

## MİLLİ EĞİTİM BAKANLIĞI ÖZEL OKULLAR SINAVI VERİLERİNİN MADDE TEPKİ KURAMI MODELLERİNE UYUMU

### THE FIT OF MINISTRY OF NATIONAL EDUCATION PRIVATE SCHOOL EXAMINATION DATA TO ITEM RESPONSE THEORY MODELS

Hüseyin H.YILDIRIM\*, Gözde ÇÖMLEKOĞLU\*, Giray BERBEROĞLU\*\*

**ÖZET:** Bu çalışmada Milli Eğitim Bakanlığı Özel Okullar Sınavında kullanılan matematik ve fen bilgisi alt testlerinin Madde Tepki Kuramı modellerine uyumu incelenmiştir. İncelemede 4728 kişilik bir örneklem kullanılmıştır. Araştırmanın bulguları, testlerdeki verilerin tek boyutluluk, yerel bağımsızlık, düşük şans başarısı sayıtlarını karşıladığını göstermektedir. Ayrıca, yetenek parametrelerinin ve madde güçlük indislerinin değişmezliği de sağlanmaktadır. Ancak Kay-kare uyum istatistikleri kullanılan modellere göre farklılıklar göstermektedir. Sonuç olarak Madde Tepki Kuramı modellerinin Özel Okullar Sınavı değerlendirilmede kullanılabilir olduğu gözlenmiştir.

**Anahtar Sözcükler:** madde tepki kuramı, özel okullar sınavı

**ABSTRACT:** In the present study, the model-data fit of Item Response Theory models to the data obtained in the Ministry of National Education, Private School Examination, Mathematics and Natural Sciences subtests were studied. Total 4728 subjects were used in the analyses. The results indicated that the unidimensionality, local independence, zero chance level assumptions were met in the data set. Also, invariance of ability parameters and item difficulty indices were achieved in the data set. On the other hand, Chi-square fit statistics gave different results across the models. The results indicated that models of Item Response Theory were feasible in evaluating the data obtained in the Private School Examinations.

**Keywords:** item response theory, private school examination

## 1. GİRİŞ

Son yıllarda, özellikle geniş ölçekli testlerin geliştirilmesi ve değerlendirilmesinde Madde

Tepki Kuramı modellerinden yararlanılmaktadır. Bu kuramın modelleri klasik test kuramına göre farklı ölçek nitelikleri vermekte, bu da elde edilen test puanlarının daha karşılaştırılabilir nitelikte olmasını sağlamaktadır. Madde Tepki Kuramı, testi alan kişilerin yeteneklerinin kullanılan test maddelerinden, madde parametrelerinin ise testi alan gruptan bağımsız olarak kestirilmesine olanak vermektedir (Wright, 1968; Hambleton, Swaminathan ve Rofers, 1991; Van der Linden ve Hambleton, 1997). Kısaca (1) kestirilen yeteneklerin kullanılan madde gruplarından bağımsız olması, (2) ölçeklenen maddelerin kullanılan gruplardan bağımsız olması, şeklinde özetlenecek bu iki önemli özellik, bilgisayar ortamında bireyselleştirilmiş testlerin hazırlanması ve geliştirilmesi, güçlükleri birbirinden farklı testlerden elde edilen puan dağılımlarının eşlenmesi, madde yanlılığının belirlenmesi gibi, testlerin uygulama alanlarına dönük çalışmaların yanı sıra, test puanlarının geçerliğine ilişkin de bazı analizlerin yapılmasına olanak tanımaktadır (Lord, 1980; Weiss, 1983).

Yabancı literatürde bu alanla ilgili çok sayıda çalışma bulunmaktadır ((Fan, 1998; Cook, 1988; Englehard, 1990; Hambleton 1993; Miller, 1988). Türkiye’de Madde Tepki Kuramı ile ilgili araştırmalar olmasına rağmen, Öğrenci Seçme ve Yerleştirme Merkezince gerçekleştirilen Öğrenci Seçme Sınavı, Milli Eğitim Bakanlığınca gerçekleştirilen Ortaöğretim Kurumları

\* Arş. Gör., ODTÜ, Eğitim Fakültesi, OFMAE, 06531 – Ankara

\*\* Prof. Dr., ODTÜ, Eğitim Fakültesi, OFMAE, 06531 – Ankara

Seçme ve Yerleştirme Sınavı ile Özel Okullar Sınavı gibi geniş ölçekli testlerde karar verme aşamasında madde tepki kuramı kullanılmamaktadır (Berberoğlu, 1988; Ertkin, 1993; Yenal, 1995; Kılıç, 1999; Çelik, 2001; Çalışkan, 2000). Ancak ileriye dönük düşünüldüğünde yukarıda adı geçen sınavların değerlendirilmesinde Madde Tepki Kuramı modellerinin kullanılması kaçınılmaz görülmektedir. Bunun için en az iki önemli neden vardır. Birincisi söz konusu testler her yıl yeniden hazırlanmakta, ancak yıldan yıla kullanılan formların birbirine denkliği istatistiksel olarak incelenmemektedir. Bu da farklı yıllarda testi alan bireylerin puanlarını karşılaştırabilir olmaktan uzaklaştırmaktadır. Madde Tepki Kuramı modelleri ve teknikleri test puanlarını denklemede önemli katkılar sağlamaktadır (Kolen, 1995; Skaggs, 1986). Ayrıca bu kurama göre elde edilen standart puanların niteliklerinin daha iyi olduğu söylenebilir (Crocker, 1986). Diğer bir neden ise, bilgisayar ortamında bireyselleştirilmiş testler ülkemizde de kullanılmaya başlandığında, belli bir alt yapının hazırlanmış olması, en azından eldeki verilerin ve soru tiplerinin modellere uygunluğunun incelenmiş olması gerekliliğidir. Özellikle grubun yeteneğine uygun testlerin geliştirilmesinde üstün yönleri olan Madde Tepki Kuramının, öğrencilerle ilgili önemli kararların verilmesi aşamasında, daha nitelikli, amaca daha iyi hizmet eden testlerin hazırlanmasına katkı vermesi beklenmekte, bu beklentiler ülkemizde konuyla ilgili daha çok sayıda araştırma yapılmasını kaçınılmaz kılmaktadır.

Bu çalışmada Milli Eğitim Bakanlığı Özel Okullar Sınavı kullanılarak Madde Tepki Kuramı modellerinin, testlerden elde edilen verilere uyumu, fen ve matematik alt testleri için incelenmiştir. Verinin kullanılan modele uyumu söz konusu testlerle ilgili yapılabilecek bilgisayar ortamında bireyselleştirilmiş test uygulamaları, test formlarının denkliği gibi çalışmalar için ön koşuldur (Albenese, 1984; Hambleton, 1991; Van der Linden, 1995). Bu çalışmada veri-model uyumunun MEB Özel Okullar Sınavında kullanılan fen ve matematik alt testlerindeki sorular açısından incelenmesi süresince gerçekleştirilen analizlerin özeti verilecektir.

### 1.1. Temel Prensipler

Madde Tepki Kuramı, örtük özellik ve madde karakteristik eğrisi kavramları üzerine kurulmuştur.

Madde Tepki Kuramı kişinin her bir maddeye doğru cevap verme olasılığını dikkate almaktadır. Madde karakteristik eğrisi, madde puanının, \_ örtük özelliği üzerinde regresyonu ile oluşan fonksiyonun grafiğidir (Bkz. Şekil 1). \_ tanımı gereği, madde karakteristik fonksiyonu, örneklem grubuna bağlı olmaksızın değişmeden kalmaktadır. Bu nedenle, bu fonksiyonu tanımlayan parametreler değişmeyen madde parametreleridir (Lord, 1968). Madde Tepki Kuramında bir soruya doğru cevap verme olasılığının yetenek düzeylerine göre dağılım fonksiyonu madde karakteristik eğrisi olarak adlandırılmakta, bu eğriyi tanımlayan farklı matematiksel denklemler de modelleri oluşturmaktadır. Sıfır ve bir şeklinde puanlanan tek boyutlu testlerde en çok kullanılan modeller lojistik modellerdir. Lojistik modeller içerisinde madde karakteristik eğrisini üç ayrı parametre ile tanımlayan model üç parametrelili lojistik modeldir ve bu modelde; madde ayırdediciliği (a), madde güçlüğü (b) ve tahminle doğru cevabı bulma olasılığı (c) olmak üzere üç parametre söz konusudur (Baker, 1987). Madde ayırdedicilik indisi (a), madde karakteristik eğrisinin b noktasındaki, yani eğrinin büküm noktasındaki eğimidir. Madde güçlüğü, bir maddenin hangi yetenek grubunda daha çok bilgi verdiğini göstermektedir (Crocker, 1986). Diğer lojistik modeller, madde karakteristik eğrisinin şekliyle ilgili bazı varsayımlarda bulunmaktadır. Bu modeller ile testi alan kişilerin cevapları arasında uygunluk sağlanabilirse, madde karakteristik fonksiyonu ile ilgili bilgilere, kişilerin maddelere verdikleri cevaplardan yola çıkılarak ulaşılmaktadır.

Lojistik modellerde, madde karakteristik fonksiyonunun genel biçimi

$$f(x) = \frac{e^x}{1+e^x} \quad (\text{Denklem I})$$

şeklinindedir. Lojistik modeller parametre sayılarına göre isimlendirildiğinde x'in madde para-

metreleri cinsinden farklı biçimlerde tanımlanmasıyla üç farklı lojistik model oluşmaktadır. Şöyle ki: 3-parametrelili model

$$P_i(\theta) = c_i + \frac{(1-c_i)e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} \quad (\text{Denklem II})$$

formülüyle tanımlanmaktadır.

$P_i(\theta)$ :  $\theta$  yetenek düzeyindeki kişiler arasından rasgele seçilen bir kişinin  $i$  maddesini doğru cevaplama olasılığı

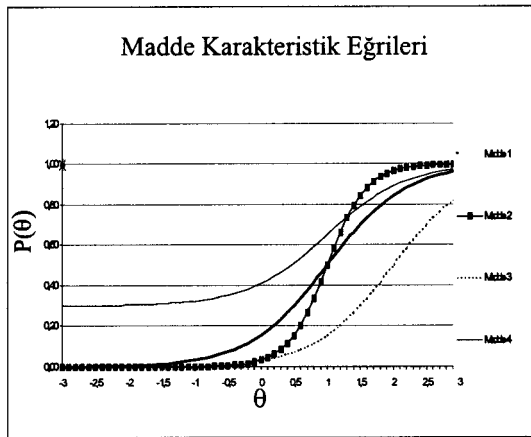
$b_i$ :  $i$  maddesi güçlük indisi

$a_i$ :  $i$  maddesi ayırdedicilik indisi

$c_i$ :  $i$  maddesi tahmin indisi' dir.

Denklem II de tüm maddeler için maddeyi tahminle doğru cevaplama olasılığı,  $c_i = 0$  olarak alınırsa 2-parametrelili modele, buna ilave olarak tüm maddelerdeki ayırdedicilik indisi ( $a_i$ ) sabit kabul edilirse 1-parametrelili modele ulaşılır.

Şekil 1 de farklı madde parametrelerine göre elde edilen örnek madde karakteristik eğrileri verilmiştir. Şöyle ki 1, 2, 3 ve 4. maddeler için (a, b, c) parametre değerleri sırasıyla (1, 1, 0), (2, 1, 0), (1, 2, 0) ve (1, 1, 0.3) şeklindedir.



Şekil1: Madde karakteristik eğrileri

Şekil 1 den görüldüğü üzere madde karakteristik eğrileri farklı madde parametreleri için farklı eğimde ve başlangıç noktasında olabilmektedir.

## 1.2.Araştırmanın Amacı

Bu çalışmada MEB Özel Okullar Sınavındaki fen ve matematik alt testlerinin üç modele göre veri-model uyumu incelenmiştir. Bu uyumda sırası ile aşağıdaki üç ana grupta ele alınan çalışmalar gerçekleştirilmiştir:

### 1. Model sayıtlarına uyum:

- Tek boyutluluk
- Yerel bağımsızlık
- Maddelerin ayırt edicilik parametrelerinin dağılımı (Bir parametrelili model için)
- Tahminle doğru cevap verme olasılığı (İki ve bir parametrelili modeller için)
- Testlerin hız testi olmaması.

### 2. Değişmezlik ilkesinin incelenmesi:

- Yetenek parametrelerinin değişmezliği.
- Madde parametrelerinin değişmezliği.

### 3. Belirlenen modellerden elde edilen teorik dağılımların gerçek durumdaki dağılımlara uyumunun incelenmesi:

- Kay-kare uyum istatistikleri.

Yukarıda amaçlanan veri-model uyumu Hambleton, Swaminathan ve Rodgers'ın önerdiği yöntemlerin bir bölümünü kapsamaktadır (Hambleton, 1985). Araştırmada her üç modele göre incelemeler yapılmış, kullanılan veriye uygun düşen modeller belirlenmeye çalışılmıştır.

## 2. YÖNTEM

### 2.1. Ölçme Aracı

Bu çalışmada, 2000 yılında İstanbul bölgesinde kullanılan Özel Okullar Sınavı verileri kullanılmıştır. Sınavda her biri 25 maddelik, Türkçe, sosyal bilgiler, matematik ve fen bilgisi olmak üzere 4 alt test bulunmaktadır. Maddeler çoktan seçmeli olup, doğru cevap 4 seçenek arasından seçilmektedir. Sorular, MEB İlköğretim Okulları Öğretim Programları kapsamındaki kavram ve ilkeler kullanılarak üst düzey zihinsel süreçleri ölçmeye yönelik hazırlanmaktadır. Özellikle fen bilgisi testinde gözlem yapma, gözlemlerden sonuç çıkarma, hipotez önerme,

deney kurma v.s gibi bilimsel yöntem sürecine dönük sorular yer almakta, bunun yanı sıra kavramsal problem çözme soruları bulunmaktadır. Matematik testinde ise, sıradan olmayan problem çözme, verilen çözüm içinde yanlış basamağı belirleme, bir problemin çözümü için yol önerme, gibi yine fen bilgisi testindeki gibi üst düzey düşünme süreçlerine yönelik sorular yer almaktadır. Sınav süresi 120 dakikadır ve düzeltilmiş ham puanlar hesaplanırken belli sayıdaki yanlış cevap toplamı toplam puandan düşülmektedir. Ancak bu çalışmada düzeltilmiş ham puanlar yerine toplam doğru cevap sayısından oluşan ham puanlar kullanılmıştır. Tablo 1 araştırmada kullanılan alt testlerle ilgili betimsel istatistikleri vermektedir.

**Tablo 1:** Fen ve matematik alt testleri betimsel istatistikleri

	Matematik Testi	Fen Testi
Arit. Ortalama.	13.25	16.15
Standart hata	0.07	0.07
Standart Sapma	5.00	4.92
Varyans	25.02	24.18
Çarpıklık	0.07	-0.33
Basıklık	-0.51	-0.56

Tablodan da görüldüğü gibi testler orta güçlüktedir. Ancak fen bilgisi testi matematik testine oranla daha kolay gözükmektedir. Cronbach alpha değerleri, matematik ve fen testi için sırasıyla 0.810 ve 0.828 olarak bulunmuştur.

## 2.2. Örneklem

Sınava yaklaşık 23000 öğrenci katılmıştır. Bu gruptan seçkisiz yolla 4728 öğrenci seçilmiştir. Seçilen örneklem tüm evrenin yaklaşık % 20 sini kapsamaktadır. Her yıl bu sınavı alan öğrencilerin aynı nitelikte olduğu ve test formlarının belli ölçülerde paralel olduğu kabul edilirse bu araştırma bulgularını geçmiş yıllarda Özel Okullar Sınavlarını alan tüm öğrencilere genellemek mümkündür.

## 2.3. İstatistiksel Analiz

Bu çalışmada Scientific Software [26] tarafından hazırlanmış olan BILOG programı kullanılmıştır. Program, En Çok Olabilirlik veya Bayesian yöntemleri ile madde parametrelerini ve yetenek düzeylerini kestirmektedir. Çalışmada ayrıca ITEMAN ve SPSS 10 paket programları da kullanılmıştır.

Analizlerde Hambleton, Swaminathan ve Rodgers'ın (Hambleton, 1985) veri-model uyumunun incelenmesi için önerdiği metotlardan yola çıkılarak aşağıda belirtilen çalışmalar yapılmıştır.

### 1. Model sayıtlarının incelenmesi:

a. Tek boyutluluk, SPSS programı kullanılarak temel bileşenler faktör çözümlemesi ile incelenmiştir. Testlerin tek boyutlu olup olmadığı Scree test ile değerlendirilmiştir (Hambleton, 1991; Crocker, 1986).

b. Yerel bağımsızlık, tüm gruptaki maddeler arası korelasyonlar, yetenek dağılımı daha dar olan alt ve üst yetenek gruplarındaki maddeler arasında elde edilen korelasyonlarla karşılaştırılarak incelenmiştir. Sorulara verilen cevap örüntülerinin birbirinden bağımsız olması anlamına gelen yerel bağımsızlık bir önceki sayıtlı olan tek boyutlulukla ilişkilidir. Test tek boyutlu ise sorularda yerel bağımsızlığa ulaşılmıştır anlamı çıkartılabilir [2,19].

c. Testteki maddelerin sabit ayırdedicilik düzeyine sahip olup olmadığı madde test puanı korelasyonu çift serili korelasyon katsayıları dağılımı incelenerek değerlendirilmiştir.

d. Maddeyi tahminle doğru cevaplama oranı, düşük yetenek grubundaki öğrencilerin testteki zor soruları cevaplama oranları incelenerek değerlendirilmiştir. Bu incelemede teorik olarak dört seçenekte tahminle doğru cevaplama olasılığı olan % 25 sınır alınmış, alt grupta madde güçlük indislerinin bu sınırın altında kalıp kalmadığına bakılmıştır.

e. Testlerin hız testi olup olmadığı her bir kişinin boş bıraktığı madde sayılarının varyansının, yanlış yaptıkları madde sayılarının varyansına oranına bakılarak değerlendirilmiştir. Bu oranın sifıra yakın olduğu durumlarda test hız testi olarak çalışmamaktadır (Gulliksen, 1950).

## 2. Değişmezlik ilkesinin incelenmesi:

a. Yetenek parametresinin değişmediğini incelemek için, maddeler kolay-zor, tek-çift olarak ikiye ayrılmış ve her bir madde grubundan kestirilen yetenek parametreleri arasındaki korelasyonlar incelenmiştir.

b. Madde parametrelerinin değişmezliğini incelemek için kişiler, seçkisiz yöntemle iki gruba ve üst-alt yetenek gruplarına ayrılmış, her bir grupta kestirilen madde parametreleri bu gruplar arasında korelasyon tekniği kullanılarak incelenmiştir.

3. Belirlenen modellerden elde edilen teorik dağılımların gerçek durumdaki dağılımlara uyumunun incelenmesi:

a. BILOG programından elde edilen Kay-kare değerlerinin anlamlılık düzeyi incelenmiştir. Burada test edilen hipotez, teorik madde karakteristik eğrileri ile gruptaki gerçek dağılımın aynı olduğu hipotezidir. Hipotezin reddedilmemesi kullanılan modelde, teorik ve gözlenen dağılımlar arasında uyum olduğuna işaret etmektedir.

## 3. BULGULAR

Araştırmanın amacında da belirtildiği gibi üç grup analiz yapılarak veri-model uyumu incelenmiştir. Sonuçlar bu gruplara göre özetlenmiştir.

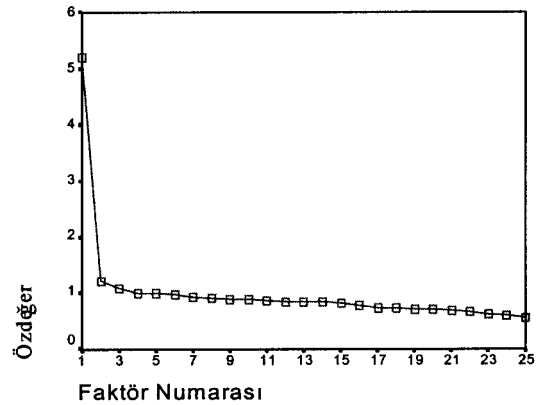
### 3.1. Model varsayımlarının geçerliği

Testlerin tek boyutlu olup olmadığını görmek için, temel bileşenler faktör çözümlemesi yapılmış, faktör eksenleri Varimax yöntemine göre döndürülmüştür. Analiz sonucunda elde edilen ilk üç faktörün özdeğerleri ve açıkladıkları varyans miktarı, matematik testinde 4.658, % 0.19; 1.251, %0.05 ve 1.168, %0.04 tür. Fen testinde ise bu değerler sırasıyla 5.2, %0.21; 1.2, %0.05; ve 1.1, % 0.04 tür. Şekil 2 de verilen Scree test grafikleri de ilk boyuttaki özdeğerlerin her iki test içinde ikinci boyuttan bir hayli büyük olduğunu, ikinci ve üçüncü boyutlardaki özdeğerlerin büyük ölçüde benzer büyüklükte olduğunu göstermektedir.

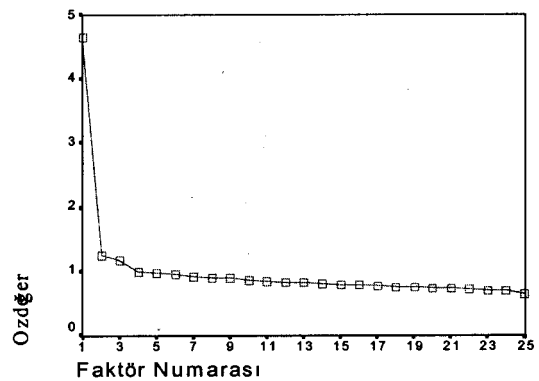
Sonuç olarak her iki testte de 1. boyuttan 2. boyuta özdeğerlerdeki keskin düşüş, testlerin tek boyutlu olduğuna işaret eden bir kanıt olarak kullanılmıştır.

Testlerde yerel bağımsızlığın sağlanıp sağlanmadığı, maddelere verilen cevapların kısıtlı yetenek dağılımında istatistiksel olarak bağımsız olması anlamına gelmektedir. Bu nedenle maddeler arasındaki korelasyonun sınırlı yetenek ranjında 0'a yaklaşması beklenmektedir. Ancak davranış bilimlerinde tam olarak sıfıra ulaşılması çok güç olduğu için aynı yetenek seviyesindeki kişilerin oluşturduğu alt gruplardan elde edilen maddeler arası korelasyonun tüm gruptan elde edilen maddeler arası korelasyona nazaran daha küçük olması yerel bağımsızlık için bir kanıt olarak kullanılmaktadır (Hambleton, 1991; Lord, 1968).

Fen Testi



Matematik Testi



Şekil2: Scree test grafikleri

Tablo 2 tüm grupta ve alt, üst yetenek gruplarında ayrı ayrı olmak üzere elde edilen maddeler arası korelasyon değerlerinin dağılımlarına ait özet bilgileri vermektedir.

**Tablo 2:** Tüm grup ve alt, üst yetenek gruplarından elde edilen maddeler arası korelasyon dağılım özeti

		Matematik	Fen
Tüm Grb.	Ort.	0.145	0.162
	Min	0.012	0.027
	Max.	0.295	0.390
Alt. Grb.	Ort.	-0.015	-0.013
	Min	-0.199	-0.304
	Max.	0.185	0.362
Üst Grb.	Ort.	-0.014	-0.011
	Min	-0.154	-0.106
	Max.	0.373	0.163

Alt grup her iki testten de 5 puan ve altında alan öğrencilerden, üst grup ise her iki testten de 22 puan ve üstünde alan öğrencilerden oluşmaktadır. Tablodan da görüldüğü gibi, hem testlerin tek boyutlu olması hem de alt ve üst yetenek gruplarındaki korelasyonların ortanca değerlerinin tüm gruptan küçük olması, veride yerel bağımsızlık sayılısının sağlandığına işaret eder niteliktedir. Bu sonucu tek boyutluluk için elde edilen bulgular da desteklemektedir.

Madde ayırdedicilikleri, her bir test için ayrı ayrı olmak üzere, her bir maddenin toplam test puanı ile olan çiftserili korelasyonları incelenerek değerlendirilmiştir. Bu sayılı yalnızca tek parametrelili model için geçerlidir. Tablo 3 ITE-MAN programı ile fen ve matematik testleri için elde edilen maddelerin ayırdedicilik indislerinin dağılımına ilişkin özet bilgileri vermektedir.

**Tablo 3:** Fen ve Matematik testlerine ait maddelerin ayırdedicilik indislerinin dağılımı özeti

	Matematik	Fen
Ortanca	0.569	0.490
Min	0.313	0.197
Max.	0.671	0.722

Her iki testte de ayırdedicilik indisleri dağılımının sabit olduğunu söylemek oldukça güçtür. Bu durum tek parametrelili modelin önemli bir sayılısı olan ayırt ediciliklerin sabit olduğu şartını sağlamamaktadır.

**Tablo 4:** Düşük yetenek grubundaki öğrencilerin testlerdeki en zor altı soruyu doğru olarak cevaplama yüzdeleri

	Madde	Tüm Gruptaki Güçlük	% Doğru
Matematik	17	0.397	9
	21	0.404	4
	23	0.287	6
Fen Bil.	09	0.485	5
	10	0.405	24
	23	0.439	5

Tablo 4 te, düşük yetenek grubundaki öğrencilerin matematik ve fen testlerindeki en zor üç soruyu doğru olarak cevaplama yüzdeleri verilmiştir. Madde tepki kuramındaki tek ve iki parametrelili modellerin önemli sayılılarından biri de tahminle doğru cevabı bulma olasılıklarının 0 olmasıdır. Çünkü bu modellerdeki madde karakteristik eğrileri 0 asimtot değerlerine sahiptir. Bu sayılıyı test etmek için, testlerdeki en zor sorularda, alt grup öğrencilerinin doğru cevap yüzdelerinin %25'in altında olup olmadığı incelenmiştir. Aslında söz konusu sınavda yanlış cevaplanan üç soru için bir doğru cevap silinmekte, böylece uygulama sırasında öğrencilerin tahminle doğru cevabı verme olasılıkları azaltılmaya çalışılmaktadır.

Düşük yetenek grubundaki öğrencilerin zor maddelerde %25 cevap verme olasılığının altında kaldığı görülmektedir. Ancak tek bir fen sorusu üst sınıra yakın doğru cevap yüzdesi vermektedir. Buna rağmen, genel olarak, hem uygulamada puan düzeltilmesi hem de Tablo 4 teki cevap verme oranları, testte tahminle doğru cevap verme olasılığının düşük olduğuna işaret etmektedir.

Testlerin hız testi olarak çalışıp çalışmadığı her bir kişinin boş bıraktığı madde sayılarının

varyansının, yanlış yaptıkları madde sayılarının varyansına oranı dikkate alınarak incelenmiştir. Öğrencilerin her bir soruyla ilgilenecek zamanlarının olması bu oranın 0'a yakın olmasını gerektirir. Bu oran matematik ve fen testleri için sırasıyla 0.89 ve 0.81 olarak elde edilmiştir. Bu oranın 1'e yaklaşması testin hız testi olduğuna işaret etmektedir. Madde tepki kuramının bu sayıltısı kullanılan veride ihlal edilmiştir.

### 3.2. Parametrelerin Değişmezliği

Madde ve yetenek parametrelerinin değişmezliği çalışılmadan önce, her iki testteki maddeler, kullanılan üç modele göre BILOG programı ile ölçeklenmiştir. Maddelerin ölçek değerlerine ilişkin özet bilgiler Tablo 5 te verilmiştir.

**Tablo 5 :** Madde parametreleri ölçek değerleri özeti

	A	B	Ortalama.	St. Sap.	Min.	Max.
Matematik	1	b	-0.18	0.97	-2.81	2.10
	2	a	0.59	0.18	0.26	0.93
		b	-0.09	1.14	-2.28	4.09
	3	a	0.79	0.31	0.34	1.86
		b	0.13	0.95	-2.15	2.78
		c	0.11	0.06	0.04	0.29
Fen Bilgisi	1	b	-0.81	1.03	-2.85	1.17
	2	a	0.71	0.26	0.23	1.24
		b	-0.73	1.02	-2.81	1.13
	3	a	0.88	0.26	0.42	1.33
		b	-0.38	1.06	-2.60	1.71
		c	0.160	0.08	0.05	0.36

A : Model

B : Madde parametreleri

Yetenek kestirimlerinin değişmezliği incelenirken, aynı grup öğrenciler için farklı güçlükteki sorulardan elde edilen yetenek parametrelerinin karşılaştırılabilirliğine bakılmıştır. Bu amaçla testteki sorular tek numaralı ve çift numaralı sorular olmak üzere ikiye bölünmüş, üç modele göre ayrı ayrı olmak üzere her öğrenci-

nin tek numaralı sorulardan elde edilen yetenek kestirimleri çift numaralı sorulardan elde edilen yetenek kestirimleri ile korelasyona sokulmuştur. Aynı işlem testteki kolay ve zor maddeler için de gerçekleştirilmiştir. Bu ayırmada maddelerin güçlüklerinin madde güçlük ortalamasının altında ve üstünde olmaları göz önünde bulundurulmuştur. Tek ve çift numaralı sorular teste eşit güçlükte iki yarı oluşturmaktadır. Tablo 6 söz konusu yetenek kestirimleri arasındaki Pearson momentler çarpımı korelasyon katsayılarını vermektedir.

**Tablo 6:** Farklı madde gruplarından kestirilen yetenekler arasındaki korelasyon katsayıları

		TekÇift maddeler	Kolay-zor maddeler
1-Prm	Mat	0.686*	0.631*
	Fen	0.722*	0.698*
2-Prm	Mat	0.704*	0.644*
	Fen	0.762*	0.728*
3-Prm	Mat	0.703*	0.645*
	Fen	0.760*	0.727*

\*p < 0.05

Matematik testindeki zor maddelerin klasik madde güçlük indisleri ortalaması 0.35, kolay maddelerin klasik güçlük indisleri ortalaması ise 0.71 dir. Fen testinde ise, zor maddelerin klasik madde güçlük indisleri ortalaması 0.33, kolay maddelerin klasik güçlük indisleri ortalaması ise 0.76 dır. Tablo 6 daki korelasyonlar incelendiğinde, kolay ve zor sorularda elde edilen korelasyonların tek ve çift numaralı sorulardan elde edilen korelasyonlardan çok farklı olmadığı görülmektedir. Madde güçlük ortalamaları birbirine yakın olan tek ve çift maddelerden elde edilen yetenek kestirimleri arasındaki değişmezliği gösteren korelasyon, madde güçlük ortalamaları birbirinden çok farklı olan kolay ve zor maddelerden elde edilen yetenek kestirimleri arasındaki değişmezliği gösteren korelasyona göre, matematik testinde ortalama 0.05, fen testinde ise ortalama 0.03 azalmıştır. Her üç modelde yetenek kestirimlerinde karşılaştırılabilir sonuçlar

elde edilmektedir. Modeller birbirleri ile karşılaştırıldığında, yetenek kestirimlerinin değişmezliği varsayımının en iyi 2-parametrelili modelde sağlandığı görülmektedir.

Madde parametrelerindeki değişmezlik için de, seçkisiz yöntemle iki ayrı grup ve üst-alt yetenek grupları oluşturulmuştur. Seçkisiz yöntemle ikiye ayrılan gruplar aynı yetenek düzeyini temsil etmektedir. Bu nedenle bu gruplarda elde edilen madde parametreleri arasındaki korelasyonlar, alt ve üst yetenek gruplarından elde edilen korelasyonlarla karşılaştırılarak madde parametreleri kestirimlerinin, kullanılan grubun yetenek düzeyinden ne kadar bağımsız olduğu incelenmiştir. Test ortalamasının altındaki grup alt, ortalamasının üstündeki grup üst yetenek düzeyi olarak tanımlanmıştır. Tablo 7, A ve B gruplarında kestirilen madde parametreleri arasındaki Pearson momentler çarpımı korelasyonlarını vermektedir.

**Tablo 7:** Kestirilmiş madde parametreleri arasındaki korelasyon katsayıları

		1-Prm		2-Prm		3-Prm	
		Mat	Fen	Mat	Fen	Mat	Fen
a prm	A	-	-	0.94	0.99	0.93	0.95
	B	-	-	0.07	0.42	0.28	0.12
b prm	A	0.99	0.99	0.99	0.99	0.99	0.99
	B	0.93	0.90	0.76	0.79	0.43	0.80
c prm	A	-	-	-	-	0.73	0.87
	B	-	-	-	-	0.10	0.31

A: Seçkisiz yöntemle seçilmiş iki grup arasında

B: Alt ve üst yetenek grupları arasında

Tablo 7 yakından incelendiğinde b parametresindeki değişmezliğin, üç parametrelili model hariç, büyük ölçüde sağlandığı görülmektedir. Ancak a ve c parametreleri için aynı şeyi söylemek mümkün gözükmemektedir. Seçkisiz yöntemle oluşturulmuş iki gruptan elde edilen madde güçlük (b) parametreleri arasındaki korelasyon, alt ve üst yetenek gruplarından elde edilen korelasyonla karşılaştırıldığında, korelasyon değerlerinin matematik testi için 2ve 3parametrelili

modelde sırasıyla 0.230, 0.559, fen testinde ise 0.210, 0.195 azaldığı görülmektedir.

Verideki cevap dağılımının kestirilen madde karakteristik eğrisine ne derece uyduğunu incelemek amacı ile Kay-kare uyum istatistikleri incelenmiştir. Tablo 8, 0.05 anlamlılık seviyesinde, her üç modele göre uyum vermeyen madde sayısını ve bu sayının tüm testteki oranını göstermektedir.

**Tablo 8:** Modele uymayan madde sayı ve yüzdeleri

		Matematik	Fen Bilgisi
1-Prm	n	18	19
	%	72	76
2-Prm	n	9	16
	%	36	64
3-Prm	n	11	7
	%	44	28

n: Modele uymayan madde sayısı

Kullanılan verinin modelle uyumunun incelenmesinde örneklem büyüklüğünden etkilendiği için yalnız Kay-kare testinin kullanılması veri-model uyumu çalışmalarında çok sağlıklı sonuçlar vermemektedir (Hambleton, 1991). Örneklemin çok büyük olduğu durumlarda, birçok maddenin modele uymadığı, dolayısıyla veri-model uyumunun olmadığı sonucuna ulaşılması mümkündür. Bu nedenle, kullanılan verinin hangi modele uygun olduğuna karar verilirken Hambleton, Swaminathan ve Rogers'ın önerdiği diğer kanıtlar da dikkate alınmalıdır (Hambleton, 1991). Tablo 8 den görüldüğü gibi Kay-kare testine göre en az uyum gösteren model 1-parametrelili modeldir. Özellikle 3-parametrelili modelde fen bilgisi testinde diğer modellere göre daha iyi uyum elde edilmiştir. Matematik testi verileri ise en iyi 2-parametrelili modele uymaktadır.



#### 4. SONUÇ

Bu çalışmada elde edilen bulgular genel olarak MEB Özel Okullar Sınavında kullanılan matematik ve fen bilgisi alt testlerindeki soruların Madde Tepki Kuramı modellerine uygun olduğunu göstermektedir. Kullanılan veri genel olarak tek boyutluluk, yerel bağımsızlık ve şans başarısından arınık olma ön sayıltılarını sağlamıştır. Buna bağlı olarak değişmezlik çalışmalarında, yetenek kestirimlerinin değişmezliği her üç model için de yeterli düzeyde sonuçlar vermektedir. Ancak madde parametrelerinin değişmezliği ilkesi yalnızca madde güçlük indisi için gözlenmiş, ayırdedicilik ve şans parametrelerinde bu özelliğe ulaşılamamıştır. Araştırmanın giriş bölümünde açıklanan bilgisayar ortamında bireyselleştirilmiş test uygulamaları, test eşleme çalışmaları gibi alanlarda Madde Tepki Kuramı modellerinin kullanılabilmesi için veri-model uyumu bir ön koşul olarak sağlanmış olmalıdır. Bu çalışma genel hatları ile Madde Tepki Kuramı modellerinin, söz konusu test için kullanılabilmesine işaret etmektedir. Ancak ayırdedicilik ve şans parametrelerindeki değişmezlik özelliklerinin elde edilememesi, özellikle üç parametreliliğin kullanılırken özen gösterilmesi gerektiğini ortaya çıkarmaktadır. Tek parametreliliğin genelde değişmezlik ilkesine uygun hareket etmekle birlikte Kay-kare uyum testlerinde beklendiği sonuçları vermemektedir. Bunun nedenleri ayırdedicilik indislerinin testteki maddeler için sabit olmaması olabileceği gibi, örneklemin Kay-kare testi için gereğinden fazla büyük olması da olabilir. Eğer testlerde ayırdedicilik indisi dikkate alınarak bir değerlendirme yapılmak isteniyorsa, iki parametreliliğin kullanılması daha uygun görülmektedir. Her üç modele göre test bilgi fonksiyonları incelendiğinde, matematik testinde tek, iki ve üç parametreliliğin modeller için en çok bilgi sırasıyla -1.78, -1.85, -1.14 yetenek düzeylerinde elde edilmektedir. Bu yetenek düzeylerinde her üç modelde elde edilen bilgi sırasıyla 5.03, 5.67, 7.09 dur. En çok bilgiyi 3 parametreliliğin model vermektedir. Fen testinde ise en çok bilgi modellere göre sırasıyla -2.14, -2.14, -1.85 yetenek düzeylerinde

elde edilmektedir. Bu yetenek düzeylerinde elde edilen bilgi ise sırasıyla 6.02, 8.43, 8.25 tir.

Bundan sonra yapılacak çalışmalarda parametrelerdeki değişmezliği etkileyen soru tiplerinin incelenmesinde, ölçeklemenin daha çok sayıda soruyla yapıldığında karşılaşılabilecek güçlüklerin araştırılmasında, elde edilen yetenek parametrelerinin seçilecek ölçüt puanlarını yordama gücünün belirlenmesinde, veri-model uyumunun farklı yöntemler kullanılarak (örneğin kalıntı analizleri, bireylerin modele uyumu gibi) irdelenmesinde yarar görülmektedir.

#### REFERANSLAR

- Albenese, M. A. & Forsyth, R. A. (1984). "The one-, two- and modified two parameter latent trait models: An empirical study of relative fit". **Educational and Psychological Measurement**, 44, 229-245.
- Baker, F. B. (1987). Methodology review: "Item parameter estimation under the one-, two-, and three-parameter logistic models". **Applied Psychological Measurement**, 11, 111-142.
- Berberoğlu, G. (1988). "Seçme amacıyla kullanılan testlerde Rasch modelinin katkıları". **Yayınlanmamış Doktora Tezi**. Hacettepe Üniversitesi, Eğitimde Ölçme ve Değerlendirme Bölümü, Ankara.
- Cook, L.L., Eignor, D. R., Taft, H. L. (1988). "A comparative study of the effects of recency of instruction on the stability of IRT and conventional item parameter estimates". **Journal of Educational Measurement**, 25: 31-45.
- Crocker, L. & Algina, J. (1986). "Introduction to classical and modern test theory" (pp. 339-371). New York: Holt, Rinehart & Winston.
- Çalışkan, M. (2000). "The fit of one-, two-, and three-parameter models of item-response theory to the ministry of national education-educational research and development directorate's science achievement test data" **Yayınlanmamış Yüksek Lisans Tezi**, ODTÜ, Eğitim Bölümleri Bölümü, Ankara.
- Çelik, D. (2001) "The fit of one-, two-, and three-parameter models of item-response theory to the ministry of national education secondary education institutions student selection and placement test data" **Yayınlanmamış Yüksek Lisans Tezi**, ODTÜ, Eğitim Bölümleri Bölümü, Ankara.
- Englehard, G., Jr., (1990). "Thorndike, Thurstone and Rasch: A comparison of their approaches to item-

- invariant measurement". American Educational Research Association'ın yıllık toplantısı. Boston.
- Ertkin, E. (1993) "Geleneksel ölçme kuramına alternatif iki yöntemin tanıtılması ve personel seçimine yönelik uygulama çalışması". **Yayınlanmamış Doktora Tezi**. İstanbul Üniversitesi, İşletme Fakültesi, İstanbul.
- Fan, X. (1998). "Item response theory and classical test theory : An emprical comparison of their item/person statistics". **Educational and Psychological Measurement** , 58: 357-381.
- Gulliksen, H. (1950). "**Theory of mental tests**". New York: John Wiley.
- Hambleton, R.K. & Swaminathan, H. (1985). "**Item response theory: Principles and applications**". Boston: Kluwer.
- Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). "**Fundamentals of item response theory**", Newbury Park, CA: Sage Publication.
- Hambleton, R. K., Zaal, J. N., & Pieters, J. M. P.(1991). "Computerized adaptive testing: Theory, applications and standards". In R. K. Hambleton & J. N. Zaal (Eds.), **Advances in educational and psychological testing: Theory and applications** (pp. 341-366). Boston: Kluwer.
- Hambleton, R. K., 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. (1995) "Meeting the measurement challenges of the 1990s and beyond: New assessment models and and methods". In T. Oakland and R. K. Hambleton (Eds.), **International Perspectives on Academic Assessment** (pp. 83-104). Boston: Kluwer.
- Kılıç, İ. (1999). "The fit of one-, two-, and three-parameter models of item-response theory to the student selection test of the student selection and placement center". **Yayınlanmamış Yüksek Lisans Tezi**., ODTÜ, Eğitim Bölümleri Bölümü, Ankara.
- Kolen, M. J. & Brennan, R. L. (1995). "**Test equating: Methods and practices**", New York: Springer.
- Lord, F. M. & Novick, M. R.(1968). "**Statistical theories of mental test scores**". Reading MA: Addison-Wesley.
- Lord, F. M. (1980). "**Applications of item response theory to practical testing problems**". Hillsdale, N.J.: Lawrence Erlbaum.
- Miller, M. D., Linn, R. L. (1988). "Invariance of item characteristic functions with variations in instructional coverage" **Journal of Educational Measurement**, 25: 205-219.
- Mislevy, R. J. , & Bock, R. D. (1986). PC-BILOG: "**Item analysis and test scoring with binary logistic models**". Scientific Software Inc.
- Skaggs, G. & Lissitz, R.L.(1986). "IRT test equating: Relevant issues and a review of recent research", **Review of Educational Research**, 56(4), 495-529.
- Wright, B. D. "Sample free test calibration and person measurement", **Proceedings of the 1967 Invitational Conference on Testing Problems**. Princeton: NJ: Educational Testing Service. (1968)
- Van der Linden, W. J. & Hambleton, R. K. (1997). "Item response theory: Brief history, common models and extensions". In van der Linden, W. J. & Hambleton, R. K. (Eds.), **Handbook of Modern Item Response Theory**. New York: Springer.
- Van der Linden, W. J. (1995). "Advances in computer applications". In T. Oakland and R. K. Hambleton (Eds.), **International Perspectives on Academic Assessment** (pp. 83-104). Boston: Kluwer Academic Publishers.
- Weiss, D. J. (1983). "**New horizons in testing**". New York: Academic Press.
- Yenal, E. (1995). "Differential item functioning analysis of the quantitative ability section of the first stage of the university entrance examination in Turkey". **Yayınlanmamış Yüksek Lisans Tezi**. ODTÜ, Eğitim Bilimleri Bölümü, Ankara.