences in the Gender Gap: Comparison Across Racial/Ethnic Groups in Education and Work. Policy Information Report, Educational Testing Service.
- Doolittle, A. E. & Cleary, T. A (1987). Gender-Based Differential Item Performance in Mathematics Achievement Items. *Journal of Educational Measurement*, 24, 2, 157-166.
- Educational Testing Service, (2009). *ETS Guidelines for Fairness Review of Assessments*. Princeton, NJ: ETS.
- Gök, B., Kalecioğlu, H. Doğan, N. (2010). Değişen Madde Fonksiyonu Belirlemede Mantel - Haenzel ve Lojistik Regresyon Tekniklerinin Karşılaştırılması. *Eğitim ve Bilim*, 35 (156), 3-16.
- Haladyna, T. M. (1997). *Writing Test Items to Evaluate Higher Order Thinking*. Allyn & Bacon.
- Hambleton, R.K., Swaminathan, H. & Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Sage Publications, California.
- Harris, A. M. & Carlton, S. T. (1993). Patterns of Gender Differences on Mathematics Items on the Scholastic Aptitude Test. *Applied Measurement in Education*, 6, 2, 137-151.

- Hidalgo, M. D. & Lopez-Pina, J. A. (2004). Differential item functioning detection and 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 Using an Effect Size Measure with the Logistic Regression Procedure for DIF Detection. *Applied Measurement in Education*, 14, 4, 329-349.
- Kristanjanjonn E., R. Aylesworth, I. McDowell & B.D. Zumbo (2005). A Comparison of Four Methods for Detecting Differeantial Item Functioning In Ordered Response Model. *Educational and Psychological Measurement*. 65,6. 935-953.
- Le, V. (1999). Identifying Differential Item Functioning on the NELS: 88 History Achievement Test. Center for the Study of Evaluation Technical Report. (ED443879)
- Mendes-Barnett, S. & Ercikan, K. (2006). Examining sources of gender DIF in mathematics assessments using a confirmatory multidimensional model approach. *Applied Measurement in Education*, 19,4, 289-304.
- Oshima T. C. & Morris, S. B. (2008). Raju's Differential Functioning of Items and Tests (DFIT). *Educational Measurement: Issues and Practice*, 27, 3, 43-50.
- Öğretmen, T. (1995). Differential Item Functioning Analysis of the Verbal Ability Section of the First Stage of the University Entrance Examination in Turkey. Yayınlanmamış Yüksek Lisans Tezi. Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Pitoniak, M., Cook, L., Cline, F. & Cahalan-Laitusis, C. (2006). *Using Differential Item Functioning to Investigate the Impact of Accomodations on the Scores of Students with Disabilities on English-Language Arts Assessments*. Educational Testing Service, NCME Presentation.
- Swaminathan, H. & Rogers, H. J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement*, 27, 4, 361-370.
- Yenal, E. (1995). Differential Item Functioning Analysis of the Quantitative Ability Section of the First Stage of the University Entrance Examination in Turkey. Yayınlanmamış Yüksek Lisans Tezi. Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Yıldırım, H. H., Berberoglu, G. (2009). Judgmental and Statistical DIF Analyses of the PISA-2003 Mathematics Literacy Items. *International Journal of Testing*, 9, 108-121.
- Yurdugül, H. & Aşkar, P. (2004). Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı'nın Cinsiyete Göre Madde Yanlılığı Açısından İncelenmesi. *Eğitim Bilimleri ve Uygulama*, 3, 5, 3-20.
- Waller, N. G. (2005) EZDIF: A computer Program for Detecting Uniform and Non uniform Differential Item Functioning with the Mantel-Haenszel and Logistic Regression Procedures.
- Zenisky, A.L., Hambleton, R.K. ve Robin, F. (2003). DIF Detection and Interpretation in Large-Scale Science Assessments: Informing Item Writing Practices. *Educational Assessment*, 9 (1&2), 61-78.
- 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., & Thomas, D. R. (1996). *A measure of DIF effect size using logistic regression procedures*. Paper presented at the National Board of Medical Examiners, Philadelphia.
- 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., & Gelin, M.N. (2005). A Matter of Test Bias in Educational Policy Research: Bringing the Context into Picture by Investigating Sociological Community Moderated (or Mediated) Test and Item Bias. *Journal of Educational Research and Policy Studies*, 5, 1-23.