

PISA 2009 OKUMA TESTİNDEN ELDE EDİLEN İKİ KATEGORİLİ VERİLERİN BILOG PROGRAMI İLE İNCELENMESİ

Arş. Gör. Derya Çakıcı Eser
Kırıkkale Üniversitesi, Kırıkkale
deryacakicieser@gmail.com

Özet

Bu çalışmada, PISA 2009 verisinin madde tepki kuramı ile analiz edilmesi amaçlanmıştır. Analizlerde kullanılan veri seti Türkiye uygulamasının okuma testine aittir. Okuma testi ortak ve farklı maddeler içeren 14 kitapçıktan oluşmaktadır. Çalışmada 1, 2, 9 ve 13. kitapçıklarda ortak olarak yer alan ve ikili puanlanan (dichotomous) 10 madde BILOG programı ile analiz edilmiştir. Sonuçta maddelerin MTK varsayımlarını (tek boyutluluk, yerel bağımsızlık, hız testi olmama) sağladığı ve 2 parametrelili lojistik modele uyum sağladığı görülmüştür. Madde test kuramına ve klasik test kuramına dayalı olarak madde parametreleri elde edilmiş, elde edilen madde ve test bilgi fonksiyonları yorumlanmıştır.

Anahtar Sözcükler: PISA, madde tepki kuramı, 2 parametrelili lojistik model.

EXAMINING THE TWO CATEGORICAL DATAS OBTAINED FROM PISA 2009 LITERACY TEST BY USING BILOG PROGRAMME

Abstract

This study aims to analyze PISA 2009 data with item response theory. The data set used in analysis is a part of PISA 2009 Turkey implementation's reading test. Reading test consists of 14 booklets. Booklets include common and different items. In this research, common dichotomous items of booklet 1, 2, 9 and 13; are analyzed with BILOG program. The results showed that; the data set satisfies item response theory assumptions (unidimensionality, local independence, unspeedness) and fits 2 parameter logistic model. Item parameters are obtained regarding to classical test theory and item response theory. Item and test information functions are interpreted.

Key Words: PISA, item response theory, 2 parameter logistic model.

GİRİŞ

Türkiye'de yıllık veya birkaç yılda bir ulusal ve uluslararası ölçekte pek çok sınav yapılmaktadır. Ulusal düzeyde yapılan sınavların büyük kısmı Ölçme, Seçme ve Yerleştirme Merkezi (ÖSYM) ve Milli Eğitim Bakanlığı (MEB) tarafından düzenlenmektedir. Bu sınavlarda amaçlar öğrencilerin/adayların ölçülmesi, seçilmesi ve yerleştirilmesi noktalarında yoğunlaşmaktadır. Örnek verecek olursak Yükseköğretime Giriş Sınavı (YGS) ile yükseköğretim kurumlarında öğrenim görmeye hak kazanacak öğrencilerin seçilmesi amaçlanırken, Lisans Yerleştirme Sınavı (LYS) ile bu öğrencilerin öğrenim görecekleri yükseköğretim puanına yerleşmelerini sağlayacak puanların elde edilmesi amaçlanır.

Uluslararası sınavlar ise düzenleme ve amaç bakımından ulusal sınavlardan farklıdır. Bu sınavlar Türkiye'nin üyesi veya ortağı olduğu topluluklar tarafından organize edilmektedir. Uluslararası Öğrenci Değerlendirme Programı (PISA) ve Üçüncü Uluslararası Matematik ve Fen Bilgisi Çalışması (TIMSS) Türkiye'nin katıldığı uluslararası eğitim araştırmalarıdır. TIMSS ilk olarak 1994-1995 yıllarında gerçekleştirilmiştir. Eğitim politikasını belirleyenlerin, öğretim programlarını hazırlayan uzmanların ve araştırmacıların kendi eğitim sistemlerinin işleyişini daha iyi anlayabilmeleri açısından bir temel sağlamak amacıyla düzenlenmiş olan TIMSS, ilk olarak 41 ülkede beşinci sınıf düzeyindeki öğrencilerin Matematik ve Fen Bilgisi başarılarını karşılaştırmıştır. Her 4 yılda bir

düzenli olarak yapılmakta ve 4. ve 8. sınıf düzeyinde Matematik ve Fen Bilgisi alanlarında testler içermektedir (EARGED, 2003). Türkiye'nin ilk olarak 2003 yılında katıldığı Uluslararası Öğrenci Değerlendirme Programı – PISA (Programme for International Student Assessment), Ekonomik İşbirliği ve Kalkınma Örgütü OECD'nin üç yıllık aralıklarla düzenlemekte olduğu ve 15 yaş grubu öğrencilerin kazandıkları bilgi ve becerilerin değerlendirilmesine yönelik yapılan bir tarama araştırmasıdır. Bu proje kapsamında ülkemiz de dahil olmak üzere 52 ülkedeki 15 yaş grubu öğrencilerinin Okuma Becerileri, Matematik okuryazarlığı ve Fen Bilimleri okuryazarlığı konu alanları ölçülmekte; ayrıca öğrencilerin motivasyonları, kendileri hakkındaki görüşleri, öğrenme biçimleri, okul ortamları ve aileleri ile ilgili bilgiler toplanmaktadır. Buna göre bu araştırmalarda, öğrencilerin öğretim stratejileri hakkındaki düşünceleri gibi konuları içeren geniş ölçekli eğitim çıktılarının elde edilmesi ve temel konu alanlarındaki performanslarının incelenmesi sağlanmaktadır (MEB, 2010).

PISA projesinde kullanılan "okuryazarlık" kavramı öğrencinin bilgi ve potansiyelini geliştirip, topluma daha etkili bir şekilde katılmasını ve katkıda bulunmasını sağlamak için yazılı kaynakları bulması, kullanması, kabul etmesi ve değerlendirmesi olarak tanımlanmakta ve bu doğrultuda ölçmeler yapılmaktadır. Üçer yıllık dönemler halinde 2000 yılından itibaren gerçekleştirilen PISA projesinde, her bir dönemde bir konu alanına ağırlık verilmektedir. Buna göre 2000 yılında okuma becerilerine, 2003 yılında Matematik okuryazarlığına, 2006 yılında yapılan uygulamada ise Fen Bilimleri alanına ağırlık verilmiştir. 2009 yılından itibaren bu döngü tekrarlanmaktadır. PISA projesi kapsamında öğrencilere çoktan seçmeli, karmaşık çoktan seçmeli, açık uçlu, kapalı uçlu gibi değişik soru türleri yöneltilmektedir. (MEB, 2010).

PISA'da maddeler öğrencilere farklı maddelerden oluşan kitapçıklar halinde verilmektedir. Bu sayede tüm öğrencilerin aynı soruları almasıyla oluşacak güvenlik tehdidinin önüne geçilmektedir. Ancak farklı maddeleri alan öğrencilerin başarılarının değerlendirilmesi, maddelere ve teste ilişkin kestirimlerim yapılması klasik test kuramı (KTK) ile istendiği gibi yapılamamaktadır. Çünkü KTK'da konu edinilen gerçek puan ve test puanı değişkenlerinin teste ve gruba bağımlıdır (Hambleton ve Jones, 1993). Bu sebeple yapılan ölçekleme işlemlerinde madde tepki kuramından (MTK) faydalanılmaktadır. Örtük özellikler kuramı olarak da adlandırılan MTK, bireylerin maddelere verdikleri cevaplara ve olasılıklarına dayanan model temelli matematiksel bir kuramdır (Embretson ve Reise, 2000). Kuram temel olarak birey, madde ve test performansının testteki maddelerin ölçtüğü performans ile bağlantısını kuran istatistiksel bir teoriye dayanır. MTK da konu edilen maddeler kesikli veya sürekli, iki kategorili veya çok kategorili, sıralı veya sırasız bir biçimde puanlanabilir. Kuram kapsamında maddeye verilen cevap ve altında bulunan yetenek veya yetenekler arasındaki ilişkiyi ortaya çıkaran pek çok yol (model) tanımlanmıştır (Hambleton ve Jones, 1993).

MTK'da iki kategorili ve çok kategorili olmak üzere 2 model tanımlanmıştır. İki kategorili modeller bir parametrelili, iki parametrelili ve üç parametrelili olmak üzere 3 lojistik model şeklinde adlandırılır. Buna göre iki kategorili lojistik modeller 1-0 şeklinde puanlanan doğru-yanlış, boşluk doldurma testleri ile kısa cevaplı, eşleştirmeli ve çoktan seçmeli testlere uygulanır. 1-0 şeklinde puanlanmayan maddeler ise çok kategorili modeller ile incelenir. Başlıca çok kategorili modeller Samejima'nın Aşamalı tepki modeli (Graded Response Model (GRM)), Master'in kısmi puan modeli (Partial Credit Model (PCM)) ve Muraki'nin genelleştirilmiş kısmi puan modeli (Generalized Partial Credit Model (GPCM)) şeklinde sıralanabilir (Embretson ve Reise, 2000; Reckase, 2009). Bu çalışmada analiz edilen veri seti iki kategorili olduğu için lojistik modeller kısaca tanıtılacaktır.

Rasch modeli olarak bilinen bir parametrelili lojistik modelde tüm maddelerin ayırt ediciliği (a) ve şans parametresi (c) aynıdır. Bu modelde testteki maddelerin güçlük parametresi (b) maddeden maddeye değişmektedir. İki parametrelili lojistik modelde maddelerin güçlüğü ve ayırt ediciliği değişim göstermekte, tüm maddelerin şans parametresi aynı değerdedir. Üç parametrelili lojistik modelde ise testteki tüm maddelerin ayırt edicilik, güçlük ve şans parametresi değişim göstermektedir (Baker, 2001). Üç lojistik modelde de madde karakteristik eğrileri elde edilir. Madde karakteristik eğrisi tüm yetenek düzeyleri için bir maddenin doğru cevaplanma olasılığını güçlük, ayırt edicilik ve şans parametresi ile birlikte veren eğrilerdir. Her bir eğri değişim gösteren madde parametrelerini içeren matematiksel bir eşitlik ile ifade edilmektedir (Embretson ve Reise, 2000). Eğriler aracılığı ile yeteneği bilinen bir bireyin maddeyi doğru cevaplama olasılığı da belirlenebilir.

İlgili alan yazında iki kategorili lojistik modellere ilişkin çok sayıda çalışma mevcuttur. Reckase (1978) 1- ve 3-parametrelili lojistik modelleri madde kalibrasyonu açısından karşılaştırmış, her iki modelin de yetenek kestirimi bakımından korelasyon gösterdiği, üç parametrelili lojistik modelin daha fazla örnekleme ihtiyaç duyduğu ancak veriye daha iyi uyum sağladığı sonuçlarına ulaşmıştır. Kingston (1985), 3 parametrelili lojistik modelin şans parametresini GRE (Graduate Record Examination) verisi üzerinden incelemiş ve c-parametresinin tahmin parametresi olarak ele alınabileceği sonucuna ulaşmıştır. Pere ve Eliseo (2007), iki parametrelili lojistik modele dayalı puanların geçerliğini KTK ve MTK karşılaştırmaları ile incelemiş ve iki kurama dayalı hesaplanan geçerlik katsayılarının benzer sonuçlar ürettiği sonucuna ulaşmışlardır.

Bu çalışmada 2009 PISA okuma testinde yer alan 10 maddenin iki kategorili lojistik modellerle incelenmesi, test için uygun modelin belirlenmesi ve analiz sonucunda elde edilen KTK ve MTK çıktılarının yorumlanması amaçlanmıştır. Çalışma gerçek veri seti üzerinden MTK ile madde kalibrasyonu içermesi ve KTK sonuçlarını içermesi bakımından önem taşımaktadır.

YÖNTEM

Bu bölümde araştırmanın türü, çalışma grubu, veri toplama araçları ve verilerin analizi kısımlarına yer verilmiştir.

Araştırmanın Türü

Bu çalışmada PISA 2009 uygulaması okuma becerileri testinden uygulamasından elde edilen veri seti madde tepki kuramının lojistik modelleri incelenmiştir. Araştırma ilgili veri setinin bilinen yöntemler analiz edilip yorumlanmasını içermesi bakımından betimseldir.

Çalışma Grubu

Araştırma PISA 2009 uygulamasına Türkiye'den katılan öğrencilere dayandırılmıştır. Ancak her öğrenci aynı maddeleri almadığı için çalışma grubu oluştururken ortak maddeleri alan öğrencilere dikkat edilmiştir. Bu doğrultuda kitapçık 1, 2, 9 ve 13'ü cevaplayan 1504 öğrenci çalışma grubunu oluşturmuştur.

Veri Toplama Araçları

Çalışmada kullanılan veriler 2009 PISA okuma testi Türkiye uygulamasından elde edilmiştir. Veri seti PISA resmi web sitesinden alınmıştır (<http://www.oecd.org/pisa/>). PISA 2009 okuma testinde 13 kitapçık ve her kitapçıkta farklı kombinasyonu olacak şekilde toplamda 131 çoktan seçmeli madde yer almaktadır. Çalışmada kitapçık 1, 2, 9 ve 13'te ortak olarak yer alan r5, r17, r18, r19, r26, r27, r28, r29, r30 ve r31 maddelerinden oluşan 10 madde analiz edilmiştir.

Verilerin Analizi

Çalışmada verilerin analizi iki aşamalı olarak gerçekleştirilmiştir. Buna göre ilk aşamada veri setine ait betimsel istatistikler belirlenmiş ve madde tepki kuramının varsayımları test edilmiştir. İkinci aşamada model veri uyumu belirlenerek, uygun bulunan modele ilişkin olarak kestirilen madde, test ve birey parametreleri elde edilmiştir. İlk aşamada STATISTICA, ikinci aşamada BILOGMG 3.0 ve SPSS programlarından faydalanılmıştır. Veriler doğru (1), yanlış (0) olacak şekilde iki kategorilidir. Veri setine ilişkin betimsel istatistikler aşağıdaki gibidir.

Tablo 1: Toplam Puanı Dayalı Betimsel İstatistikler

	Veri Seti
Birey	1504
Madde	10
Ortalama	5,646
Std. Sapma	2,218
Çarpıklık Katsayısı	0,63
Basıklık Katsayısı	0,126

Testte 10 maddenin yer alması sebebiyle alınabilecek en yüksek puan 10'dur. Betimsel istatistiklere göre ortalama 5,646'dır. Ayrıca çarpıklık ve basıklık değerlerine göre veri seti olsa sivri ve sağa çarpık bir grafik

oluşturmaktadır. Ancak bu değerlerin +1,-1 aralığında olması sebebi ile veri setinin normal dağılıma sahip olduğu söylenebilir (Tabachnick ve Fidell, 2007).

Madde Tepki Kuramının Varsayımları ve Test Edilmesi

Veri setinin madde tepki kuramı ile incelenmesine geçmek için öncelikle varsayımların karşılanıp karşılanmadığının araştırılması gereklidir. Burada öncelikle karşılanması gereken varsayımlar:

1. Tek boyutluluk,
2. Yerel Bağımsızlık,
3. Testin hız testi olup olmadığının kontrol edilmesidir. Aşağıda varsayımlar tek tek ele alınmış ve kontrol edilmiştir.

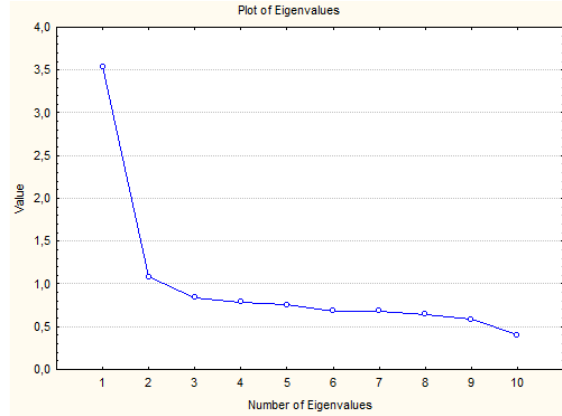
Tek boyutluluk

Tek boyutluluk testle sadece bir yeteneğin ölçülmesidir. Bu varsayımın tam olarak karşılanmasının güç olmasına karşılık, karşılanması için test performansını etkileyen baskın bir faktörün olması yeterlidir. Tek boyutluluğun test edilmesi için faktör analizi yapılması yeterlidir (Crocker ve Algina, 1986). Ancak burada dikkat edilmesi gereken verinin iki kategorili olmasından dolayı tetrakorik korelasyondan faydalanarak faktör analizi yapılması gerektiğidir. Bunun için STATISTICA programı ile faktör analizi yapılmış ve aşağıdaki çıktılar elde edilmiştir.

Tablo 2: Faktör Analizi Çıktıları

Faktör	Özdeğer	Varyans Yüzdesi	Yığılmalı Yüzde
1	3,538411	35,38411	35,38411
2	1,079566	10,79566	46,17976

Şekil 1: Özdeğer Grafiği



Tablo 2’de yer alan faktör analizi çıktılarına göre ilk faktör toplam varyansın %35,39’unu açıklamaktadır. Bu faktöre ilişkin özdeğer ikinci faktöre ilişkin özdeğerin 3 katından fazladır. Ayrıca özdeğer grafiğine göre birinci boyuttan sonra keskin bir düşüş izlenmektedir. Bu göstergeler sonucunda veri setimizin tek boyutlu olduğuna karar verilmiş ve analizlere devam edilmiştir.

Yerel Bağımsızlık

Yerel bağımsızlık varsayımı tek boyutluluk varsayımı ile ilişkilidir. Yerel bağımsızlık bir bireyin farklı maddelere verdiği tepkilerin istatistiksel olarak bağımsız olması demektir. Bunun bir sonucu olarak bireyin bir maddedeki performansı diğer maddedeki performansını etkilememelidir. Sabit bir yetenek düzeyinde bulunan bireylerin elde ettikleri test puanları da aynı olmalıdır. Ancak maddeler istatistiksel olarak bağımsız değil ise aynı yetenek düzeyindeki bireylerin test puanları farklı olacaktır. Bu durumda test performansı birden fazla yetenek ile açıklanmakta; aynı zamanda yerel bağımsızlık varsayımı ihlal edilmektedir. Bu varsayımın karşılanmaması tek boyutluluk varsayımının da karşılanmayacağı anlamına gelir. Yerel bağımsızlık madde setindeki ilişkilerin tek yetenek ile açıklanması olduğundan; tek boyutluluk varsayımının karşılanması yerel bağımsızlık varsayımının da

sağlandığı şekilde yorumlanabilir (Hambleton ve Swaminathan, 1985:23). Bu sebeplerden dolayı PISA 2009 okuma testinden elde edilen verilerin yerel bağımsız olduğu kabul edilmiştir.

Testin Hız Testi Olup Olmadığının Kontrol Edilmesi

Testin hız testi olup olmadığı testte yer alan maddelere ulaşılma durumuna bakarak kontrol edilir. Analiz edilen veri setinde 1504 bireyin cevaplandığı 10 madde vardır. Maddeler 1-0 şeklinde kodlandığından testteki maddeleri boş bırakan veya maddelere ulaşamayan birey yoktur. Bu sebeple testin hız testi olmadığı söylenebilir.

BULGULAR

Bu bölümde maddelerin lojistik modellere uygunluğu incelenmiş, klasik test kuramı ve madde tepki kuramına dayalı madde parametreleri elde edilmiştir.

Test Maddelerinin Lojistik Modellere Uygunluğunun İncelenmesi

Testin uyumlu olduğu lojistik modeli belirlemek üzere veri seti 1-, 2- ve 3-parametrelili lojistik modellere göre analiz edilmiştir. Elde edilen -2LogLikelihood (-2LL) değerlerinin farklarından faydalanarak ki-kare testi yapılmıştır. Kullanılan -2LL istatistiklerinin değerleri ve her modelde yer alan parametre sayıları tablo 3'te verilmiştir.

Tablo 3: Model Uyum İstatistikleri

	1PL	2PL	3PL
-2LL	16549,4858	16427,4886	16542,6945
PARAMETRE SAYISI	10	20	30

Tablo 3'te yer alan -2LL değerlerinin farkları 10 serbestlik derecesi ve 0,05 önem derecesinde ki-kare tablosu ile karşılaştırılmıştır. Burada -2LL farklarını bulmak üzere aşağıdaki eşitliklerden yararlanılmıştır.

$$X^2 = -2 \log \text{likelihood}_{1PL} - 2 \log \text{likelihood}_{2PL} = 121,9972$$

$$X^2 = -2 \log \text{likelihood}_{3PL} - 2 \log \text{likelihood}_{2PL} = 115,2059$$

Yapılan işlemler sonucunda -2PL modelin ($\chi^2=121,9972$; $sd=10$) veri setine uyum sağladığına karar verilmiş ve analizlerin 2PL modele ait çıktılara göre yorumlanmasına karar verilmiştir. Ancak MTK'da tüm maddelere karşılık gelen model uyumunun yanında her bir maddenin model ile uyumuna da odaklanılmaktadır (Demars, 2010). Maddelere ilişkin model veri uyumunun 2PL model ile incelenmesi sonucunda ilgili serbestlik derecelerinde aşağıdaki ki-kare değerleri elde edilmiştir.

Tablo 4: Madde Uyum İstatistikleri

	χ^2	sd		χ^2	sd
Madde 1	41352,00*	9.0	Madde 6	49.2*	8.0
Madde 2	30.2*	7.0	Madde 7	41517,00*	8.0
Madde 3	38.2*	7.0	Madde 8	41300,00*	8.0
Madde 4	41503,00*	7.0	Madde 9	54.0*	4.0
Madde 5	41500,00*	8.0	Madde 10	41541,00*	7.0

* $p<0,01$

2PL modele uyum gösteren madde istatistiklerine göre elde edilen ki-kare değerleri tablo değerlerinden büyüktür. Bu sonuçlara göre analizde yer alan maddelerin hepsi model ile uyumlu çıkmıştır ($p<0,01$). Bu sonuç veri seti ile 2PL modelin uyum gösterdiği bulgusunu desteklemektedir.

Klasik Test Kuramına Dayalı Olarak Elde Edilen Madde Parametreleri

Modele karar verilmesinin ardından çıktılar yorumlanmaya devam edilmiştir. BILOG çıktıları arasında klasik madde parametreleri kestirimi de yer almaktadır. KTK'ya dayalı madde istatistikleri aşağıdaki tabloda verilmiştir.

Tablo 5: Klasik Test Kuramına Dayalı İstatistikler

Madde No	Cevaplandırılan Sayısı	Doğru Cevaplandırılan Sayısı	Doğru Cevaplandırılan Yüzdesi	Lojit	Madde-Test Pearson Korelasyonu	Madde-Test Nokta Çift Serili Korelasyonu
1	1504	1173.0	78.0	-0.74	0.129	0.180
2	1504	235.0	15,6	0.99	0.267	0.404
3	1504	375.0	24,9	0.65	0.313	0.427
4	1504	1155.0	76.8	-0.70	0.282	0.390
5	1504	1127.0	74.9	-0.64	0.291	0.397
6	1504	445.0	29,6	0.51	0.282	0.372
7	1504	755.0	50.2	0.00	0.338	0.423
8	1504	951.0	63.2	-0.32	0.345	0.442
9	1504	1138.0	75.7	-0.67	0.449	0.615
10	1504	907.0	60.3	-0.25	0.360	0.457

Yukarıdaki istatistiklere göre en kolay maddenin %78,0 doğru cevaplanma oranı ile madde1'dir. En zor maddenin ise %15,6 doğru cevaplanma oranı ile madde18'dir. Test puanları sürekli madde puanları yapay süreksiz olduğu için madde-test puanları arasındaki korelasyonu nokta-çift serili korelasyon katsayısına dayalı olarak incelenebilir. En büyük nokta çift serili korelasyon katsayısı değerine sahip madde 9. madde (0.615) iken, en düşük korelasyona sahip madde 1. maddedir (0.180).

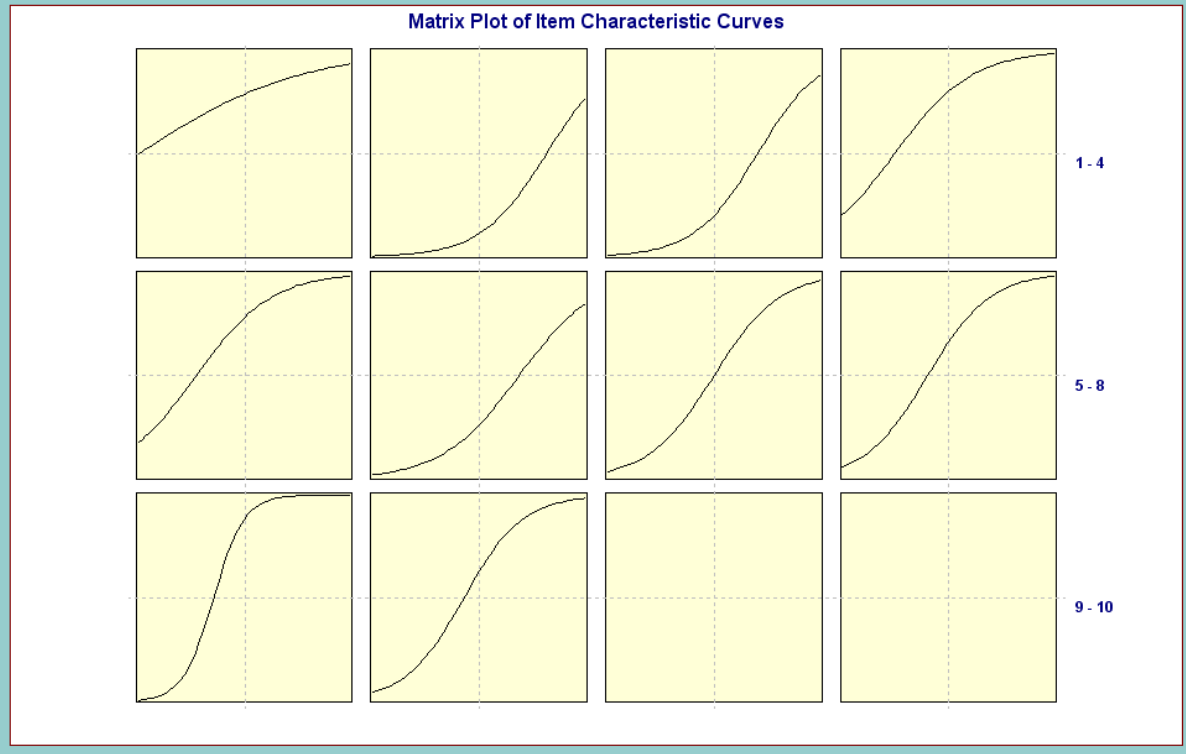
Madde Tepki Kuramına Dayalı Olarak Kestirilen Madde Parametreleri

Veriler 2PL model ile incelendiği için analizler sonucunda ayırıcılık ve madde güçlüğü parametreleri kestirilmiştir. Elde edilen çıktılar aşağıdaki gibidir.

Tablo 6: MTK'ya Dayalı Madde Parametreleri

	a	b		a	b
Madde 1	0,254	-3,059	Madde 6	0,527	1,142
Madde 2	0,623	1,941	Madde 7	0,617	-0,01
Madde 3	0,644	1,257	Madde 8	0,652	-0,619
Madde 4	0,527	-1,565	Madde 9	1,32	-0,907
Madde 5	0,539	-1,411	Madde 10	0,692	-0,461

Madde ayırıcılık katsayısı 1,32 ile 0,254 arasında; madde güçlüğü parametresi ise -3,059 ile 1,142 arasında değerler almaktadır. Maddeler için çizilen karakteristik eğriler ayırıcılık ve güçlük parametrelerinin ikisi hakkında da bilgi verir. Elde edilen eğriler aşağıdaki gibidir.



Şekil 2: Madde Karakteristik Eğrileri Matrisi

Madde karakteristik eğrilerinde de görüldüğü gibi madde9 ayırıcılık yönünden; madde 7 güçlük yönünden ideale yakın görünüme sahiptir. MTK ile elde edilen güçlük parametresi KTK ile elde edilen doğru cevaplama oranına paralel sonuçlar üretmiştir. Buna göre MTK'ya göre güçlüğü en yüksek madde (madde2) KTK'ya göre en düşük doğru cevaplanma oranına; en kolay madde (madde1) KTK'ya göre en yüksek doğru cevaplanma oranına sahiptir.

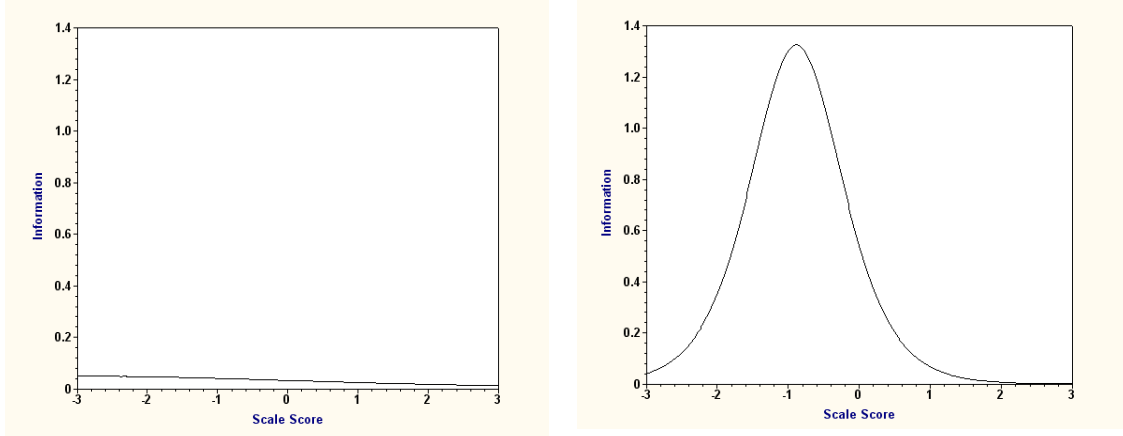
Madde ve Test Bilgi Fonksiyonu

Madde bilgi fonksiyonu maddenin ilgilenilen özelliğe ilişkin olarak ne kadar anlam taşıdığını verir. Bilgi fonksiyonlarına ilişkin olarak kestirilen standart hata değerleri fonksiyonun ne kadar hata ile kestirildiğini verir ve sıfıra yakın olması istenir. Bilgi fonksiyonunun değeri arttıkça maddeye ilişkin kestirilen hata düşmektedir (Hambleton ve Swaminathan, 1985). Benzer tanımlar test bilgi fonksiyonu için de yapılabilir (Embretson ve Reise, 2000). Aşağıda madde bilgi fonksiyonunun maksimum değerleri ve maddelerin bu değeri aldığı theta düzeyleri verilmiştir.

Tablo 7: Madde Bilgi Fonksiyonu

	Max. Bilgi Miktarı	(SE*)	Max. Bilgi Theta	(SE*)
Madde 1	0,049	0,0153*	-2,982	0,4570*
Madde 2	0,2951	0,0634*	1,8898	0,1598*
Madde 3	0,3161	0,0600*	1,2237	0,1012*
Madde 4	0,2113	0,0441*	-1,5267	0,1409*
Madde 5	0,2212	0,0432*	-1,3764	0,1237*
Madde 6	0,2111	0,0406*	1,1114	0,1073*
Madde 7	0,2897	0,0497*	-0,0109	0,0601*
Madde 8	0,3239	0,0560*	-0,6047	0,0685*
Madde 9	1,3261	0,2808*	-0,8852	0,0555*
Madde 10	0,3644	0,0608*	-0,4509	0,0614*

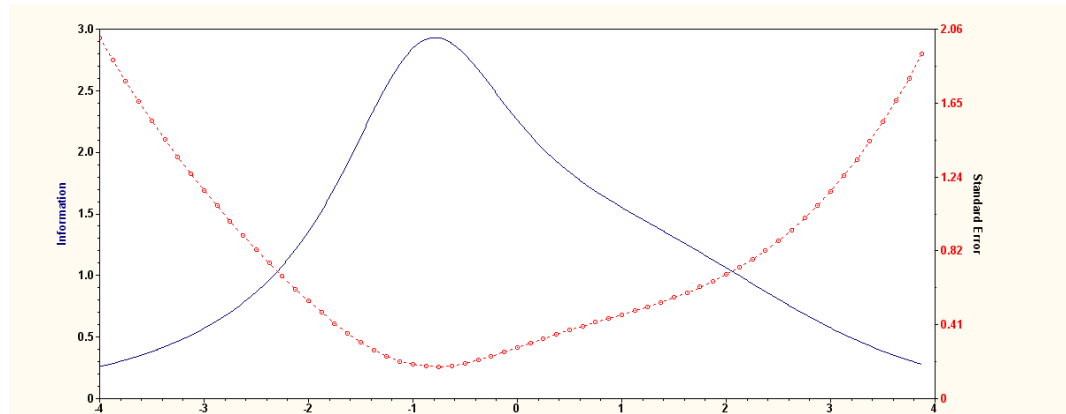
Bilgi fonksiyonuna ilişkin çıktılarına göre en fazla bilgiyi -0,89 theta düzeyinde madde9 vermektedir. Bilgi fonksiyonuna ilişkin hata miktarı en küçük değerini madde1'de -2,982 theta yetenek düzeyinde almaktadır. Madde9 ve madde1'e ilişkin madde bilgi fonksiyonları aşağıda verilmiştir.



Madde 1

Madde 9

Şekil 3: Madde Bilgi Fonksiyonu Eğrileri



Şekil 4: Test Bilgi Fonksiyonu

Analizlere dayalı olarak aşağıdaki test bilgi fonksiyonu elde edilmiştir. Bu fonksiyonda sürekli çizgi test bilgisini, kesikli çizgi standart hatayı temsil etmektedir.

Test bilgi fonksiyonu grafiğine göre testin sağladığı maksimum bilgi -0,750 theta düzeyindedir ve değeri 2,9269'dir. Standart hata bu noktada minimum değerini almaktadır. Testin tamamı için elde edilen marjinal güvenilirlik katsayısı 0,678'dir.

Madde Parametrelerinin Değişmezliği

MTK ile yapılan kestirimlerin sonucunda parametre değişmezliğinin gösterilmesi gerekir. Değişmezlik madde parametrelerinin gruptan bağımsız olmasıdır. Hambleton ve Swaminathan (1985), madde parametrelerinin değişmezliğinin evrenden çekilen iki veya daha fazla alt gruptan (örneğin erkekler ve kızlar, üst ve alt grup gibi) elde edilen madde parametrelerinin karşılaştırılarak kontrol edilebileceğini belirtmişlerdir. Madde parametrelerinin değişmezliğinin karşılanıp karşılanmadığını belirlemek üzere; veri setinde yer alan cinsiyet değişkenine dayalı olarak, veriler kızlar-erkekler şeklinde ikiye bölünmüş ve parametre kestirimi tekrarlanmıştır. Veri setleri ile yapılan analizler sonucunda elde edilen madde parametreleri aşağıda verilmiştir.

Tablo 8: Alt Gruplarda Elde Edilen Madde Parametreleri

	Erkekler		Kızlar	
	a	b	a	b
Madde 1	0,254	-2,679	0,303	-2,981
Madde 2	0,601	2,095	0,65	1,788
Madde 3	0,649	1,362	0,636	1,156
Madde 4	0,473	-1,645	0,628	-1,434
Madde 5	0,534	-1,214	0,555	-1,62
Madde 6	0,668	0,816	0,442	1,55
Madde 7	0,682	0,022	0,559	-0,048
Madde 8	0,658	-0,45	0,639	-0,819
Madde 9	1,210	-0,86	1,386	-0,978
Madde 10	0,659	-0,341	0,721	-0,591

Varsayımın karşılanabilmesi için kızlardan elde edilen parametreler ile erkeklerden elde edilen parametrelerin korelasyon göstermesi gereklidir. Elde edilen parametrelerin korelasyonu aşağıda verilmiştir.

Tablo 9: Alt Gruplarda Elde Edilen Madde Parametreleri Korelasyonu

	Madde Güçlük Parametresi (b)	Ayrırt Edicilik Parametresi(a)
Erkekler-Kızlar	0.909**	0.975**

** $p < 0,01$

Yukarıda görüldüğü gibi iki alt grupta da kestirilen madde güçlük ve ayrırt edicilik parametreleri yüksek korelasyon göstermektedir. Madde güçlük parametreleri arasındaki korelasyon 0,909 ($r < 0,01$) ve ayrırt edicilik parametreleri arasındaki korelasyon 0,975 ($p < 0,01$) olarak kestirilmiştir. Değerlerin 1'e yakın olması madde parametrelerin gruptan bağımsız olarak kestirildiği ve değişmezlik varsayımının sağlandığı anlamına gelmektedir.

SONUÇ

Bu çalışma ile PISA 2009 okuma testi Türkiye datasından elde edilen 10 maddelik ve 1504 kişilik veri seti madde tepki kuramına ait 3 lojistik model ile test edilmiştir. Lojistik modeller test edilmeden önce veri seti MTK varsayımları bakımından test edilmiştir. Veri setine ilişkin STATISTICA ile gerçekleştirilen faktör analizine göre tek boyutlu, dolayısıyla da yerel bağımsız olduğu gösterilmiştir. Maddelerin cevaplanma oranına bakarak hız testi olup olmadığı kontrol edilmiş, tüm bireylerin tüm maddeleri cevaplaması açısından hız testi olmadığına karar verilmiştir. Daha sonra lojistik modellere göre analizler yapılmış; sonuçta -2LL farklarına dayalı olarak veri setine en uygun modelin 2PL model olduğuna karar verilmiştir.

2PL modele göre ayrırt edicilik katsayısı 1,32 ile 0,254 arasında; madde güçlüğü parametresi ise -3,059 ile 1,142 arasında değerler almaktadır. Güçlük parametresine bakımından iki uçta yer alan (en güç, en kolay) maddeler hem KTK hem de MTK'ya göre aynı maddeler olarak kestirilmiştir. Madde bilgi fonksiyonuna göre en yüksek bilgi -0,89 theta düzeyinde 1,33 değeri ile madde9 tarafından sağlanmaktadır. Test bilgi fonksiyonu ise en yüksek değerini -0,75 theta düzeyinde 2,93 değerini almaktadır.

Madde parametrelerinin değişmezliği için cinsiyete dayalı olarak erkekler ve kızlar alt grupları oluşturulmuştur. Alt grupların analizi ile madde parametreleri kestirilmiştir, kestirilen güçlük parametrelerin korelasyonu 0,909 ($p < 0,01$) ve güçlük parametresinin korelasyonu 0,975 ($p < 0,01$) olarak elde edilmiştir. Parametre korelasyonlarının aldığı değerler ile parametrelerin değişmez olduğu gösterilmiştir.

Sonuç olarak PISA 2009 okuma testinden elde edilen 1504 birey ve 10 maddelik veri seti madde tepki kuramının varsayımlarını karşılamaktadır ve veriler 2 parametrelili lojistik model ile analiz etmeye uygundur.

Not: Bu çalışma 25-27 Nisan 2013 tarihlerinde Antalya’da 28 Ülkenin katılımıyla düzenlenen “International Conference on New Trends in Education - ICONTE-2013”da sözlü bildiri olarak sunulmuştur.

KAYNAKÇA

Baker, F. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation. (ERIC Document Reproduction Service No. ED458219).

Crocker, L. and Algina, J. (1986). *Introduction to Classical And Modern Test Theory*. N.Y: CBS Collage Publishing Company.

Demars, C. (2010). *Item response theory [Series in Understanding Statistic: Measurement]*. New York: Oxford University Press.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Ferrando, P. J. And Chico, E. (2007). The External Validity of Scores Based on the Two-Parameter Logistic Model: Some Comparisons between IRT and CTT. *Psicologica: International Journal of Methodology and Experimental Psychology*, v28 n2 p237-257.

Hambleton, R.K. ve Jones, R.W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.

Hambleton, R. K. ve Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Kluwer Nijhoff Publishing, Baston.

Kingston, N. M. (1985). Assessing Guessing Behavior Using the Three-Parameter Logistic Model. Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, April 1-4, 1985).

Lord, F. M. (1953). The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*, Vol 13, 517-549.

PISA 2006 Nihai Raporu (2010) Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı.

Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.

Tabachnick, B. G. & Fidell, L.S. (2007). *Using Multivariate Statistics (5th Edt)*. USA: Pearson.