

# Klasik Test Teorisine ve Örtük Özellikler Teorisine Göre Kestirilen Madde Parametrelerinin Karşılaştırılması Üzerine Ampirik Bir Çalışma

Adnan KAN\*

*Özet* – Bu araştırmada OKS (Orta Öğretim Kurumları Sınavı) Türkçe alt testine ait maddeler klasik test teorisi (KTT) ve örtük özellikler teorisine (ÖÖT) göre analiz edilmiştir. Araştırmanın amacı, KTT ve ÖÖT’ye göre kestirilen madde istatistikleri arasında nasıl bir ilişki olduğunu ortaya koymaktır. Araştırmada veri toplama aracı olarak OKS Türkçe alt testi kullanılmış ve araştırma bu sınavı alan 553108 öğrenci üzerinde yürütülmüştür. Toplanan verilerin öncelikle ÖÖT’nin varsayımlarını karşılayıp karşılamadığı test edildikten sonra, her iki teoriye göre madde istatistikleri ve madde parametreleri BILOG ve ITEMAN programları yardımıyla kestirilmiştir. Daha sonra her iki teoriye göre kestirilen madde parametreleri ve istatistikleri arasındaki ilişkiyi belirlemek için Spearman Brown sıra farkları korelasyon katsayısı kullanılmıştır. Araştırma sonucunda elde edilen bulgular göstermiştir ki; OKS Türkçe alt testinin belirtilen iki ölçme teorisine dayalı olarak kestirilen madde parametreleri ve istatistikleri büyük benzerlik göstermektedir.

*Anahtar kelimeler:* Klasik test teorisi, örtük özellikler test teorisi, madde parametreleri, madde istatistikleri.

*Abstract – An Empirical Study on the Comparison of Predicted Item Parameters with respect to Classical and Item Response Test Theories* – In this study, the item statistics of OKS were analyzed according to two different theories frameworks called classical test theory (CTT) and item response theory (IRT). The purpose of this study is to bring out what are the empirical relationships between CTT and IRT based item statistic. OKS Turkish subtest was used as a data collection instrument and study was implemented on 553108 students who took the exam. Firstly, the assumption of IRT were tested on collected data. Secondly, the item statistics was estimated through BILOG and ITEMAN computer programs. After all, in order to determine the relationship between item statistics and parameter that was estimated through two different theories, Spearman Brown correlation coefficient was used. Research results indicated that the item statistics derived from these two different theories frameworks are quite comparable.

*Key words:* Classical test theory, item response theory, item parameters, item statistics.

## Giriş

Eğitimde ve psikolojide bir çok yapı ya da özellik doğrudan gözlenemez. Bu nedenle bu tür değişkenleri ölçmek için iki alternatif ölçme teorisi geliştirilmiştir. Bunlar sırasıyla

---

\* Adnan Kan, Öğretim Üyesi, Mersin Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü. E-mail: <adnankan@mersin.edu.tr>.

gerçek puan teorisi olarak da bilinen klasik test teorisi (KTT) ve örtük özellikler teorisidir (ÖÖT). Eğitimde ve psikolojide test geliştirme aşamasında madde ve test istatistiklerini kestiren bu iki teoriden yararlanır. Eğitimle uğraşan çoğu araştırmacı KTT'ne daha aşinadır. KTT'nin en temel avantajı bir çok test durumuna uygulanmasını kolaylaştıran zayıf teorik varsayımlara sahip olmasıdır (Hambleton ve Jones, 1993). Testi uygulama ve parametreleri kestirme kolaylığı, daha az varsayım gerektirmesi ve üzerinde ÖÖT'ye göre daha çok çalışılmış olmasından dolayı test geliştirmede KTT'ye dayalı yöntemler daha çok tercih edilmektedir (Kelecioğlu, 2001). Bu sebeple, KTT, eğitimde kullanılan birçok testin geliştirilmesinde ve ÖÖT'nin oluşturulmasına ve yorumlanmasına temel teşkil etmiştir. Klasik test teorisinde iyi bir test geliştirmek için hareket noktası, madde güçlük indeksi (p) ve madde ayıricılık gücü indeksidir ( $r_{jx}$ ). Bu istatistikler yardımıyla teste ait özellikler kestirilebilir (Doğan ve Tezbaşaran, 2003; Osterlind, 1989).

Bir test geliştirilirken hangi teorinin daha avantajlı olacağı hala üzerinde tartışılan ve sınıranan bir konudur. Fakat KTT ve ÖÖT arasında bir takım farklılıklar vardır. KTT'ne göre maddeye, teste, veya bir bireye ait parametreler testi alan gruba bağlıdır. Bu sebeple herhangi bir referans olmaksızın sunulan madde karakteristikleri çok fazla bir bilgi sağlamaz. KTT'nin sınırlılığının odak noktası bir ya da birkaç grup için ölçmelerin birleştirilmesine ve yorumlanmasına temel teşkil edecek ölçüt bulunamamasıdır. KTT'nin aksine ÖÖT'nde odak noktası test değil, maddedir. ÖÖT dayandığı teorik matematiksel modelin, maddenin işleyişi (maddeyi doğru cevaplama olasılığı) ve altındaki yapı arasındaki ilişkiyi tanımlayabileceği savına sahiptir. ÖÖT'nin bu iki karakteristik özelliği bir arada ele alınırsa, bu ölçmelerin orijinal bir referans çerçevesi açısından sınırlandırılmaması anlamındadır. Bu ÖÖT'nin en büyük avantajı ve üstün yönlerinden birisidir. ÖÖT'ne göre, bir dizi madde üzerinden bir cevaplayıcı grubundan elde edilen bilgi, maddeleri tanımlayan, teorik modele ait parametreleri kestirmede kullanılır. Bununla beraber, bu değerler bir kez kestirildiğinde, orijinal referans çerçevesinin (orijinal madde seti ve orijinal cevaplayıcı grubu) sınırlayıcılığının kalktığı kabul edilir. Bir dizi madde başka bir cevaplayıcı grubuna verilebilir ve cevaplar her bir cevaplayıcıyı tek boyutlu bir ölçek üzerinde yerleştirilmek için kullanılabilir. Bu KTT'nde olduğu gibi madde karakteristiklerinin sadece orijinal grup veya orijinal madde seti için geçerli olmadığı anlamını taşır (Loyd, 1988).

ÖÖT'de madde ve test istatistiklerini kestirmek için bir çok model kullanılmaktadır. Bu modeller cevap kategorilerinin ve kestirilen madde parametresinin sayısına göre değişiklik gösterir. ÖÖT'ye ait modeller farklı matematiksel fonksiyonlardan yararlanırlar. Bazı modeller normal *ogive* fonksiyonunu bazıları ise lojistik fonksiyonları kullanırlar.  $\theta$  (*theta*) olarak adlandırılan yetenek düzeyi ve kestirilen madde parametrelerinin kombinasyonu (güçlük indeksi-b parametresi, ayırdedicilik gücü indeksi-a parametresi ve şans parametresi-c) madde karakteristik eğrilerinden (MKE) de görülebileceği gibi maddenin doğru cevaplandırılma olasılığını belirler

(Hambleton ve Swaminathan, 1985; Hambleton ve diğeri, 1991). Bir maddeyi doğru cevaplama olasılığı ile bireyin yetenek düzeyi arasındaki ilişkiyi ifade eden MKE gruptan gruba değişmez ve bu eğriyi belirleyen parametreler de sabit parametreler olarak kalır (Lord ve Novick, 1968). Ancak bu değişmezliğin sağlanması bazı şartlara bağlıdır. Bu şartlar ÖÖT'nin varsayımlarını oluşturur. Bunlar, tek boyutluluk, yerel bağımsızlık ve normallik varsayımlarıdır (Lord ve Novick, 1968; Hambleton ve Swaminathan, 1985). Her iki teoremin ortak varsayımı, test puanlarının normal dağılıma sahip olmasıdır. Normallik varsayımının karşılanması durumunda ÖÖT'deki  $a$  ve  $b$  parametreleri ile KTT'deki madde güçlük indeksi ( $p$ ) ve ( $r_{jx}$ ) arasında geçiş sağlanabilmektedir (Lord ve Novick, 1968; Crocker ve Algina, 1986).

Bu iki teoremin madde seçme yada test geliştirme prosedürleri göz önüne alındığında aynı sonuçları üretip üretmeyeceği merak konusu olmuştur. Bu çalışma ile bu ve benzeri sorulara cevap aranmış, çalışmanın bulgularına ve daha önceki literatüre dayalı yorumlamalar yapılmıştır.

## Yöntem

### *Verilerin Toplandığı Grup ve Ölçme Aracı*

Araştırmada veri toplama aracı olarak, Milli Eğitim Bakanlığı, Eğitim Teknolojileri Genel Müdürlüğü tarafından yapılan 2001 Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı (OKÖSYS) Türkçe alt testi kullanılmıştır. Araştırma bu testi alan 553.108 öğrenci üzerinde yürütülmüştür.

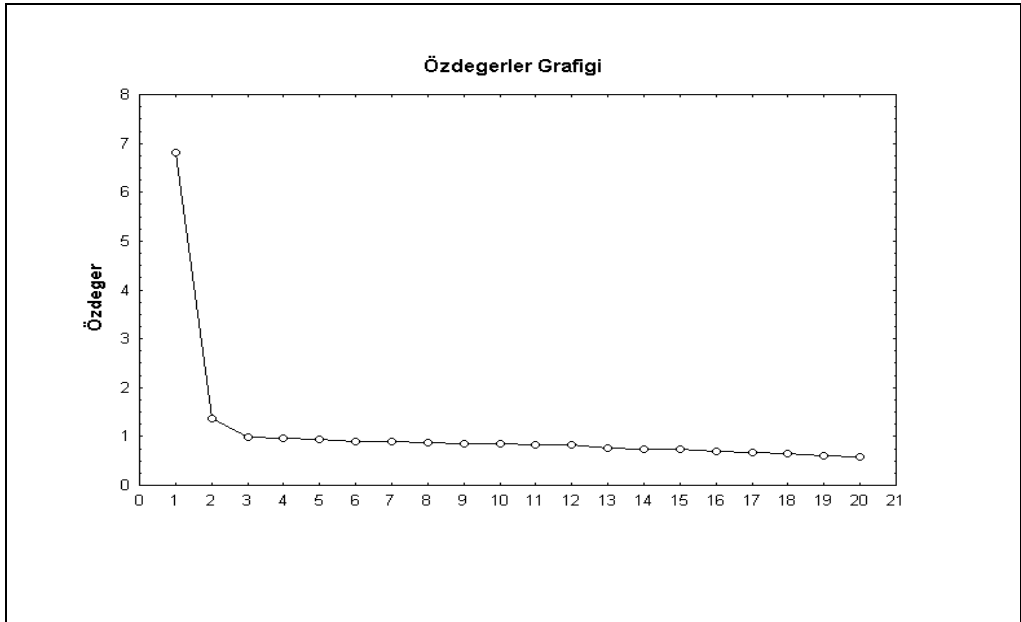
### *Verilerin Analizi*

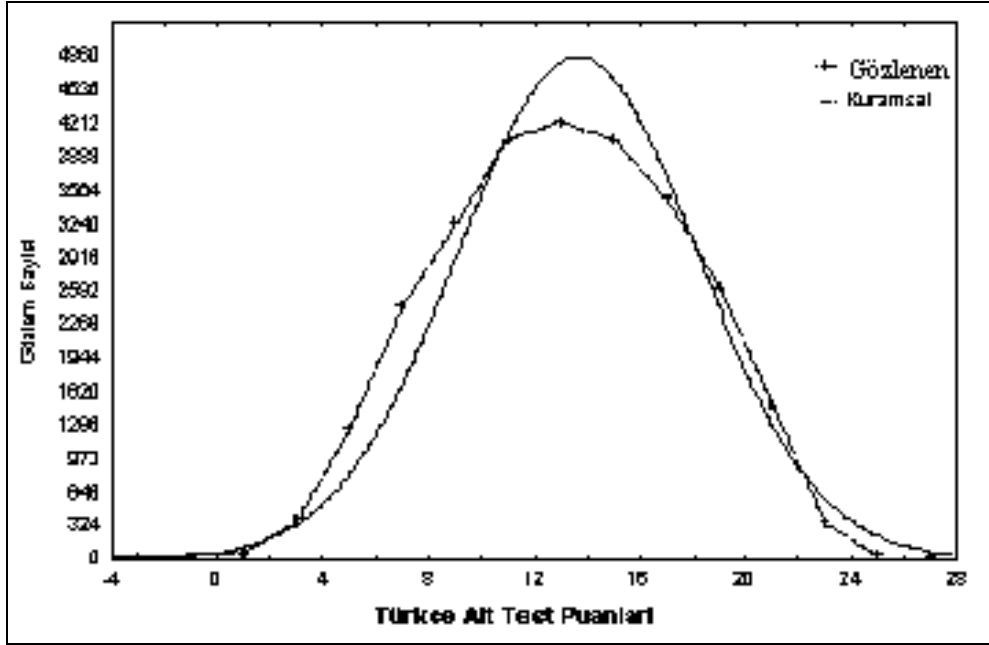
Verilerin analizi üç aşamada gerçekleştirilmiştir. Birinci aşamada, evrene ait betimsel istatistikler hesaplanmış ve Tablo 1'de verilmiştir. Tablo 1'de yer alan betimsel istatistikler incelendiğinde, Türkçe alt testinin başarı puanları dağılımı, standart normal dağılıma göre çok hafifçe sola çarpık ve basıktır. Çarpıklık değerinin yaklaşık 0 bulunmuş olması verilerin oluşturduğu dağılımın simetriğe çok yakın bir şekil aldığına göstergesi olarak kabul edilebilir. İkinci aşamada, evrenden elde edilen verilerin ÖÖT'nin temel varsayımlarını karşılayıp karşılamadığı test edilmiştir. Tek boyutluluk ve yerel bağımsızlık varsayımının karşılanıp karşılanmadığını belirlemek için temel bileşenler analizi yapılmış ve Türkçe puanlarına ilişkin özdeğer-bileşen grafiği Şekil 2'de verilmiştir. Grafiğe göre birinci faktör çok yüksek bir özdeğere sahip olmakla birlikte, bu faktörden sonra gelen ilk özdeğer birden bire düşmektedir. Bu testin dominant (baskın) tek faktöre sahip olduğunu göstermektedir. Testin tek boyutlu olması, ÖÖT varsayımlarından tek boyutluluğun yanında yerel bağımsızlık varsayımının da karşılandığını kabul etmek için yeterli kanıt olarak görülmektedir (Lord, 1980; Hambleton ve Swaminathan, 1985).

**Tablo 1:** Türkçe Alt Test Puanları Dağılım Bilgileri

Madde Sayısı	25
Yanıtlayıcı Sayısı	553.108
Ortalama	13,62
Varyans	20,79
Çarpıklık	-0,04
Basıklık	-0,67
Güvenirlilik	0,77
Ölçmenin Std. Hatası	2,17
Ortalama Güçlük	0,55

Betimsel istatistikler ve Şekil 2 incelendiğinde, standart normal dağılıma göre çok hafif sola çarpık ve basık olduğu gözlenebilir. Çarpıklık değerinin yaklaşık 0 bulunmuş olması sebebiyle evrendeki dağılımın ÖÖT'nin normallik varsayımını karşıladığı kabul edilebilir. Üçüncü aşamada ise ÖÖT'ye ve KTT'ye göre madde istatistikleri ve parametreleri BILOG ve ITEMAN paket programları aracılığıyla kestirilmiş ve daha sonra araştırmanın problemine cevap bulmak amacıyla ÖÖT'ye ve KTT'ye göre Madde istatistikleri ve parametreleri arasındaki ilişkiyi belirlemek için Spearman'ın sıra farkları korelasyonu hesaplanmıştır.

**Şekil 1:** Türkçe Alt Testi Özdeğerler Çizgi Grafiği



Şekil 2: Türkçe Alt Testi Sıklık Dağılım Grafiği

## Bulgular ve Yorum

Araştırmada öncelikle OKS Türkçe alt testine ilişkin madde parametreleri ve istatistikleri KTT'ye ve ÖÖT'ye göre belirlenmiş ve Tablo 2'de verilmiştir.

Tablo 2 incelendiğinde KTT'ye göre kestirilen OKS Türkçe alt testine ait madde güçlük indekslerinin 0,23 (madde 15) ile 0,92 (madde 2) arasında değerler aldığı ve ortancasının 0,55 olduğu, ayırıcılık gücü indekslerinin ise ortancası 0,40 olmak üzere, 0,11 (madde 16) ile 0,56 (madde 21) arasında değiştiği gözlenmiştir. Öte yandan ÖÖT'ye göre kestirilen a parametrelerinin ortancası 0,51 olmak üzere, 0,12 (madde 16) ile 0,87 (madde 2) arasında değiştiği, b parametrelerinin ise ortancası -0,23 olmak üzere -2,13 (madde 2) ile 2,56 (madde 15) arasında değiştiği gözlenmiştir. Tablo ve bulgular dikkatle incelendiğinde, KTT'ye ve ÖÖT'ye göre kestirilen madde parametrelerinin birlikte değişim gösterdiği gözlenebilir. Her iki teoriye göre en düşük ayırıcılık gücüne 16. madde sahipken, en kolay maddenin de her iki teoriye göre 2. madde olduğu gözlenebilir.

Ayrıca OKS Türkçe alt testinde yer alan maddelerin KTT ve ÖÖT'ye dayalı olarak kestirilen madde ayırıcılık gücü ve madde güçlük değerleri arasındaki ilişki Spearman'ın sıra farkları korelasyon katsayısı ile incelenmiş ve sonuçlar Tablo 3'te verilmiştir.

**Tablo 2:** OKS'ye İlişkin KTT ve ÖÖT'ye Ait Madde Parametre Değerleri

Madde	<i>p</i>	$r_{jx}$	<i>a</i>	<i>b</i>
1	0,45	0,35	0,41	0,33
2	0,92	0,49	0,87	-2,13
3	0,76	0,52	0,77	-1,17
4	0,26	0,21	0,25	2,56
5	0,33	0,40	0,49	0,98
6	0,47	0,29	0,33	0,21
7	0,69	0,55	0,83	-0,80
8	0,55	0,26	0,30	-0,45
9	0,65	0,33	0,40	-1,01
10	0,62	0,52	0,71	-0,53
11	0,39	0,41	0,51	0,62
12	0,38	0,35	0,41	0,75
13	0,76	0,49	0,70	-1,20
14	0,64	0,38	0,46	-0,82
15	0,23	0,26	0,31	2,51
16	0,46	0,11	0,12	0,87
17	0,42	0,54	0,75	0,32
18	0,55	0,50	0,66	-0,23
19	0,79	0,40	0,54	-1,67
20	0,54	0,36	0,43	-0,22
21	0,72	0,56	0,84	-0,93
22	0,42	0,48	0,63	0,35
23	0,78	0,52	0,80	-1,26
24	0,24	0,28	0,35	2,14
25	0,61	0,57	0,84	-0,46

**Tablo 3:** KTT ve ÖÖT'ye Dayalı Olarak Kestirilen Madde Ayırıcılık Gücü ve Madde Güçlük Değerleri Arasındaki İlişki

		ÖÖT	
		<i>a</i>	<i>b</i>
KTT	$r_{jx}$	0,961**	-
	<i>p</i>	-	0,981**

p&lt;0,01

Tablo 3'e göre, OKS Türkçe alt testinde yer alan maddelerin KTT ve ÖÖT'ye dayalı olarak kestirilen madde ayırıcılık güçleri arasındaki korelasyon katsayısı 0.961 iken madde güçlük indeksleri arasındaki korelasyon katsayısı ise 0.981 olup elde edilen bu katsayılar 0.01 düzeyinde manidardır. Bu bulgu elde edilen madde ayırıcılık gücü ve madde güçlük değerlerinin birbirleriyle tutarlı olduğu, farklı kuramlara göre

kestirilen madde ayırıcılık gücü ve madde güçlük değerleri bakımından kuramlar arasında büyük bir farklılığın olmadığı şeklinde yorumlanabilir. Fakat korelasyon katsayılarının büyüklüğü göz önüne alınırsa, her iki teoriye göre kestirilen madde güçlük indekslerinin madde ayırıcılık gücü değerlerine göre birbiriyle daha tutarlı olduğu söylenebilir.

## Sonuç ve Öneriler

Bu çalışmada, OKS Türkçe alt testine ilişkin madde parametreleri Klasik Test Teorisi (KTT) ve Örtük Özellikler Teorisine (ÖTT) göre analiz edilerek elde edilen madde parametreleri arasındaki ilişkiler test edilmiştir. Böylece bu iki kuramın madde seçme ve test geliştirme prosedürü açısından benzer sonuçlar üretilip üretilmediği belirlenmeye çalışılmıştır.

OKS Türkçe testinde yer alan maddelerin klasik test kuramı ve örtük özellikler kuramına göre kestirilen madde ayırıcılık güçleri ve güçlük indeksleri arasındaki ilişkiler *Spearman* sıra farkları korelasyon katsayısı ile incelenmiştir. Sonuç olarak her iki teoriye göre kestirilen madde güçlük indeksleri arasındaki korelasyon, 0,981 iken madde ayırıcılık gücü indeksleri arasındaki korelasyon ise 0,961 olarak bulunmuştur. Her iki kurama göre kestirilen madde ayırıcılık güçlerinin ve madde güçlük indekslerinin birbirine benzediği diğer bir deyişle, benzer sonuçlar ürettiği söylenebilir. Ancak madde ayırıcılık gücü indekslerine ilişkin korelasyonların madde güçlük indekslerine ilişkin korelasyonlardan görece daha düşük olduğu söylenebilir. Bu sonuçlar, Gelbal (1994), Erden (1997), Fan (1998), Stage (1998), Somer (1998), Nartgün (2002)'nin ulaştığı sonuçları desteklemektedir. Söz konusu çalışmalarda da test maddelerine ait klasik kurama ilişkin  $p$  ve  $r_{jx}$  istatistikleri ile örtük özellikler teorisine ilişkin  $a$  ve  $b$  parametreleri arasındaki ilişkiler incelenmiş ve oldukça yüksek korelasyonlar elde edilmiştir. Elde edilen sonuçlara dayanılarak KTT ve ÖÖT arasındaki kuramsal farklılıklara rağmen, madde istatistikleri ve madde parametreleri açısından uygulamada elde edilen sonuçların fazlasıyla benzerlik gösterdiği sonucuna ulaşılmıştır.

Bu çalışmanın sonuçlarına dayanarak, test geliştirme ve madde seçme sürecinde, testi oluşturacak maddelerin her iki teoriye göre de hemen hemen aynı olacağı başka bir deyişle, her iki teorinin de madde seçme ve test geliştirme sürecine ilişkin olarak benzer sonuçlar ürettiği söylenebilir. Bu sebeple teste madde seçme ve test geliştirme sürecine ilişkin koşullar sağlandıktan sonra testi oluşturacak maddelerin seçilmesi aşamasında eğitimciler yada uygulayıcılar iki teoriden herhangi birini kullanabilirler. Örtük özellikler teorisine göre test geliştirmek için uygulamanın klasik test teorisine göre daha büyük grup gerektirmesinden dolayı, puan dağılımına ilişkin koşullar sağlandıktan sonra klasik test teorisine göre test geliştirme daha pratik bir yol olarak önerilebilir. (Kelecioğlu, 2001). Bununla birlikte, klasik test teorisinde karşılığı

bulunmayan madde ve test bilgi fonksiyonları da madde seçme ve test geliştirmenin ÖÖT'ye göre yapıldığı durumlarda başvuru önemli referans noktalarıdır. Bu açıdan bakıldığında, ÖÖT'ye dayalı test geliştirecek araştırmacıların bu parametreleri de göz önüne almaları gerekmektedir. Bu çalışma madde seçme ve test geliştirmede önemli yer tutan madde istatistiklerinin iki teoriye göre karşılaştırılması üzerine yapılandırılmış olup, klasik test teorisinde bu parametrelerin karşılığı olmadığı için karşılaştırma yapılamamıştır. Bu durum araştırmanın sınırlılığını oluşturmaktadır.

### Kaynakça

- Crocker, L ve Algina, J. (1986). *Introduction to classical and modern test theory*. Florida: Harcourt Brace Jovanovich College Publishers.
- Doğan, N. ve Tezbaşaran, A.A. (2003). Klasik test kuramı ve örtük özellikler kuramının örneklem bağlamında karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 25, 58-67
- Erden, D. (1997). Örtük özellikler ve klasik test teorisine yaklaşımına dayalı olarak geliştirilen Likert tipi tutum ölçeğinin psikometrik özelliklerinin karşılaştırılması. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi, Ankara.
- Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item-person statistics. *Educational and Psychological Measurement*, 58(3), 357-381
- Gelbal, S. (1994). P madde güçlük indeksi ile rash modelinin bir parametresi ve bunlara dayalı yetenek ölçüleri üzerine bir karşılaştırma. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi, Ankara.
- 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*. Boston: Kluwer Academic Publishers.
- Hambleton, R.K., Swaminathan, H., ve Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, California: Sage Publications.
- Kelecioğlu, H. (2001). Örtük özellikler teorisindeki b ve a parametreleri ile klasik test teorisindeki p ve r istatistikleri arasındaki ilişki. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 20, 104-110.
- Lord, F.M. ve Novick, M.R. (1968). *Statistical theories of mental test scores*. New York: Addison-Wesley Publishing Company.
- Lord, F.M. (1980). *Application of item response theory to practical testing problem*. New Jersey: Lawrence Erlbaum Associates Publishers.
- Loyd, B.H. (1988). Implication of the item response theory for the measurement practitioner. *Applied Measurement in Education*, 2, 135-143.
- Nartgün, Z. (2002). Aynı tutumu ölçmeye yönelik Likert tipi ölçek ile metrik ölçeğin madde ve ölçek özelliklerinin klasik test kuramı ve örtük özellikler kuramına göre incelenmesi. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi, Ankara.
- Osterlind, S.J. (1989). *Constructing test items*. Boston: Kluwer Academic Publishers.



- Stage, C. (1998). Item analysis based on item response theory and on classical test theory: A comparison. The International Swe SAT Conference, San Diego.
- Somer, O. (1998). Kişilik testlerinde klasik ve modern test kuramları ile madde analizi. *Türk Psikoloji Dergisi*, 13(41), 1-15.