

Standart MIMIC ve Arınık Çapa Maddeli MIMIC Yöntemleriyle Gerçekleştirilen DMF Çalışması*

A DIF Study by using Standard MIMIC and MIMIC With a Pure Anchor

Pelin TAMCI¹, Hülya KELECİOĞLU²

¹ Hacettepe Üniversitesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme
ABD e-posta: pelintamci@gmail.com

² Hacettepe Üniversitesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme
ABD e-posta: hulyaebb@hacettepe.edu.tr

Makale Türü/Article Types: Araştırma Makalesi/ Research Article

Makalenin Geliş Tarihi: 21.04.2024

Yayına Kabul Tarihi: 30.05.2024

ÖZ

Bu çalışmanın amacı değişen madde fonksiyonu (DMF) belirleme yöntemlerinden standart MIMIC (M-ST) ve arınık çapa maddeli MIMIC (M-AÇM) yöntemlerinin benzerlik ve farklılıklarını belirlemektir. Araştırma PISA 2012 verileri ile 900 ve 450 kişilik örneklem üzerinde referans-odak grup oranları 550/350, 450/450, 300/150, 225/225 olacak şekilde yürütülmüştür. M-ST ve M-AÇM yöntemleri ile belirlenen DMF'li maddeler karşılaştırıldığında, tüm koşullarda M-AÇM yöntemiyle daha az sayıda DMF'li madde belirlendiği görülmüştür. 26 nolu maddenin tüm koşullarda her iki DMF belirleme yöntemiyle de DMF gösterdiği belirlenmiştir. En çok DMF'li madde sayısının M-ST yönteminde 900 kişilik örneklem büyüklüğü için referans-odak oranı 550/350 olduğunda ortaya çıktığı görülmüştür.

Anahtar Sözcükler: Standart MIMIC, arınık çapa maddeli MIMIC, DMF

ABSTRACT

The aim of this study is to determine the similarities and differences between standard MIMIC (M-ST) and MIMIC with a pure anchor (M-PA) methods, which are differential item function (DIF) determination methods. The research was conducted on samples of 900 and 450 people, with reference-focus group ratios of 550/350, 450/450, 300/150, 225/225. When the items with DIF

* **Ahntılama:** Tamcı, P. ve Kelecioğlu, H. (2024). Standart MIMIC ve arınık çapa maddeli MIMIC yöntemleriyle gerçekleştirilen DMF çalışması. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 44(2), 1001-1022.

determined by the M-ST and M-PA methods were compared, it was seen that fewer DIF items were determined by the M-PA method in all conditions. It was seen that item 26 showed DIF in all conditions with both DIF determination methods. It was observed that the highest number of items with DIF occurred when the reference-focus ratio was 550/350 for a sample of 900 people in the M-ST method.

Keywords: Standard MIMIC, MIMIC with a pure anchor, DIF

GİRİŞ

PISA, zorunlu eğitim döneminin sonuna yaklaşan 15 yaşındaki öğrencilerin modern toplumlara tam katılım için gerekli olan bilgi ve becerileri ne derecede edindiklerini değerlendiren OECD (Ekonomik İş birliği ve Kalkınma Örgütü) tarafından gerçekleştirilen uluslararası bir eğitim araştırmasıdır. Üç yıllık döngüler halinde gerçekleştirilen PISA uygulamasında matematik okuryazarlığı, fen okuryazarlığı ve okuma gibi temel okul konularına odaklanılmaktadır. Her bir PISA uygulamasında bu üç alandan biri ana alan olarak belirlenmektedir (OECD, 2023). PISA farklı dillere, kültürlere ve okul sistemlerine sahip olan pek çok ülkede uygulanmaktadır. PISA uygulamasıyla amaçlanan değerlendirmelerin yapılabilmesi için birbirinden farklı özelliklere sahip ülkelere yönelik birbirine eşit olan test takımlarının geliştirilmesi için çeşitli çalışmalar yapılmaktadır (OECD, 2010). Geliştirilen test takımlarındaki maddelerin, ölçülmek istenen özellik dışındaki faktörlerden kaynaklı olarak herhangi bir gruba yanlılık sağlamaması kritiktir. Uluslararası düzeyde kritik çıktıları olan ve farklı kültürlerde uygulanan PISA'da, maddelerin yanlılık göstermemesi, elde edilen sonuçların güvenilir ve geçerli olması bağlamında büyük öneme sahiptir. O nedenle çalışmanın devamında madde yanlılığı, değişen madde fonksiyonu (DMF) ve bu çalışmada ele alınacak DMF belirleme yöntemleri ele alınmaktadır.

Madde yanlılığı, testin ölçmeyi amaçladığı özellik dışındaki bazı etkenlerden dolayı aynı yetenek düzeyindeki alt gruplarda maddenin doğru cevaplanma olasılığının birbirinden farklı olması durumunda ortaya çıkar (Zumbo, 1999). Yanlı bir maddenin aynı yetenek düzeyine sahip bireyler tarafından doğru cevaplandırılma olasılığı, testle ölçülen özellik dışındaki başka özelliklere bağlı olabilmektedir.

Madde yanlılığı inceleme çalışmaları, değişen madde fonksiyonu (DMF) gösteren maddelerin belirlenmesiyle başlar. DMF, madde yanlılığını araştırmak için kullanılan bir indekstir. Maddenin DMF göstermesi, farklı gruplardaki bireylerin ölçülen özellik bakımından eşlendikten sonra maddeyi doğru cevaplama olasılıklarının farklılaşmasıdır (Camili ve Shepard, 1994; Clauser ve Mazor, 1998). DMF incelemesi için geliştirilmiş pek çok yöntem bulunmaktadır. Bu yöntemler arasında Mantel-Haenszel (Holland ve Thayer, 1988), lojistik regresyon (Swaminathan ve Rogers, 1990), SIBTEST (Shealy ve Stout, 1993), Lord'un ki-kare testi (Lord, 1980), MIMIC (Finch, 2005; Oort, 1998) yöntemleri bulunmaktadır. MIMIC yöntemi DMF belirlemek için diğer yöntemlerden sonra geliştirildiğinden bu yöntemlere nazaran daha az çalışılmış ve incelenmiştir.

MIMIC modelleri maddelerin tek biçimli DMF gösterip göstermediğini kontrol etmede kullanılmaktadır. Araştırmacılar DMF'nin yönünü ve büyüklüğünü belirlemek için bu yöntemden elde edilen model parametrelerinin MTK'daki madde güçlük katsayılarına nasıl dönüştürülebileceğini göstermişlerdir (Gallo, Anthony ve Muthen, 1994). DMF bağlamında MIMIC modeli aşağıdaki gibi tanımlanmıştır (Finch, 2005, Wang, Shih ve Yang, 2009):

$$y_i^* = \lambda_i \theta + \beta_i z + \varepsilon_i \quad (1)$$

Denklemden y_i^* , i maddesi için gizil cevap değişkenini; θ gizil özelliği; λ_i , i değişkeni için faktör yükünü; z , DMF'yle ilişkili olduğundan düşünülen grup üyeliği değişkenini; β_i yanıtla grup değişkenini ilişkilendiren eğimi, ε_i tesadüfi hatayı temsil etmektedir. β_i , z 'nin y_i^* 'ye etkisini temsil etmektedir ve $\beta_i=0$ olması i maddesinin z grup değişkenleri boyunca homojen olduğunu ve maddenin DMF'li olmadığını göstermektedir. $\beta_i \neq 0$ olduğunda ise θ 'ya ek olarak z 'den de y_i^* 'ye direkt bir etki söz konusudur ve madde DMF göstermektedir diye yorumlanmaktadır.

Denklemin 1'deki z değişkeni çoklu grup değişkenleri içeren bir vektör olduğundan MIMIC yöntemi, birden fazla grup değişkeniyle eş zamanlı DMF analizi yapılmasına olanak sağlar. Ayrıca z kategorik ya da sürekli bir değişken olabilir. Bu sebeplerden dolayı MIMIC yöntemi; Mantel-Haenszel ve SIBTEST vb. tek bir kategorik grup

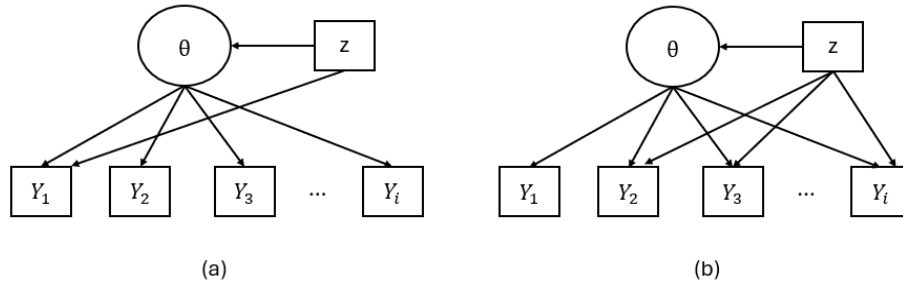
değişkeniyle DMF analizi yapılmasına imkân veren geleneksel yöntemlere göre daha esneklerdir (Wang, Shih ve Yang, 2009).

DMF çalışmalarında odak ve referans grupları aynı metriğe yerleştirebilmek için eşleştirme değişkeni kullanılmaktadır (Wainer, Sireci, ve Thissen, 1991). Eşleştirme değişkeni olarak kullanılacak maddeler DMF açısından incelenen testte yer alan maddeler olduğunda eşleştirme değişkeni arınık (DMF’li madde içermeyen) olmayabilir. İsaletli DMF analizlerinin yapılabilmesi için de eşleştirme değişkeninin arınık olması gerekmektedir (Holland ve Thayer, 1988). Bu amaçla Shih ve Wang (2009), eşleştirme kriteri olarak kullanılmak üzere saf bir kısa çapa oluşturmayı önermiş arınık çapa maddeli MIMIC yöntemini (M-AÇM, MIMIC method with a pure anchor) geliştirmişlerdir.

Standart MIMIC Yöntemi

Standart MIMIC (M-ST) yöntemiyle i maddesinin DMF incelemesi yapılırken, referans ve odak grubun bu maddedeki performansını karşılaştırabilmek için testteki diğer tüm maddeler gruplar arası ortak bir metrik oluşturabilmek amacıyla eşleştirme kriteri olarak kullanılmaktadır. Örneğin; 20 maddelik bir testte madde 1 için DMF incelemesi yapılırken β_1 ’in kestiriminde madde 1’in gruplar arasında farklı parametrelere sahip olmasına izin verilir. Ancak madde 2’den madde 20’ye kadarki diğer tüm maddeler gruplar arasında eşit parametrelere sahip olacak şekilde sınırlandırılır. Şekil 1(a)’da görüldüğü üzere grup üyeliği değişkeni z ’den Y_1 ’e doğru bir ok vardır ve bu ok Y_1 ’in β parametresinin olduğunu ve incelenen madde olduğunu göstermektedir. z ’den diğer maddelere ok olmaması, bu maddelerin β parametresinin olmadığını dolayısıyla da DMF’siz ve de eşleştirme değişkeni olarak görev alacakları şekilde sınırlandırıldıklarını göstermektedir. Bu inceleme aşamasında β_1 istatistiksel olarak 0’dan farklıysa madde 1’in DMF gösterdiği kabul edilmektedir. Benzer şekilde madde 2 için DMF incelemesi yapılırken madde 1 ve madde 3’ten madde 20’ye kadar olan maddeler gruplar arasında eşit parametrelere sahip olacak şekilde sınırlandırılırken β_2 ’nin kestiriminde madde 2’nin gruplar arasında farklı parametrelere sahip olmasına izin verilir. Eğer β_2 istatistiksel olarak 0’dan farklıysa madde 2’nin DMF gösterdiği kabul edilmektedir. Testteki diğer

maddeler için de analiz bu şekilde devam etmektedir (Wang, Shih ve Yang, 2009, Wang ve Shih, 2020). M-ST yönteminde bu şekilde uygulanan gruplar arasında ortak metrik oluşturma yöntemine all-other-item (AOI) denilmektedir (Wang, 2004; Wang ve Yeh, 2003). Bu yöntemde çalışılan madde dışındaki tüm test maddeleri gruplar arasında eşit parametrelere sahip olacak şekilde sınırlandırılarak çapa görevi görür ve yalnızca DMF incelemesi yapılan madde gruplar arasında farklı parametrelere sahip olabilir. Çalışılan her maddenin kendine ait bir ortak metriği olduğundan testteki madde sayısı kadar ortak metrik olacaktır. Testte DMF gösteren bir madde olmadığı sürece eşleştirme değişkenlerinin hepsinin eş zamanlı olarak DMF'den arınık olması mümkün değildir. Testteki DMF'li madde sayısı ne kadar fazlaysa eşleştirme değişkenlerinin arınıklığı da o kadar kötü olmaktadır. Bu nedenle AOI yöntemi de aynı oranda kötü performans göstermektedir.



Şekil 1. (a) M-ST Yöntemi ve (b) M-AÇM Yöntemi

Arınık Çapa Maddeli MIMIC Yöntemi

Testteki tüm maddeler eşleştirme değişkeni olarak kullanılırsa eşleştirme değişkenindeki DMF'li olma durumu ölçek saflaştırma yöntemleriyle telafi edilemeyecek kadar büyük olabilir. Eşleştirme kriterindeki DMF'li maddelerin DMF incelemesindeki etkisini ortadan kaldırmanın uygun bir yolu testteki DMF göstermeyen küçük bir madde grubunu seçip arınık bir eşleştirme değişkeni (çapa) oluşturmak ve testteki diğer maddelerin DMF incelemesini bu arınık eşleştirme kriteriyle yapmaktır. Burada DMF analizi iki aşamada gerçekleşmektedir. Birinci adımda DMF göstermeme olasılığı en yüksek olan maddeler

belirlenmektedir. İkinci adımda bu maddeler çapa görevi görmekte ve diğer maddeler için eş zamanlı olarak tek bir bilgisayar kalibrasyonu ile DMF analizi yapılmaktadır. Bu metoda Arınık Çapa Maddeli MIMIC Yöntemi (M-AÇM) denilmektedir (Shih ve Wang, 2009, Wang ve Shih, 2010). Gruplar arasında ortak bir metrik sağlamak amacıyla DMF göstermeyen çapa maddeleri seçmek için Shih ve Wang (2009) iteratif MIMIC (M-IT) yöntemini önermiştir. Yaptıkları simülasyon çalışmalarıyla bu yöntemin DMF'siz maddeleri belirlemede yüksek performans gösterdiğini, özellikle de 4 taneye kadar DMF'siz madde seti belirlemede mükemmel yakın performans gösterdiğini görmüşlerdir. Ayrıca çapa madde sayısının fazla olmaması gerektiğini (maddelerin hepsinin DMF'siz olması pek mümkün olmayacağı için) belirterek dört madde ya da test uzunluğunun %10 veya %20'si kadar madde içeren kısa çapalar kullanılmasını önermişlerdir.

M-IT yöntemiyle DMF göstermeme olasılığı en yüksek olan çapa maddelerin bulunduğu M-AÇM yönteminin ilk aşamasının adımları aşağıdaki gibidir (Shih ve Wang, 2009):

1. Madde 1 çapa olarak kabul edilir ve diğer tüm maddeler için MIMIC yöntemiyle DMF incelemesi yapılır. Her bir madde için DMF katsayısı elde edilir.
2. Bir sonraki madde çapa olarak kabul edilir ve diğer tüm maddeler için MIMIC yöntemiyle DMF incelemesi yapılır. Her bir madde için DMF katsayısı elde edilir.
3. Testteki son madde de çapa kabul edilene kadar ikinci adım tekrar edilir.
4. Bütün tekrarların ardından her bir madde için DMF katsayılarının mutlak değerlerinin ortalaması hesaplanır. Ve en düşük ortalama değere sahip maddeler (çalışmada kullanılacak çapa madde sayısı kadar) çapa maddeler olarak belirlenir.

DMF göstermeme olasılığı yüksek olan çapa maddeler belirlendikten sonra ikinci aşamada M-AÇM yöntemi Şekil 1b'de gösterildiği gibi gerçekleştirilmektedir. Şekil 1b'de madde 1'in çapa madde olduğu ve testteki diğer tüm maddelerin DMF açısından incelendiği görülmektedir. Grup üyeliği değişkeni z 'den $Y_2 \dots Y_i$ 'ye doğru oklar varken Y_1 'e doğru ok yoktur. Bu da $Y_2 \dots Y_i$ maddelerinin kendi β parametrelerinin olduğunu ve

de DMF açısından incelenen maddeler olduğunu, Y_1 'in ise β parametresinin olmadığını ve de çapa madde olduğunu göstermektedir. M-AÇM modelinde birden fazla çapa madde olabileceğinden z grup üyeliği değişkenin çapa olan maddelere uzanan oklar kaldırılıp bu maddeler için β kestirimi yapılmayarak model birden fazla çapa madde olan durumlara uyarlanabilmektedir (Wang ve Shih,2010).

Amaç

Bu çalışmada PISA 2012 uygulamasında 4 numaralı kitapçıktaki 34 adet çoktan seçmeli matematik sorusunun anadili İngilizce olup olmama durumuna göre DMF gösterip göstermediği M-ST ve M-AÇM yöntemleriyle analiz edilmiştir. İki yöntemle elde edilen sonuçların örneklem büyüklüğü ve odak/referans grubu oranlarına göre nasıl farklılaştığı incelenmiştir. Shih ve Wang (2009) çalışmasında M-AÇM yöntemini tanıtmış, simülasyon verisiyle farklı koşullar altında M-AÇM yöntemi için elde edilen güç ve I. tip hata oranlarını karşılaştırmıştır. Wang ve Shih (2010), çok kategorili simülasyon verisi kullanarak M-ST, M-AÇM ve ölçek arındırmalı MIMIC yöntemlerinin performanslarını karşılaştırmıştır. Literatürde, M-ST ve M-AÇM yöntemlerinin iki kategorili veri için performanslarının karşılaştırıldığı herhangi bir çalışmaya rastlanmamıştır. M-ST ve M-AÇM yöntemleri DMF belirlemek için diğer yöntemlere göre daha sonra geliştirildiğinden literatürde bu iki yöntemin incelendiği çalışma sayısı yeterli değildir. Bu nedenle farklı koşullar altında bu iki yöntemin sonuçlarının karşılaştırıldığı bu çalışmanın araştırmacılara hangi yöntemin daha etkili olacağı konusunda yol göstereceği düşünülmektedir.

YÖNTEM

Araştırmanın Türü

Bu çalışmada PISA 2012 uygulamasında 4 numaralı kitapçıktaki 34 adet çoktan seçmeli matematik sorusunun anadili İngilizce olup olmama durumuna göre DMF gösterip göstermediği analiz edilmiştir. DMF belirleme yöntemi olarak M-ST ve M-AÇM kullanılmıştır. Çalışma, var olan durumu belirlediğinden dolayı betimsel bir çalışmadır.

Çalışma Grubu

PISA 2012 uygulamasına katılan ve aynı dil ailesinde yer alan farklı anadillere sahip ülkeler, araştırmada kullanılacak uygun çalışma gruplarını tespit etmek amacıyla, DMF analizlerinde referans ve odak grubu belirlemede kriter olarak kullanılacak olan anadili İngilizce olup olmama durumuna göre incelenmiştir. Tablo 1'deki verilere göre uygulamaya katılan öğrenci sayısı ve matematik okuryazarlığı başarı sıralaması bakımından birbirine yakın olan ülkeler belirlenmiştir. Buna göre anadili İngilizce olan ülkeleri (AİÖÜ) temsil etmek üzere ABD ve Birleşik Krallık, anadili İngilizceden farklı olan ülkeleri (AİFOÜ) temsil etmek üzere Çekya, Danimarka, Fransa, İzlanda ve Slovakya seçilmiştir.

Tablo 1. Çalışma Grubuna Dâhil Edilen Ülkelerle İlgili Bilgiler

Ülke	PISA 2012'ye Katılan Öğrenci Sayısı	PISA 2012 Matematik Okuryazarlığı Başarı Sırası	Anadil
ABD	4953	36	İngilizce
Birleşik Krallık	12239	26	İngilizce
Çekya	5327	24	İngilizce Değil
Danimarka	7481	22	İngilizce Değil
Fransa	4613	25	İngilizce Değil
İzlanda	3508	27	İngilizce Değil
Slovakya	4650	35	İngilizce Değil

4 numaralı kitapçığı çözmüş ve matematik testine verdikleri cevaplarda kayıp veri bulunmayan öğrencilerden oluşan 566'sı AİÖÜ'den ve 513'ü AİFOÜ'den oluşan iki grup elde edilmiştir. Bu iki grup üzerinden Tablo 2'de belirtilen 4 farklı çalışma grubu oluşturulmuştur. Çalışma grupları, örneklem büyüklüğü açısından aralarında yaklaşık ½ oranı olması ve referans-odak gruplarının eşit ve eşit büyüklükte olmaması koşullarına göre belirlenmiştir. Çalışma gruplarının belirlenmesi aşamasında SPSS 15 paket programı kullanılmıştır.

Tablo 2. Çalışma Gruplarının Frekans Dağılımı

Çalışma Grubu	AİOÜ Sayısı	AİFOÜ Sayısı	Toplam
Çalışma Grubu 1	550	350	900
Çalışma Grubu 2	450	450	900
Çalışma Grubu 3	225	225	450
Çalışma Grubu 4	300	150	450

PISA 2012’de ‘Matematik Okuryazarlığı’ alanı öne çıktığı için tüm kitapçıklar matematik dersi açısından incelenmiştir. Bu incelemeye göre, ‘‘Matematik Okuryazarlığı’’ ölçen başarı testlerindeki, uygulanmamış (N/A ve 7 kodlu) ve kısmi puanlamaya tabi tutulan maddeler çıkarıldığında en fazla madde sayısına sahip kitapçığın 4 numaraları kitapçık olduğu tespit edilmiştir. 4 numaralı kitapçıkta ‘‘Matematik Okuryazarlığı’’ testinde yer alan, uygulanmamış ve kısmi puanlamaya tabi tutulmuş maddeler çıkarıldığında 34 madde kaldığı görülmüş ve çalışma bu 34 madde için elde edilen veriler üzerinden yürütülmüştür.

DMF analizleri için Mplus 6.12 (Muthén ve Muthén, 1998, 2010) programı ve R programındaki MplusAutomation paketi (Hallquist ve Wiley, 2018) kullanılmıştır.

DMF analizlerine başlamadan evvel verilerin boyut yapısının belirlenmesi sürecin kritik bir basamağıdır (Dorans ve Holland, 1993). Çalışmada kullanılan verilerin tek boyutluluk varsayımını sağlayıp sağlamadığını kontrol etmek amacıyla faktör analizi uygulanmıştır.

Veri yapısının faktör analizine uygun olup olmadığını belirlemek için Bartlett küresellik testi değeri ve Kaiser-Meyer-Olkin (KMO) katsayısı incelenmiştir.

Tablo 3. KMO ve Bartlett Testi Sonuçları

Test		Katsayı
KMO		0.94
Bartlett Küresellik Testi	Ki-Kare	6077.69
	Sd	561
	p	.00

Field (2009)'a göre, KMO katsayısının 0.5'ten büyük olması ve Bartlett küresellik testinden elde edilen ki-kare istatistiğinin anlamlı çıkması verinin faktör analizi uygun olduğu anlamına gelmektedir. Tablo 3 incelendiğinde KMO katsayısının 0.94 olduğu ve Bartlett testi sonucu elde edilen ki-kare istatistiğinin anlamlı ($\chi^2= 6077.69$, $p < .01$) bulunduğu görülmektedir. Buna göre verinin faktör analizine uygun olduğu yorumu yapılabilir.

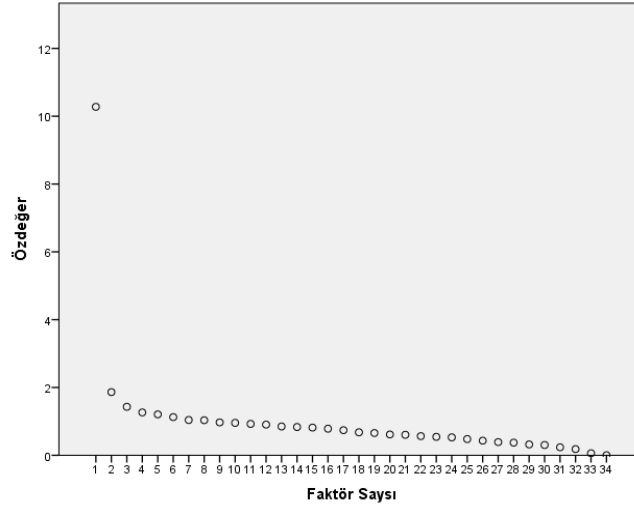
Tablo 4. Paralel Analiz Sonuçlarına Göre Açıklanan Varyans Yüzdeleri

No	Açıklanan Varyans Yüzdeleri		
	Gerçek Veri	Tesadüfi Veri	Tesadüfi Verinin 95. Yüzdelik Dilimi
1	30.60	6.0	6.60
2	5.60	5.60	6.10
3	4.30	5.30	5.70
4	3.80	5.10	5.40
5	3.60	4.80	5.10
6	3.30	4.60	4.90
7	3.10	4.50	4.70
8	3.10	4.30	4.50
9	2.90	4.20	4.40
10	2.90	4.00	4.20
11	2.80	3.90	4.00
12	2.70	3.70	3.90
13	2.50	3.60	3.70
14	2.50	3.40	3.60
15	2.40	3.30	3.40
16	2.30	3.20	3.30
17	2.20	3.00	3.10
18	2.00	2.70	3.00
19	2.00	2.70	2.90
20	1.80	2.60	2.80

21	1.70	2.40	2.60
22	1.70	2.30	2.50
23	1.60	2.10	2.30
24	1.50	2.00	2.20
25	1.30	1.90	2.10
26	1.20	1.70	1.90
27	1.10	1.50	1.80
28	1.00	1.40	1.60
29	0.90	1.20	1.50
30	0.80	1.00	1.30
31	0.60	0.80	1.10
32	0.20	0.60	0.90
33	0.00	0.40	0.60
34	0.00	0.00	0.00

Verinin boyut sayısını belirlemek amacıyla gerekleřtirilen paralel analiz sonuları Tablo 4'te sunulmuřtur. Buna gre, ilk satır iin gerek verinin aıkladıđı varyans yzdesi deđerinin tesadfi verinin aıkladıđı varyans yzdesinden byk olduđu, fakat ikinci satırdaki varyans yzdesi deđerlerinden itibaren tesadfi verinin aıkladıđı varyans yzdesinin gerek verinin aıkladıđı varyans yzdesinden eřit ya da kk olduđu grlmektedir. Buna gre, verinin tek boyutlu olduđu sylenebilir.

Ařađıda sunulan Őekil 2'de, veriye ait yama-eđim grafiđi verilmiřtir. Grafik incelendiđinde birinci zdeđerden sonra keskin bir dřř olduđu grlmektedir. Bu durum da verinin tek boyutlu olduđunu destekler niteliktedir.



Şekil 2. Yamaç - Eğim Grafiği

BULGULAR

PISA 2012 matematik testi 4 numaralı kitapçıktaki 34 maddenin anadil değişkenine göre M-ST ve M-AÇM yöntemleriyle yapılan DMF analizi sonuçları Tablo 5, Tablo 6, Tablo 7 ve Tablo 8’de verilmiştir.

Tablo 5. Çalışma Grubu 1 için DMF Analiz Sonuçları

Madde No	M-ST		M-AÇM	
	Beta Estimate	p	Beta Estimate	p
1	-0.05	0.58	-0.01	0.91
2	-0.16	0.09	-0.12	0.29
3	0.20	0.01*	0.23	0.03*
4	0.16	0.18	0.19	0.15
5	0.09	0.23	0.13	0.22
6	-0.16	0.07	-0.12	0.29
7	0.01	0.88	0.04	0.65
8	-0.42	0.00*	-0.37	0.01*
9	0.34	0.01*	0.36	0.01*
10	-0.04	0.69	-	-
11	0.18	0.06	0.21	0.05
12	0.04	0.78	0.07	0.59

13	0.11	0.16	0.14	0.12
14	-0.30	0.01*	-0.25	0.03*
15	-0.17	0.04*	-0.13	0.15
16	-0.20	0.02*	-0.16	0.13
17	-0.02	0.82	0.03	0.82
18	-0.02	0.81	-	-
19	-0.08	0.44	-0.03	0.82
20	0.07	0.40	0.10	0.31
21	-0.03	0.72	0.01	0.90
22	0.43	0.00*	0.46	0.00*
23	-0.07	0.38	-0.03	0.76
24	0.26	0.01*	0.29	0.00*
25	0.07	0.57	-	-
26	0.31	0.00*	0.33	0.00*
27	-0.06	0.47	-	-
28	-0.17	0.04*	-0.12	0.24
29	0.07	0.51	0.10	0.39
30	-0.08	0.32	-0.04	0.72
31	-0.25	0.01*	-0.20	0.12
32	-0.00	0.96	0.04	0.71
33	0.13	0.28	0.17	0.23
34	0.13	0.11	0.16	0.10

$p < .05$,

- ile gösterilen maddeler çapa olarak kullanıldığından beta kestirimi yapılmamıştır.

Çalışma grubu 1 için M-ST yöntemiyle yapılan DMF analizi sonucunda 3, 8, 9, 14, 15, 16, 22, 24, 26, 28 ve 31 numaralı maddeler DMF göstermektedir. M-AÇM yöntemiyle yapılan DMF analizi sonucunda ise 3, 8, 9, 14, 22, 24 ve 26 numaralı maddeler DMF göstermektedir.

Tablo 6. Çalışma Grubu 2 için DMF Analiz Sonuçları

Madde No	M-ST		M-AÇM	
	Beta Estimate	p	Beta Estimate	p
1	-0.07	0.39	-0.09	0.37
2	-0.13	0.17	-0.14	0.20
3	0.20	0.01*	0.17	1.00
4	0.13	0.28	0.11	0.38
5	0.04	0.63	0.01	0.94
6	-0.31	0.01*	-0.32	0.01*
7	0.05	0.50	0.03	0.75
8	-0.35	0.00*	-0.37	0.00*
9	0.22	0.09	0.21	0.11

10	-0.01	0.91	-	-
11	0.15	0.13	0.12	0.25
12	-0.06	0.65	-0.08	0.56
13	0.07	0.36	0.06	0.54
14	-0.34	0.00*	-0.35	0.00*
15	-0.21	0.01*	-0.22	0.02*
16	-0.09	0.30	-0.11	0.31
17	-0.05	0.55	-0.07	0.50
18	0.03	0.69	-	-
19	-0.05	0.65	-0.08	0.57
20	0.16	0.05	0.14	0.17
21	-0.01	0.90	-0.03	0.73
22	0.46	0.00*	0.42	0.00*
23	-0.04	0.63	-0.06	0.57
24	0.24	0.00*	0.21	0.03*
25	0.11	0.38	-	-
26	0.36	0.00*	0.34	0.00*
27	-0.02	0.86	-	-
28	-0.15	0.05	-0.17	0.08
29	0.05	0.64	0.03	0.82
30	-0.08	0.30	-0.11	0.33
31	-0.21	0.03*	-0.23	0.07
32	-0.01	0.86	-0.04	0.70
33	0.06	0.61	0.04	0.78
34	0.18	0.02*	0.16	0.09

$p < .05$

- ile gösterilen maddeler çapa olarak kullanıldığından beta kestirimi yapılmamıştır.

Çalışma grubu 2 için M-ST yöntemiyle yapılan DMF analizi sonucunda 3, 6, 8, 14, 15, 22, 24, 26, 31 ve 34 numaralı maddeler DMF göstermektedir. M-AÇM yöntemiyle yapılan DMF analizi sonucunda ise 6, 8, 14, 15, 22, 24 ve 26 numaralı maddeler DMF göstermektedir.

Tablo 7. Çalışma Grubu 3 için DMF Analiz Sonuçları

Madde No	M-ST		M-AÇM	
	Beta Estimate	p	Beta Estimate	p
1	-0.12	0.29	-0.07	0.62
2	-0.07	0.57	-0.03	0.87
3	0.09	0.38	0.14	0.31
4	-0.07	0.66	-	-

5	0.12	0.27	0.17	0.23
6	-0.17	0.17	-0.11	0.45
7	0.04	0.72	-	-
8	-0.34	0.01*	-0.28	0.06
9	0.35	0.04*	0.39	0.03*
10	0.06	0.69	0.11	0.53
11	0.16	0.24	0.20	0.17
12	0.08	0.63	0.13	0.49
13	-0.10	0.37	-	-
14	-0.31	0.01*	-0.26	0.09
15	-0.25	0.03*	-0.21	0.11
16	-0.02	0.87	0.03	0.80
17	-0.01	0.92	0.05	0.74
18	0.11	0.35	0.15	0.28
19	-0.12	0.44	-0.04	0.84
20	0.27	0.02*	0.31	0.02*
21	0.03	0.78	0.09	0.55
22	0.57	0.00*	0.61	0.00*
23	-0.01	0.95	0.04	0.75
24	0.12	0.25	0.17	0.20
25	0.22	0.21	0.26	0.17
26	0.32	0.00*	0.35	0.01*
27	-0.06	0.65	-	-
28	-0.18	0.10	-0.11	0.45
29	-0.05	0.72	-0.01	0.98
30	-0.19	0.09	-0.13	0.40
31	-0.44	0.00*	-0.36	0.04*
32	-0.10	0.34	-0.03	0.83
33	0.07	0.68	0.12	0.53
34	0.23	0.03*	0.27	0.04*

$p < .05$

- ile gösterilen maddeler çapa olarak kullanıldığından beta kestirimi yapılmamıştır.

Çalışma grubu 3 için M-ST yöntemiyle yapılan DMF analizi sonucunda 8, 9, 14, 15, 20, 22, 26, 31 ve 34 numaralı maddeler DMF göstermektedir. M-AÇM yöntemiyle yapılan DMF analizi sonucunda ise 9, 20, 22, 26, 31 ve 34 numaralı maddeler DMF göstermektedir.

Tablo 8. Çalışma Grubu 4 için DMF Analiz Sonuçları

Madde No	M-ST		M-AÇM	
	Beta Estimate	p	Beta Estimate	p
1	-0.11	0.39	0.00	1.00
2	-0.07	0.62	0.06	0.74
3	0.31	0.00*	0.43	0.01*
4	0.18	0.29	0.27	0.18
5	0.07	0.52	0.21	0.22
6	-0.17	0.23	-0.06	0.72
7	0.12	0.29	0.23	0.15
8	-0.37	0.02*	-0.25	0.18
9	-0.11	0.63	-	-
10	0.21	0.17	0.32	0.09
11	0.25	0.08	0.36	0.04*
12	-0.28	0.23	-0.15	0.56
13	0.21	0.09	0.29	0.05
14	-0.31	0.02*	-0.17	0.29
15	-0.10	0.38	-	-
16	-0.18	0.15	-0.06	0.72
17	0.05	0.70	0.19	0.35
18	0.05	0.67	0.15	0.34
19	-0.20	0.17	-0.04	0.85
20	-0.11	0.40	0.00	1.00
21	-0.04	0.72	0.07	0.69
22	0.21	0.14	0.32	0.09
23	0.00	0.99	0.11	0.46
24	0.27	0.02*	0.39	0.02*
25	-0.09	0.64	-	-
26	0.45	0.00*	0.52	0.00*
27	-0.19	0.15	-0.11	0.44
28	-0.28	0.02*	-0.14	0.43
29	-0.05	0.78	-	-
30	-0.06	0.65	0.07	0.69
31	-0.28	0.05	-0.14	0.47
32	-0.05	0.68	0.10	0.58
33	0.26	0.17	0.36	0.10
34	0.12	0.34	0.21	0.20

$p < .05$,

- ile gösterilen maddeler çapa olarak kullanıldığından beta kestirimi yapılmamıştır.

Çalıřma grubu 4 için M-ST yöntemiyle yapılan DMF analizi sonucunda 3, 8, 14, 24, 26 ve 28 numaralı maddeler DMF göstermektedir. M-AÇM yöntemiyle yapılan DMF analizi sonucunda ise 3, 11, 24 ve 26 numaralı maddeler DMF göstermektedir.

TARTIřMA VE SONUÇ

Bu çalıřmada PISA 2012 uygulamasında 4 numaralı kitapçıkta 34 adet çoktan seçmeli matematik sorusunun anadili İngilizce olup olmama durumuna göre DMF gösterip göstermediđi M-ST ve M-AÇM yöntemleriyle analiz edilmiřtir.

M-ST yöntemine göre 900 kiřilik çalıřma grubunda; R/O oranı 550/350 olduđunda 34 maddenin %32'si (11 madde); R/O oranı 450/450 olduđunda ise %29'u (10 madde) DMF'li bulunmuřtur. DMF'li bulunan maddelerin 7'si ortaktır.

450 kiřilik örnekleme R/O oranı 225/225 olduđunda 34 maddenin %26'sı (9 madde), R/O oranı 300/150 olduđunda ise %18'i (6 madde) DMF'li bulunmuřtur. Bu grupta DMF'li bulunan maddelerin 3 tanesi ortaktır. R/O oranları aısından bakıldıđında, örnekleme sayısı daha küçük ve R/O oranı 1 olduđunda belirlenen DMF'li madde sayısının R/O oranının 1'den büyük olduđu kořuldan daha fazla olduđu görölmektedir. Örnekleme büyüklüđü daha yüksek olan grupta ise R/O oranları eřit olan grupta daha fazla DMF'li madde belirlenmiřtir.

M-AÇM yöntemine göre 900 kiřilik örnekleme; R/O oranı 550/350 ve 450/450 olan her iki grupta 34 maddenin %21'i (7 madde) DMF'li bulunmuřtur. Bu maddelerin 5 tanesi ortaktır. Örnekleme büyüklüđü 450 kiři olduđunda R/O oranı 225/225 olan grupta 34 maddenin %18'i (6 madde), R/O oranı 300/150 olan grupta ise %12'si (4 madde) DMF içermektedir. Bu maddelerin 1 tanesi ortaktır. R/O oranları aısından bakıldıđında 900 kiřilik örnekleme için belirlenen DMF'li madde sayısının eřit olduđu, 450 kiřilik örnekleme ise R/O oranı 1 olduđunda tespit edilen DMF'li madde sayısının daha fazla olduđu görölmektedir.

M-ST ve M-AÇM yöntemleri ile belirlenen DMF'li maddeler karşılaştırıldığında, dört çalışma grubunda da M-AÇM yöntemiyle daha az sayıda DMF'li madde belirlendiği görülmektedir. M-ST yöntemi için dört farklı koşulda ortaya çıkan DMF'li maddelere bakıldığında 8, 14 ve 26 nolu maddelerin, M-AÇM yönteminde ise 26 nolu maddenin ortak şekilde DMF gösterdiği görülmüştür. Buna göre 26 nolu maddenin tüm koşullarda her iki DMF belirleme yöntemiyle de DMF gösterdiği sonucu ortaya çıkmıştır. Ayrıca en çok DMF'li madde sayısı M-ST yönteminde 900 kişilik örneklem büyüklüğü için R/O oranı 550/350 olduğunda belirlenmiştir.

Sonuçlara bakıldığında örneklem büyüklüğü arttıkça her iki yöntemle de belirlenen DMF'li madde sayısının artış gösterdiği görülmektedir. Shih ve Wang (2009) yaptığı simülasyon çalışmasında örneklem büyüklüğü arttıkça M-ST ve M-AÇM yöntemlerinin güç oranlarının arttığını raporlamıştır. Çalışma sonuçları bu açıdan ele alındığında örneklem büyüklüğü arttıkça DMF'li olan maddelerin daha isabetli tespit edildiği yorumu yapılabilir.

Tamcı (2018)'in aynı çalışma gruplarını kullanarak SIBTEST yöntemiyle yaptığı DMF analizi sonucunda elde ettiği DMF'li maddelerin ve bu maddelerin sayılarının M-ST yöntemiyle elde edilen sonuçlara, M-AÇM'yle elde edilen sonuçlara nazaran daha benzer olduğu görülmüştür. Bu durumun M-AÇM yönteminde eşleştirme kriteri olarak arınık maddelerin (4 madde) kullanılmasından kaynaklandığı söylenebilir.

Bu çalışmada grup değişkeni olarak anadili İngilizce olup olmama durumu incelenmiştir. Bunun dışında farklı dil ailesine ait anadil, cinsiyet, okul türü, sosyoekonomik durum vb. değişkenlerle de çeşitli çalışmalar yürütülebilir. İki yöntemin sonuçları yalnızca örneklem büyüklüğü ve R/O oranı açısından incelenmiştir. Testteki madde sayısı, M-AÇM yönteminde kullanılan arınık çapa madde sayıları açısından da farklı koşullar altında incelemeler yapılabilir. Ayrıca yöntemlerin gücü ve I. tip hatası, gerçek verilerdeki koşullar için simülasyon çalışmaları yapılarak incelenebilir. Böylece hangi yöntemin hangi koşullarda daha etkili olduğuna karar verilebilir.

KAYNAKLAR

- Camilli, G., & Shepard, L. (1994). *Methods for identifying biased test items*. London: Sage.
- Clauser, B., & Mazor, K. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17, 31-44. doi: 10.1111/j.1745-3992.1998.tb00619.x
- Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland, & H. Wainer, *Differential Item Functioning* (s. 35-66). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Finch, H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel-Haenszel, SIBTEST, and the IRT likelihood ratio. *Applied Psychological Measurement*, 29, 278-295. doi: 10.1177/0146621605275728
- Field, A. (2009). *Discovering Statistics Using SPSS*. London: SAGE.
- Gallo, J. J., Anthony, J. C., & Muthen, B. O. (1994). Age differences in the symptoms of depression: A latent trait analysis. *Journal of Gerontology: Psychological Sciences*, 49, P251-P264. doi: 10.1093/geronj/49.6.p251
- Hallquist, M. N. & Wiley, J. F. (2018). MplusAutomation: An R package for facilitating large-scale latent variable analyses in mplus. *Structural Equation Modelling*, 25, 621-638. doi: 10.1080/10705511.2017.1402334.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer, & H. I. Braun (Eds.). *Test validity*, 129-145. Hillsdale, NJ: Erlbaum
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale NJ: Erlbaum.
- Muthén, L. K. & Muthén, B. O. (1998, 2010). *Mplus user's guide (6th ed.)*. Los Angeles, CA: Muthén & Muthén
- OECD (2023), "What is PISA?", in *PISA 2022 Assessment and Analytical Framework*, OECD Publishing, Paris. doi: <https://doi.org/10.1787/82f078d7-en>
- OECD (2010). *Translation and adaptation guidelines for PISA 2012*. Budapeşte: OECD Publishing.
- Oort, F. J. (1998). Simulation study of item bias detection with restricted factor analysis. *Structural Equation Modeling*, 5, 107-124.

- Shealy, R. & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as bias/DIF. *Psychometrika*, 58(2), 159-194. doi:10.1007/BF02294572
- Shih, C.-L. & Wang, W.-C. (2009). Differential item functioning detection using the multiple indicators, multiple causes MIMIC method with a pure short anchor. *Applied Psychological Measurement*, 33, 184-199. doi: 10.1177/0146621608321758
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361-370.
- Tamcı, P. (2018). *Kayıp Veriyle Başa Çıkma Yöntemlerinin Değişen Madde Fonksiyonu Üzerindeki Etkisinin İncelenmesi* (Yüksek lisans tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Wainer, H., Sireci, S., & Thissen, D. (1991). Differential testlet functioning: Definitions and detection. *Journal of Educational Measurement*, 28, 197-219.
- Wang, W.-C., & Yeh, Y.-L. (2003). Effects of anchor item methods on differential item functioning detection with the likelihood ratio test. *Applied Psychological Measurement*, 27, 479-498. doi: 10.1177/0146621603259902
- Wang, W.-C. (2004). Effects of anchor item methods on the detection of differential item functioning within the family of Rasch models. *Journal of Experimental Education*, 72, 221-261. <https://doi.org/10.3200/JEXE.72.3.221-261>
- Wang, W.-C., Shih, C.-L., & Yang, C.-C. (2009). The MIMIC method with scale purification for detecting differential item functioning. *Educational and Psychological Measurement*, 69(5), 713-731. <https://doi.org/10.1177/0013164409332228>
- Wang, W. C., & Shih, C. L. (2010). MIMIC methods for assessing differential item functioning in polytomous items. *Applied Psychological Measurement*, 34(3), 166-180. doi: 10.1177/0146621609355279
- 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

SUMMARY

Aim

In this study, whether the 34 multiple choice mathematics items in booklet 4 in the PISA 2012 application showed DIF depending on whether the native speaker is English or not was analyzed by standard MIMIC (M-ST) and MIMIC with a pure anchor (M-PA) method. It was examined how the results obtained according to the two methods differ according to sample size and reference/focus group ratios.

Method

PISA 2012 data was used in this study. In order to determine the appropriate study groups, the countries participating in the PISA 2012 were examined according to whether their native language English, which would be used as a criterion in determining the focus and reference group in the DIF analyses. As a result of the analysis, countries that were close to each other in terms of mathematics achievement rankings and the number of students participating in the application were identified. Accordingly, the USA and the United Kingdom were selected to represent countries whose native language is English, and Slovakia, Iceland, France, Czechia and Denmark were selected to represent countries whose native language is other than English. A group of students were obtained, consisting of students who took booklet number 4 and who had no missing data was observed in the mathematics test. From these groups, study groups were formed with sample sizes of approximately $\frac{1}{2}$ between them. When the items that were not applied (coded N/A and 7) and those that were subject to partial scoring were removed from the achievement test measuring "Mathematical Literacy", it was seen that booklet 4 had the highest number of items. Therefore, in this study, when the items that were not applied and were subject to partial scoring were removed from the "Mathematical Literacy" test in booklet number 4, the data obtained for the remaining 34 items were used.

Results

According to M-ST method, in the study group of 900 people, when the reference-focus group ratio is 550/350, 32% of 34 items (11 items), when the reference-focus group ratio is 450/450, 29% of 34 items (10 items) are found to have DIF. In the sample of 450 people, 26% of the items (9 items) are found to have DIF when reference-focus group ratio is 225/225, and 18% (6 items) are found to have DIF when reference-focus ratio is 300/150.

According to M-PA method, in the study group of 900 people, 21% of 34 items (7 items) are found to have DIF in both groups with reference-focus group ratios of 550/350 and 450/450. In the sample of 450 people, 18% of the items (6 items) are found to have DIF when reference-focus group ratio is 225/225, and 12% (4 items) are found to have DIF when reference-focus ratio is 300/150.


Discussion and Conclusion

According to M-ST method, when the sample size is smaller and the reference-focus group ratio is 1, the number of items with DIF is higher than when the reference-focus group ratio is greater than 1. When the sample size is higher, more DIF items are identified in the group with equal reference-focus ratios.

In the M-PA method, it is seen that the number of items with DIF for the sample of 900 people is equal. In the sample of 450 people and when the reference-focus group ratio is 1, the number of items with DIF is higher.

When the items with DIF determined by the two methods are compared, it is seen that fewer items with DIF are determined by the M-PA method in all four study groups. It is concluded that item 26 showed DIF in all conditions with both DIF determination methods.

ORCID

Pelin Tamcı  ORCID 0009-0005-3677-5443

Hülya Kelecioğlu  ORCID 0000-0002-0741-9934

Araştırmacıların Katkı Oranı Beyanı

Bu çalışmanın planlanması, yürütülmesi ve yazılı hale getirilmesinde araştırmacılar eşit oranda katkı sağlamıştır.

Destek ve Teşekkür Beyanı

Bu araştırmada herhangi bir kurum, kuruluş ya da kişiden destek alınmamıştır.

Çatışma Beyanı

Araştırmacıların, araştırma ile ilgili diğer kişi ve kurumlarla herhangi bir kişisel ve finansal çıkar çatışması yoktur.

Etik Kurul Beyanı

Bu araştırma, PISA 2012 verilerine dayalı bir çalışma olduğu için etik kurul izni gerektirmemektedir.