



Değişen Madde Fonksiyonu Belirlemede MTK-Olabilirlik Oranı, Ordinal Lojistik Regresyon ve Poly-Sıbtest Yöntemlerinin Karşılaştırılması

*Arş.Grv. Çiğdem Arıkan Akın
Hacettepe Üniversitesi-Türkiye
cakinarikan@hacettepe.edu.tr*

Özet

Bu çalışmada PISA 2012 kapsamında uygulanan öğrenci anketinde yeralan Matematik Öğrenme bölümü altındaki matematik çalışma etiği tutum maddelerinin cinsiyete göre değişen madde fonksiyonu gösterip göstermediği incelenmiştir. DMF analizleri Madde Tepki Kuramı-Olabilirlik Oranı (MTK-OO), Ordinal Lojistik regresyon (OLR) ve Poly-SIBTEST yöntemleriyle gerçekleştirilmiştir. Çalışmanın örneklemini, matematik çalışma etiği tutum maddelerine yanıt veren 3217 bireyden SPSS programı aracılığıyla random olarak seçilen 1000 (500 kadın, 500 erkek) birey oluşturmaktadır. Verilerle ilgili madde analizleri MTK ile yapılmıştır. Cinsiyete göre DMF analizlerinde Poly-SIBTEST yöntemine göre 3. Maddenin yüksek düzeyde erkekler lehine, MTK-OO yöntemine göre 3. Maddenin orta düzeyde ve OLR yöntemine göre 3. Maddenin yüksek düzeyde tek biçimli olmayan DMF gösterdiği görülmüştür.

Anahtar Sözcükler: PISA, Değişen Madde Fonksiyonu, MTK-OO, Poly-SIBTEST, OLR

GİRİŞ

Türkiye'nin de üyesi olduğu Ekonomik İşbirliği ve Kalkınma Teşkilatı-OECD (Organisation for Economic Cooperation and Development) tarafından düzenlenen PISA (Uluslararası Öğrenci Başarılarını Değerlendirme Programı) ve TIMSS (Uluslararası Matematik ve Fen Eğilimleri Araştırma) sınavları uluslararası düzeydeki en kapsamlı eğitim araştırmalarındandır. PISA uygulamasına 2012 yılında 65 ülkeden 28 milyon öğrenciyi temsilen yaklaşık 510 bin öğrenci katılmıştır (MEB, 2013). 2000 yılında ilk olarak uygulanan PISA'ya, Türkiye 2003 yılından bu yana katılmaktadır. PISA uygulaması, 15 yaş öğrencilerinin okuma, matematik ve fen konu alanlarındaki bilgi ve becerilerinin yanı sıra uygulanan çeşitli anketlerle öğrencilerin öğrenme, motivasyon, ilgi, ev ve okul ortamı, aileleri hakkında da bilgi toplamaktadır (MEB, 2013). PISA uygulamasında okuma, matematik ve fen konu alanlarından her dönem birine ağırlık verilmekte olup; 2012 PISA'da ağırlık alanı 2003 yılında olduğu gibi matematiktir. PISA uygulamalarında öğrencilere fen, matematik ve okuma alanlarının bulunduğu 13 farklı kitapçıktan biri ve bir öğrenci anketi verilmektedir.

PISA'nın temel amacı, temel eğitimin sonunda öğrencilerin yetişkinliklerinde gereksinim duyacakları temel becerileri ne kadar edinebildiğini ölçmektir (ERG, 2010). PISA'da yeralan konu alanları için farklı soru türleri bulunmaktadır. Bu





soruların yaklaşık yarısı çoktan seçmeli, diğerleri ise karmaşık çoktan seçmeli ve açık uçludur. PISA uygulamalarında kullanılan bir çok anketin ölçekleme işlemleri Madde Tepki Kuramı (MTK) çerçevesinde yapılmaktadır. MTK, bireyin gözlenen test performansı ile gözlenemeyen yetenek arasındaki ilişkiyi matematiksel olarak açıklar (Embretson ve Reise, 2000). MTK'ında iki kategorili ve çok kategorili maddeler için bir çok model bulunmaktadır. İki kategorili maddeler için, 1 parametrelili Lojistik Model (1PLM), Rasch Modeli, 2 PLM ve 3 PLM; çok kategorili maddeler için Aşamalı Tepki Modeli (Graded Response Model), Kısmi puan modeli (Partial Credit Model (PCM) Adlandırılmalı Tepki Modeli (Nominal Response Model-NRM), Genelleştirilmiş Kısmi Puan Modeli (Generalized Partial Credit Model (GPCM) örnek olarak verilebilir (Hambleten ve Swaminathan, 1995; Embretson ve Reise, 2000).

Bu çalışmadaki çok kategorili veriler Aşamalı Tepki Modeli (ATM) ile kalibre edildiğinden bu modelle ilgili bilgi verilecektir. Samejima (1969) tarafından geliştirilen Aşamalı Tepki Modeli, iki veya daha fazla sıralı kategoriye sahip maddeler için uygundur. 2 PLM'in genelleştirilmiş hali olan model, dolaylı MTK modeli olarak düşünülebilir. Çünkü, belirli bir kategori için bireyin cevaplama olasılığı iki aşamada gerçekleşir (Hambleten ve Swaminathan, 1995; Embretson ve Reise, 2000). ATM'de her madde için bir madde eğim parametresi (α_1) ve $j=1, \dots, m$ kategoriler arasında eşik parametresi (β_{ij}) vardır. Eşik parametre sayısı, kategori sayısının bir eksiğine eşittir. İki aşamada gerçekleşen ATM'nde ilk olarak (1) numaralı eşitlik yardımıyla her bir maddeye ait kategori eşik değerleri için işlem karakteristik eğrileri (OCC) hesaplanır.

$$P_{ix}^* = \frac{\exp[\alpha_1(\theta - \beta)]}{1 + \exp[\alpha_1(\theta - \beta)]} \quad (1)$$

İkinci aşamada ise (2) numaralı eşitlikle kategorik tepki eğrileri (CRC) hesaplanır. Kategorik tepki eğrileri belirli kategoride bireyin cevaplama olasılığını gösterir.

$$P_{ix}(\theta) = P_{ix}^*(\theta) - P_{i(x+1)}^*(\theta) \quad (2)$$

Genel olarak, eğim parametresi yükseldikçe, işlem karakteristik eğrisi dikleşir (Embretson ve Reise, 2000).

Dünyada birçok ülkede uygulanan ve sonuçlarıyla ülkelerin eğitim sistemlerinin güçlü ve zayıf yönlerini belirleyen, eğitim sistemleri için geleceğe dönük politikalar geliştirmesini sağlayan PISA sonuçlarının bireyler arasındaki farklılıkları ayırt edebilmesi ve minimum hata içermesi gerekmektedir. Yani, ölçme sonuçlarıyla ilgili güvenilirlik ve geçerlik kanıtlarının sunulması gereklidir. Geçerliğe karşı önemli tehditlerden biri madde ve test yanlılığıdır (Clauser ve Mazor, 1998). Aynı yetenek düzeyindeki bireylerin, testten veya testte yeralan maddelerden aynı





puanı alması beklenir. Ancak, testin koşullarından ya da maddenin bazı özelliklerinden dolayı testi alan aynı yetenek düzeyindeki alt gruplarda maddenin doğru cevaplandırılma olasılığı farklılaşmasına değişen madde fonksiyonu (DMF) denir (Zumbo, 2007; Finch ve French, 2007). DMF, aynı test puanına, aynı yetenek düzeyine sahip bir gruptaki bireylerin sistematik olarak belirli bir test maddesini çözme olasılığı bakımından farklılık göstermesi olarak tanımlanır (Doğan, Guerrero ve Tatsuoka, 2005). Hulin, Drasgow ve Parsons (1983), tutum maddelerinde DMF'yi; aynı yetenek düzeyinde farklı alt gruplardaki bireylerin, bir maddeye olumlu tutum gösterme olasılığının farklı olması olarak tanımlamışlardır (Akt. Johanson ve Dodeen, 2003). Tutum anketlerinde DMF'nin varlığı, ölçek sonuçlarının doğruluğunu olumsuz yönde etkiler.

DMF analizi için öncelikle cinsiyet, etnik köken, ırk, kırsal/kent yaşamı ve dil gibi değişkenler kullanılarak alt grupların belirlenmesi gereklidir. Fakat, testlerin belirtilen alt gruplara göre farklılaşmaması, bireyleri sadece yetenek düzeylerine göre ayırt etmesi gerekmektedir. Belirlenen alt gruplardan biri odak grup, diğeri ise referans grup olarak belirlenir. Odak grupta yeralan bireylerin performansları, referans grupta yeralan bireylerin performansı ile karşılaştırılır. Eğer belirli bir maddenin cevaplanma olasılığı belirlenen alt grup için diğer gruptan farklı ise, maddenin DMF gösterdiği söylenir (Doğan, Guerrero ve Tatsuoka, 2005). Maddeler tek biçimli ve tek biçimli olmayan DMF olarak iki şekilde DMF gösterirler. Tek biçimli DMF, tüm yetenek düzeylerinde maddenin doğru cevaplanma olasılığının bir gruba daha avantajlı olması, tek biçimli olmayan DMF de ise cevaplanma olasılığı farklı yetenek düzeylerinde farklı gruplar lehine işlemesi, avantaj göstermesidir (Zumbo, 1999; Finch ve French, 2007). DMF, maddenin yanlı olduğunun kanıtlarından biridir, ancak yeterli değildir. Eğer bir madde DMF gösteriyorsa, yanlı olup olmadığını belirlemek için amprik değerlendirmeler ve içerik analizi yapılmalı uzman görüşü alınmalıdır (Zumbo, 1999).

Bu çalışmada DMF belirleme yöntemlerinden SIBTEST (Stout, Roussos; 1995), Ordinal Lojistik Regresyon (Zumbo,1999) ve Madde Tepki Kuramına dayalı MTK-Olabilirlik Oranı (MTK-OO) (Thissen, 2001) kullanılmıştır.

MTK-Olabilirlik Oranı: Thissen, Steinberg ve Wainer (1993) tarafından geliştirilen bu yöntem, Madde Tepki Kuramı Modellerinden 3 Parametrelili Lojistik Model ve Aşamalı Tepki Modeli için DMF olabilirlik oranını hesaplar ve IRTLRF yazılımını kullanır (Thissen 2001). Bu yöntemde, maddelerin DMF içerip içermediğine ilişkin oluşturulan modellerin olabilirlik oranları karşılaştırılır. Bunun için sınırlandırılmış ve genişletilmiş modeller oluşturulur. Sınırlandırılmış modelde, testteki tüm maddelerin madde parametreleri eşit olduğuna ilişkin olabilirlik oranı hesaplanır (IIAllEqual). Bu modelde maddelerin hiçbirinde DMF olmadığı varsayılır. Genişletilmiş modelde ise, DMF incelenen i. maddenin parametrelerinin odak ve referans grupta farklı, diğer maddelerin parametrelerinin eşit olduğuna ilişkin olabilirlik oranı elde edilir (IIIItemNotEqual).

$G^2 = -2(\text{IIAllEqual} - \text{IIIItemNotEqual})$ değeri ile test edilen i. maddenin tüm parametreleri için DMF olup olmadığı test edilir. χ^2 dağılımı gösteren ve serbestlik





derecesi madde parametresi sayısı olan G2 değeri manidarsa, maddede DMF olduğu kabul edilir. Maddede DMF bulunduğu, sırasıyla madde parametrelerinin biri hariç tutularak teste devam edilir (Thissen, 2001).

IRTLRDIF yazılımı ikili, çoklu ve karma puanlanan testlerde kullanılabilir. Bu yazılım madde tepki modeli parametreleriyle ilgili hipotezlerin doğrudan test edilmesini gerektirdiğinden, ölçülen yapılar ile ilgili farklı ilişkilerden ve farklı güçlüklerden oluşan veya farklı tahmin oranlarından oluşan DMF'yi belirleyebilir. DMF analizi, referans ve odak grup için madde parametrelerinin eşit olduğu sınırlandırılmış duruma ait -2loglikelihood değeri ile analiz edilen i. maddenin eşit, diğer madde parametrelerinin kısıtlandığı duruma ait -2loglikelihood değeri arasındaki fark olabilirlik oranı (G2) olarak adlandırılır. Ki-kare dağılımı gösteren G2 değerine ait serbestlik derecesi, iki duruma ait madde parametrelerinin farkına eşittir (Thissen, 2001). G2 değerine ait ölçütler Greer (2004) tarafından aşağıdaki gibi verilmiştir.

$3,84 < G^2 < 9,4$ İhmal edilebilir düzeyde

$9,4 \leq G^2 < 41,9$ Orta düzeyde

$G^2 \geq 41,9$ Yüksek düzeyde

Poly-SIBTEST: SIBTEST, iki kategorili puanlanan maddelerde DMF belirlemek için Shelay ve Stout (1993) tarafından önerilen istatistiksel bir metottur. Çok boyutlu veriler için geliştirilmesine rağmen, tek boyutlu veriler için de kullanılmaktadır. Bu yöntem standartlaştırma yönteminden, 1. Tip hata kontrolünde regresyon temelli düzeltme kullanması yönünden farklılık göstermektedir. Buna ek olarak SIBTEST yardımıyla küçük bir soru öbeğinde veya birkaç soruluk bir ölçüde yer alan maddeler üzerinden de DMF analizi yapılabilir (Clauser ve Mazor, 1998).

Poly-SIBTEST ise Chang, Mazzeo ve Roussos (1995) tarafından çok kategorili puanlanan maddelerde DMF belirlemek için geliştirilmiştir. *Poly-SIBTEST* için birkaç varsayım gerekli olduğundan parametrik olarak kabul edilmemektedir. Referans ve odak grupların bağımsız olması, tek boyutluluk ve DMF'den bağımsız olmasıdır. Ancak madde tepki fonksiyonu veya θ dağılımına ilişkin kabul yoktur (Woods, 2011). *Poly-SIBTEST*, DMF miktarını β etki büyüklüğünü kullanarak belirler. β değerlerine ait belli ölçütler belirlenmiştir. Eğer $|\beta| < 0,059$ ise ihmal edilebilir düzeyde (A), $0,059 \leq |\beta| < 0,088$ aralığında değer almışsa orta düzeyde (B) ve $|\beta| \geq 0,088$ 'e eşit ve büyük değer almışsa yüksek düzeyde (C) DMF olduğu anlamına gelmektedir (Roussos ve Stout, 1996). Elde edilen pozitif β değeri referans grubu lehine, negatif değeri ise odak grup lehine maddenin DMF gösterdiğini belirtir.

Ordinal Lojistik Regresyon (OLR): Çok kategorili verilerde DMF gösteren maddelerin belirlenmesinde kullanılan yöntemlerden biri de ordinal lojistik regresyondur. Zumbo (1999) OLR yöntemini, hem ikili hem de sıralı verilerde kullanılabilir olması, etki büyüklüğünü hesaplayabilmesi ve uygulama sürecini kolaylaştırması nedeniyle tercih edilmesi gerektiğini belirtmiştir. OLR yöntemi, sıralı veriler için daha önce Miller ve Spray (1993) tarafından önerilmiştir, ancak Zumbo





(1999) bir adım daha atarak etki büyüklüğünü de hesaplamıştır. Zumbo, OLR yönteminde yığılmalı lojit modelini kullanır. Bu modele göre i bireyin j ve altındaki bir puan kategorisinde puan alma olasılığı aşağıdaki gibidir.

$$\text{logit}[P(Y_i \leq j)] = \alpha_j + b_1 \text{toplam} + b_2 \text{grup} + b_3(\text{tot} * \text{grup}) \quad (3)$$

Eşitlik 3'de, α_j modelin kesişim katsayısı, b_1 , b_2 ve b_3 ise modelin eğim katsayılarıdır.

OLR yönteminde ilk olarak toplam puan, daha sonra grup değişkeni, son olarak da toplam puan ve grup değişkeni etkileşiminin eklenmesiyle oluşan 3 modele ait ki-kare değerleri elde edilir. Üçüncü aşamadan elde edilen ki-kare değeri (serbestlik derecesi=3) ile 1. aşamadan elde edilen ki-kare değeri (serbestlik derecesi=2) arasındaki fark, 2 serbestlik derecesine ait ki-kare dağılımı ile karşılaştırılır, anlamlı ise DMF'nin varlığından söz edebiliriz. DMF'nin miktarı ise etki büyüklüğü olan R2 değeri ile hesaplanır. OLR yönteminde hem tek biçimli ve hem de tek biçimli olmayan DMF belirlenebilir. Tek biçimli DMF'yi belirlemek için toplam puan değişkeninin modellendiği ilk aşama (model 1) ile toplam puan değişkeni ve grup değişkeninin modellendiği ikinci aşamadaki (model 2) R2 değerleri arasındaki fark hesaplanır. Tek biçimli olmayan DMF için ise toplam puan değişkeni ve grup değişkeninin modellendiği model 2 ile toplam puan değişkeni, grup ve etkileşim değişkenlerinin modellendiği üçüncü aşama (model 3) için elde edilen R2 değerleri arasındaki fark hesaplanır (Zumbo, 1999). DMF miktarını yorumlamada Zumbo ve Thomas (1996) tarafından önerilen ölçütler aşağıda yer almaktadır.

A düzey DMF $\Delta R^2 < 0,13$ DMF yoktur ya da ihmal edilebilir düzeydedir.

B düzey DMF $0,13 \leq \Delta R^2 < 0,26$ Orta düzey DMF vardır

C düzey DMF $\Delta R^2 \geq 0,26$ Yüksek düzeyde DMF vardır.

OLR yöntemi hem tek biçimli hem de tek biçimli olmayan DMF'yi belirleyebildiği için bu çalışmada tercih edilmiştir.

Ülkemizde iki kategorili veriler için DMF belirlemeye yönelik çalışmalar son yıllarda sıklıkla yapılmaktadır (Atar ve Kamata, 2011; Yurdugül, 2003; Doğan ve Öğretmen 2006; Karakaya ve Kutlu, 2012; Karakaya, 2012; Kan, 2007; Özdemir, 2003; Öğretmen, 1995; Gök, Kelecioğlu ve Doğan, 2010; Atalay Kabasakal ve Diğerleri, 2010; Demir ve Köse, 2014). Ancak çok kategorili verilerde DMF belirlemeye yönelik çok az çalışma bulunmaktadır (Kabasakal ve Kelecioğlu, 2012, Asil, 2010; Başoğlu ve Öğretmen, 2013; Gök, Kabasakal ve Kelecioğlu, 2015).

Bu araştırmada, poly-SIBTEST, MTK-OO ve OLR yöntemleri kullanılarak elde edilen DMF sonuçları cinsiyete göre karşılaştırılmıştır. Hem gözlenen puana hem de MTK'na dayalı yöntemlerin birbiriyle karşılaştırılmasıyla yöntemlerin farklılıkları ve benzerlikleri belirlenmeye çalışılmıştır. Çok kategorili maddeler için DMF belirlemede yukarıda belirtilen bu üç yöntemin karşılaştırıldığı çalışmalar literatürde çok az olduğundan, çalışma bu yönüyle önemlidir.





Bu amaçla PISA 2012 matematik çalışma etiği anketinde yeralan maddelerin cinsiyete göre DMF gösterip göstermediği poly-SIBTEST, MTK-OO ve OLR yöntemleri ile incelenmiştir. Ayrıca, DMF gösteren maddelerin yöntemlere göre değişip değişmediğine bakılmıştır.

YÖNTEM

Araştırmada, PISA 2012 uygulamasındaki matematik öğrenme anketinin alt boyutu olan matematik çalışma etiği bölümünde yeralan 9 maddenin değişen madde fonksiyonu (DMF) gösterip göstermediği incelenmiştir. Çalışmada var olan durumun belirlenmesi amaçlandığından, betimsel bir araştırmadır.

Araştırmanın evrenini, PISA 2012 uygulamasına katılan 15 yaş grubu öğrenciler oluşturmaktadır. PISA 2012 uygulamasının Türkiye örneğinde anket maddelerine 3217 birey yanıt vermiştir. 3217 bireyden kayıp veriler temizlenmiş ve geriye kalan 3130 bireyden SPSS programı aracılığıyla random olarak 1000 (500 kadın, 500 erkek) birey seçilmiştir. Örnekleme ait betimsel istatistikler Tablo 1’de verilmiştir.

Tablo 1

Örnekleme ait betimsel istatistikler

Birey Sayısı	1000 (Toplam)	500 (Erkek)	500 (Kadın)
Madde Sayısı	9	9	9
Ortalama	18,64	18,88	18,39
Standart Sapma	5,56	5,594	5,52
Basıklık	0,448	0,407	0,490
Çarpıklık	0,446	0,436	0,383
Cronbach α	0,899	0,901	0,895

Tablo 1’ de örneklemlere ilişkin hesaplanan Cronbach α güvenilirliğinin yeterli olduğu görülmektedir. Cinsiyete ait ortalama ve standart sapma değerlerinin de birbirine yakın olduğu söylenebilir. Çarpıklık ve basıklık değerleri incelendiğinde (-1,+1) aralığında çıktığı görülmektedir. Bu da verilerin normal dağıldığını gösterir.

PISA 2012 öğrenci anketinde yeralan matematik çalışma etiği tutum maddelerine verilen yanıtlar, 2014 yılı Ocak ayında OECD (<http://pisa2012.acer.edu.au/>) sitesinden indirilmiştir. Anket maddeleri EK-1’de verilmiştir. 2012 PISA öğrenci anketinde, Matematik çalışma etiğine ilişkin 9 madde bulunmaktadır. Bu maddelere verilebilecek cevaplar 4 kategoriden oluşmaktadır. Yanıtlar, Kesinlikle Katılıyorum (1), Katılıyorum (2), Katılmıyorum (3) ve Kesinlikle Katılmıyorum (4) olarak kodlanmıştır. Analiz sonuçlarının daha kolay anlaşılabilmesi amacıyla kodlamalar; Kesinlikle Katılıyorum (4), Katılıyorum (3), Katılmıyorum (2) ve Kesinlikle Katılmıyorum (1) olarak değiştirilmiştir.

Araştırma verilerinin analizi iki aşamada gerçekleştirilmiştir. Veri analizinin ilk aşamasında, MTK’nın varsayımlarını karşılayıp karşılayamadığı (SPSS programıyla)





incelenmiştir. Verilerin MTK'nın tek boyutluluk ve yerel bağımsızlık varsayımını sağlayıp sağlamadığı belirlenmiştir. Bunun için verilere SPSS.21 programıyla Açıklayıcı Faktör analizi (AFA) yapılmış ve AFA'ne ait çıktılar Tablo 2 'de verilmiştir.

Tablo 2

Ölçeğe Ait Faktörlerin Özdeğerleri ve Varyans Açıklama Oranları

Maddeler	Özdeğerler		
	Toplam	% Varyans	Toplamlı Yüzde %
1	5,144	57,160	57,160
2	,825	9,169	66,328

Tablo 2'de, özdeğeri 1'den büyük olan tek faktör olduğu ve tek başına toplam varyansın %57,16'sını açıkladığı görülmüştür. Yani, anketin tek boyutlu olduğu söylenebilir. MTK'nın diğer bir varsayımı olan yerel bağımsızlık da, tek boyutluluk varsayımı ile ilişkilidir. Yerel bağımsızlık, testte yeralan maddelerin cevaplanma olasılıklarının birbirinden bağımsız olmasıdır. Ayrıca maddelerin testteki sırası bireyin performansını etkilememelidir (Hambleton ve Swaminathan,1985). Çalışmadaki anket tek boyutlu olduğundan yerel bağımsızlık varsayımının da karşılanmış olduğu kabul edilmiştir.

Anketin uyumlu olduğu modeli belirlemek için, veri daha çok anketlerde kullanılan Adlandırılmalı ve Aşamalı Tepki Modellerine göre analiz edilmiştir. Elde edilen -2LogLikelihood (-2LL) değerlerinin farkları dikkate alınarak ki-kare testi yapılmış, verinin Aşamalı Tepki Modeline uyumlu olduğu görülmüştür. Elde edilen -2LL istatistiklerinin değerleri ve her modelde yeralan parametre sayıları Tablo 3'te verilmiştir.

Tablo 3

İki Modele Ait -2loglikelihood Değerleri

	Aşamalı Tepki Modeli	Adlandırılmalı Tepki Modeli
Parametre sayısı	36	54
-2loglikelihood	3192,2	3217,5

$$\chi^2 = -2\loglikelihood_{nominal} - (-2\loglikelihood_{graded})$$
$$= 3217,5 - 3192,2 = 25,3$$

Bulunan 25,3 değeri, 18 serbestlik derecesinde χ^2 kritik değeri (28,869) ile karşılaştırıldığında bu değer anlamsız olduğu görülmektedir. Bu da verinin aşamalı tepki modeline (ATM) uygun olduğu anlamına gelmektedir.

Ankette yeralan 9 maddenin ATM ile yapılan analiz sonucu elde edilen ayırt edicilik değerleri, eşik değerleri ve maddelerin bilgi verdiği maksimum yetenek düzeyleri Tablo 4'de verilmiştir.





Tablo 4

Madde Parametrelerine Ait Değerler

Maddeler	A	B1	B2	B3	Max. Bilgi Verdiği θ Düzeyi	Gözlenen-beklenen Frekans Uyumu
Madde 1	2,10	-1,99	-0,73	0,77	-1	Uyumlu
Madde 2	2,48	-1,70	-0,11	1,12	0	Uyumlu
Madde 3	2,45	-2,15	-1,16	0,35	-2	Uyumlu
Madde 4	2,21	-1,89	-0,40	1,03	-0,4	Uyumlu
Madde 5	2,60	-1,83	-0,48	0,78	-0,4	Uyumlu
Madde 6	2,70	-1,87	-0,57	0,74	-0,6	Uyumlu
Madde 7	2,60	-2,18	-1,27	0,44	-1,4	Uyumlu
Madde 8	2,09	-2,09	-0,72	0,74	-0,8	Uyumlu
Madde 9	1,93	-1,80	-0,35	1,06	-0,4	Uyumlu

Tablo 4’de maddeye ait ayırt edicilik değerleri, eşik değerleri, maximum bilgi verdiği yetenek düzeyi ve gözlenen-beklenen frekans uyumunun olup olmadığı yer almaktadır. Tablo incelendiğinde, madde ayırt edicilik değerlerinin 1,93 ile 2,70 arasında değiştiği görülmektedir. En düşük ayırt ediciliğe sahip olan maddenin 9.Madde olduğu, en yüksek ayırt edicilik değerine sahip maddenin ise 6. madde olduğu görülmektedir. Eşik değerlerine bakıldığında %50 olasılıkla en fazla yetenek düzeyi gerektiren kategori 2 .maddenin üçüncü kategorisidir. Maddelere ilişkin en fazla bilginin ise -1,4 ile 0 aralığında elde edildiği görülmektedir.

Ayrıca kategorilerindeki gözlenen frekanslara dayanarak elde edilen gözlenen oranlar ile beklenen oranları arasında büyük bir farklılık olmadığı, bütün maddelerin uyumlu olduğu görülmektedir.

İkinci aşamada ise ankette yeralan 9 maddenin cinsiyete göre DMF gösterip göstermediği incelenmiştir. Poly-SIBTEST için SIBTEST yazılımı, OLR için OLR yöntemi için Zumbo (1999) tarafından yazılan makro, MTK-OO için IRTLRFIF yazılımı kullanılmıştır. Analizlerde, erkekler odak, kadınlar referans grup olarak ele alınmıştır.

BULGULAR

PISA 2012 Matematik çalışma etiği anketinde yeralan 9 maddenin cinsiyete göre DMF gösterip göstermediğini belirlemek için poly-SIBTEST, MTK-OO ve OLR yöntemleri ile incelenerek uyumları karşılaştırılmıştır.

Cinsiyete göre Poly- SIBTEST analizi sonuçları Tablo 5’de, MTK-OO analizi sonuçları Tablo 6’da ve OLR analizi sonuçları Tablo 7’de verilmiştir.



Tablo5

Cinsiyete Göre Poly-SIBTEST Sonuçları

Madde No	Poly SIBTEST	
	Beta Estimate	p
1	-0,016	0,023
2	0,054	0,021
3	-0,122	0,020
4	0,016	0,023
5	-0,022	0,022
6	0,067	0,021
7	-0,028	0,019
8	0,043	0,023
9	0,025	0,025

Referans grup: Kadın (n=500) Odak grup: Erkek (n=500)

Tablo 5 incelendiğinde; POLYSIBTEST sonuçlarına göre; $p < 0.05$ anlamlılık düzeyinde ölçekte yeralan tüm maddelerin farklı fonksiyonlaştığı görülmektedir. DMF düzeylerine baktığımızda ise; 1,2,4,5,6,7,8 ve 9. maddelerin ihmal edilebilir (A) düzeyde, 3.maddenin ise yüksek düzeyde (C) erkekler lehine farklılaştığı görülmektedir.

Tablo 6

MTK-OO Analizi Sonuçları

Maddeler	Hipotez Testi	G ²
1	All	3,04
2	All	3,7
3	All	14,04
3	a	2,4
3	b	11,5
4	All	2,64
5	All	1,04
6	All	5,34
6	a	0,01
6	b	5,23
7	All	1,14
8	All	2,64
9	All	6,24
9	a	3,31
9	b	2,93

Tablo 6'da MTK-OO'na ait bulgular yer almaktadır. Elde edilen G2sonuçlara göre 3, 6 ve 9. maddelerin farklı fonksiyonlaştığı, bu maddelerden 6. ve 9. maddelerin ihmal edilebilir (A) düzeyde, 3. maddenin ise orta düzeyde (B) DMF gösterdiği görülmektedir.

Tablo 7

Cinsiyete Göre OLR Sonuçları

Madde No	DMF X ²	Model1 R ²	Model2 R ²	Model3 R ²
1	10,05*	60,16	60,28	60,28
2	1,271	67,52	67,52	67,68
3	14,511*	64,22	64,42	64,85
4	0,566	61,75	61,79	61,76
5	0,085	66,72	66,72	66,74
6	0,675	68,06	68,09	68,11
7	1,862	66,16	66,26	66,27
8	2,285	57,69	57,87	57,88
9	5,116	58,74	59,03	59,04

(Sd=2; $\chi^2=9,21$)

Tablo 7'deki OLR sonuçlarına göre; dokuz maddeden sadece iki tanesinin ki-kare değeri 2 serbestlik derecesinde anlamlıdır. DMF gösteren maddelerin DMF miktarlarını hesaplamak için modellere ait R² değerleri farkları göz önüne alındığında (tek biçimli DMF için Model2 R² - Model 1 R² ve tek biçimli olmayan DMF için Model3 R² - Model2 R² hesaplanır); 1. maddenin ihmal edilebilir (A) düzeyde tek biçimli DMF, 3. maddenin ise yüksek düzeyde (C) tek biçimli olmayan DMF gösterdiği görülmektedir.

SONUÇ, TARTIŞMA VE ÖNERİLER

Bu araştırmada, OECD tarafından hazırlanan PISA 2012 uygulamasında yeralan matematik bilimlerine yönelik anketinde yeralan matematik çalışma etiğiyle ilgili 9 tutum maddesi aşamalı tepki modeli" ne göre incelenmiş ve bu maddelerin DMF gösterip göstermediği araştırılmıştır. DMF, bireylerin cinsiyet (kadın-erkek) değişkeni temel alınarak POLY-SIBTEST, MTK-OO ve OLR yöntemleri kullanılarak yapılmıştır. Elde edilen bilgiler şu şekilde özetlenebilir.

Yapılan DMF analizlerine incelendiğinde, 3. maddenin Poly-SIBTEST analizi sonucuna göre yüksek düzeyde, MTK-OO analizi sonuçlarına göre orta düzeyde, OLR sonucuna göre de yüksek düzeyde tek biçimli olmayan DMF içerdiği görülmüştür.

Tekniklerden elde edilen sonuçlar karşılaştırıldığında; poly-SIBTEST, OLR ve MTK-OO yöntemlerinde de 3. maddenin DMF gösterdiği, ancak DMF düzeyinin farklı olduğu görülmüştür. OLR ile Poly-SIBTEST yöntemlerinde ise 2 maddenin ortak DMF verdiği, ancak bu maddelerin ihmal edilebilir düzeyde olduğu görülmüştür. DMF gösteren 3. maddeye baktığımızda MTK-OO ve Poly-SIBTEST yönteminde tek biçimli DMF gösterirken, OLR yönteminde yüksek düzeyde tek



biçimli olmayan DMF gösterdiği görülmüştür. Atar ve Kamata (2011), simülasyon verileriyle farklı koşullarda MTK-OO ve OLR yöntemlerine ait I. Tip hata oranlarını karşılaştırdıkları çalışmada MTK-OO yönteminin OLR yöntemine göre daha güçlü olduğunu bulmuşlardır. MTK-OO ve poly-SIBTEST yöntemleri tek biçimli DMF'yi belirlerken, OLR yöntemi tek biçimli ve tek biçimli olmayan DMF'yi belirleyebilmektedir. Bu nedenle OLR yöntemi DMF'nin doğasına ilişkin daha fazla bilgi vermektedir.

Elde edilen sonuçlar incelendiğinde; kullanılan yöntemlere göre DMF gösteren maddeler için benzer sonuçlar verse de, DMF miktarlarının farklı olduğu gözlenmiştir. Bu farklılığın nedeninin ise yöntemlerin DMF düzeylerine ilişkin sınıflamalarına ait değer aralıklarının farklılığından kaynaklandığı düşünülmektedir.

PISA gibi geniş ölçekli sınavlarda yeralan maddelerin DMF gösterip göstermediklerine ilişkin yapılan çalışmalarda (Costa ve Araújo; 2012, Ayan; 2011, Asil; 2010, Le; 2009, Huang; 2010; Geske ve Ozala; 2010, Yıldırım ve Berberoğlu, 2009, Atalay Kabasakal ve Kelecioğlu; 2012 ve Gök, Atalay Kabasakal ve Kelecioğlu; 2015) uygulamalarda yeralan maddelerin DMF gösterdiği bulunmuştur. Elde edilen bu sonuçlar çalışmamızla da uyumludur. Asil (2010) ve Gök, Atalay Kabasakal ve Kelecioğlu (2015) PISA uygulamasında yeralan maddelerin DMF göstermesinin nedenini çeviri ve uyarlama yanlışlıklarından kaynaklandığını belirtmişlerdir. Van de Vijver ve Tanzer (2003), Hambleton, Merenda ve Spielberger (2005), madde yanlılığının özellikle uluslar arası karşılaştırmalarda kullanılan ölçme araçlarında; yetersiz çeviri, maddenin yazılış tarzı ve maddenin orjinal dildeki anlamının çeviride belirsizleşmesi gibi nedenlerle DMF'nin ortaya çıktığını vurgulamışlardır. PISA, TIMMS ve PIRLS gibi uluslar arası düzeyde yapılan uygulamaların geçerliğinin artırılması için uygulamalarda yeralan maddelerin DMF gösterip göstermediği incelenmelidir.

Bu çalışmada maddelerin DMF gösterip göstermediği incelenmiş, yanlılık çalışması yapılmamıştır. Yapılacak diğer araştırmalarda DMF gösteren maddelerin yanlı olup olmadığı yargısal süreçle belirlenebilir. Ayrıca PISA 2012 de yeralan anketler farklı kültür, okul, bölge, sosyo ekonomik düzey vb. değişkenlere göre DMF analizleri yapılarak karşılaştırılabilir. Ayrıca bu çalışmada çok kategorili veriler için MTK-OO, Poly-SIBTEST ve OLR yöntemleri kullanılmıştır, yapılacak diğer çalışmalarda MIMIC, Genelleştirilmiş Mantel-Hanzsel, GLLAMM gibi farklı DMF belirleme yöntemleri ile benzer çalışmalar yapılabilir. Testteki DMF'li madde sayısı, örneklem büyüklüğü, test uzunluğu, yetenek dağılımları gibi faktörler simülasyon koşulları alınarak farklı yöntemler için I. Tip hata oranlarının karşılaştırıldıkları çalışmalar da yapılabilir.





KAYNAKÇA

- Asil, M. (2010). Uluslar arası Öğrenci Değerlendirme Programı (PISA) 2006 Öğrenci Anketinin Kültürler Arası Eşdeğerliğinin İncelenmesi. Yayınlanmamış Doktora Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Ayan, C. (2011). PISA 2009 Fen Okuryazarlığı Alt Testinin Değişen Madde Fonksiyonu Açısından İncelenmesi, Yayınlanmamış Yüksek Lisans Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Atar, B. ve Kamata, A. (2011). MTK Olabilirlik Oranı testi ve Lojistik regresyon Değişen madde Fonksiyonu Belirleme Yöntemlerinin Karşılaştırılması. Hacettepe Üniversitesi Eğitim Fakültesi Dergisi 41: 36-47
- Bakan Kalaycıoğlu, D. ve Kelecioğlu, H. (2011). Öğrenci Seçme Sınavı'nın madde yanlılığı açısından incelenmesi. Eğitim ve Bilim, 36 (161), 3-12.
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 221-256). Westport: American Council on Education & Praeger Publishers.
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedure to identify differential item functioning test items. Educational Measurement: Issues and Practice, 17, 31-44.
- Costa, P.D. & Araújo L. (2012). Differential Item Functioning (DIF): What Functions Differently for Immigrant Students in PISA 2009 Reading Items? European Commission Joint Research Centre. İtalya.
- Doğan, E., Guerrero, A. & Tatsuoka, K. (2005). Using DIF to investigate strength and weakness in mathematics achievement profiles of 10 different countries. The annual meeting of the National Council on Measurement in Education-NCME (2005,12-14 April). Montreal, Canada.
- Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Eğitim Reformu Girişimi (2010). PISA 2009 Sonuçlarına İlişkin Değerlendirme. (http://erg.sabanciuniv.edu/sites/erg.sabanciuniv.edu/files/PISA2009DeğerlendirmeNotu_Final_08022010.pdf, Erişim Tarihi: 06.05.2014).
- Finch, W. H. ve French, F.B. (2007). Detection of Crossing Differential Item Functioning A Comparison of Four Methods. Applied Psychological Measurement. Educational and Psychological Measurement 67.4.
- Gök, B., Atalay Kabasakal, K. ve Kelecioğlu, H. (2015). PISA 2009 Öğrenci Anketi Tutum Maddelerinin Kültüre Göre Değişen Madde Fonksiyonu Açısından İncelenmesi. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 5/1, Yaz 2014, 72-87.
- Hambleton, R. K. & De Jong, J.H.A.L. (2003). Advances in translating and adapting educational and psychological tests. Language Testing, 20(2).
- Hambleton, R. K & Swaminathan, H. (1985). Item response theory: Principles and applications. Norwell, MA: Kluwer Academic Publishers.
- Hambleton, R. K., Merenda, P. F. ve Spielberger, C. D. (eds.) (2005). Adapting educational and psychological tests for cross-cultural assessment. Mahwah, NJ Lawrence Erlbaum.
- Huang, X. (2010). Differential Item Functioning: The Consequence of Language, Curriculum, or Culture?, Unpublished PhD dissertation, University of California, Berkeley: USA
- Johanson, G. A., & Dodeen, H. (2003). An Analysis of Sex-related Differential Item Functioning in Attitude Assessment. Assessment & Evaluation in Higher Education, 28.
- Le, L.T. (2009). Investigation gender differential item functioning across countries ABD test languages for PISA science items. International Journal of Testing. 9(2), 122-133.
- MEB (2013). PISA 2012 Uluslararası Öğrenci Değerlendirme Programı Ulusal Ön Raporu. MEB, Ankara.





- 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.
- Özdemir, D. (2003). Çoktan seçmeli testlerde iki kategorili ve önsel ağırlıklı puanlamanın diferansiyel madde fonksiyonuna etkisi ile ilgili bir araştırma. *Eğitim ve Bilim*, 25; 37-44.
- Stout, W. F., & Roussos, L. A. (1995). SIBTEST Users Manual, (2nd ed.). Unpublished manuscript, University of Illinois at Urbana-Champaign, Illinois.
- Thissen, D. (2001). IRTLRDIF (Version 2.0b) [Computersoftware]. Chapel Hill: L. L. Thurstone Psychometric Laboratory, University of North Carolina.
- Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and Equivalence in Cross-Cultural Assessment. *European Review of Applied Psychology*, 54, 119-135.
- Zumbo, B. D. (1999). A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-type (Ordinal) Item Scores. Ottawa ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.
- Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, Where it is now, and Where It Is Going. *Language Assessment Quarterly*, 4(2), 223-233.





Comparison of IRT Likelihood Ratio Test, Poly-Sibtest and Logistic Regression DIF Detection Procedures

*Res. Ass. Çiğdem Arıkan Akın
Hacettepe University-Turkey
cakinarikan@hacettepe.edu.tr*

Extended Abstract

Problem and Purpose: PISA, which are implemented in many countries, identify strengths and weaknesses of education systems of such countries and pave the way for developing future-oriented policies for education systems. For this reason, PISA results need to be able to enlighten differences among individuals and include minimum error. In other terms, reliability and validity evidences regarding measuring results must be presented. One important threat against validity is item and test bias (Clauser and Mazor, 1998). The individuals at the same level of ability are normally expected to get the same score from the test or items in a test. However, probability of correct answers in subgroups of the same ability level might differ due to conditions of the test or some attributes of the item. This is called differential item functioning (DIF) (Zumbo, 2007; Finch and French, 2007). DIF refers to systematic differentiation of individuals in a group with the same test score and ability level in terms of probability of answering a certain test item (Doğan, Guerrero and Tatsuoka, 2005). Regarding items of attitude, Hulin, Drasgow and Parsons (1983) define DIF as differing probability of showing positive attitude towards a certain item by individuals in different subgroups with the same ability level (cited by Johanson and Dodeen, 2003). In attitude scales, existence of DIF affects accuracy of measuring results in a negative way. The study aims at identify whether or not items in PISA 2012 maths work ethics questionnaire differ by gender was examined with poly-SIBTEST, MTK-OO and OLR methods. Also it was investigated whether items showing DIF change depending on methods.

Method: The responses given for attitude items in maths work ethics questionnaire of PISA 2012 student questionnaire were downloaded from the OECD web site (<http://pisa2012.acer.edu.au/>) in January 2014. In Turkey sample of PISA 2012 application, 3217 15-year old students responded to the maths work ethics attitude items. After removing missing data of 3217 participants, 1000 people were selected among 3130 individuals by using SPSS on a random basis (500 females and 500 males). 9 items in the questionnaire were checked to see whether they show DIF by gender. SIBTEST software was used for Poly-SIBTEST method, the macro developed by Zumbo (1999) was used for OLR method, and the IRTLRDIF software was used for MTK-OO. During data analysis, males were used as focus group and females as reference group.





Conclusion and Recommendations: According to the DIF analysis results, item 3 shows non-uniform DIF at high level, medium and high level as a result of Poly-SIBTEST, MTK-OO and OLR analysis, respectively. While MTK-OO and poly-SIBTEST methods determine uniform DIF, OLR is able to determine both uniform and non-uniform DIF. Therefore, OLR indicates more details about nature of DIF. Obtained results show that although used methods produce similar results for items with DIF, DIF amounts are found to be different. This difference is thought to be due to the difference of value intervals belonging to classifications regarding DIF levels. Asil (2012), point out that DIF in PISA items often arises from translation and adaptation errors. In relation with item bias, Van de Vijver and Tanzer (2003), and Hambleton, Merenda and Spielberger (2005) argue that DIF arises especially in questionnaires used in international comparisons due to reasons such as insufficient translation, mistyping of items and ambiguous translation of the item.





EK-1:

Thinking about the mathematics you do for school: to what extent do you agree with the following statements?

		<i>Strongly agree</i>	<i>Agree</i>	<i>Disagree</i>	<i>Strongly disagree</i>
1)	I finish my homework in time for mathematics class.				
2)	I work hard on my mathematics homework.				
3)	I am prepared for my mathematics exams.				
4)	I study hard for mathematics quizzes.				
5)	I keep studying until I understand mathematics material.				
6)	I pay attention in mathematics class.				
7)	I listen in mathematics class.				
8)	I avoid distractions when I am studying mathematics.				
9)	I keep my mathematics work well organised.				

