

Öğrenci Başarısının Öğretmen Yapımı Bir Testle Klasik Test Kuramı ve Madde Tepki Kuramı Yöntemleriyle Elde Edilen Puanlara Göre Karşılaştırılması

Ümit ÇELEN*

Eren Can AYBEK**

Eskişehir Osmangazi Üniversitesi

Özet

Bu çalışmanın amacı, sınıf içi ders başarısını ölçmeyi amaçlayan bir ölçme aracından, KTK ve MTK'ya dayalı olarak elde edilen öğrenci başarıları arasındaki ilişkiyi ve her iki kurama göre ölçme aracını oluşturan madde sayılarını karşılaştırmaktır. Amaç doğrultusunda betimsel türde bir araştırma yöntemi tercih edilmiştir. Eskişehir Osmangazi Üniversitesi'nde öğrenim gören ve ölçme ve değerlendirme dersini alan 231 öğrenci, araştırma grubunu oluşturmuştur. Araştırma kapsamında 40 çoktan seçmeli maddeden oluşan, ölçme ve değerlendirme dersindeki başarıyı ölçmeyi amaçlayan bir test geliştirilmiş ve uygulanmıştır. Test, klasik test kuramı ve madde tepki kuramı yöntemleriyle puanlanmış ve elde edilen puanların karşılaştırılması yapılmıştır. Analiz sonuçlarına göre, testteki tüm maddeler her iki kurama göre puanlandığında, elde edilen puanlar arasındaki ilişki pozitif, oldukça yüksek ve manidar bulunmuştur. Madde analizleri yapıldığında, klasik test kuramına göre beş, madde tepki kuramına göre ise 14 maddenin testten çıkarılması gerektiği görülmüş ve madde atılarak puanlama yeniden yapıldığında, elde edilen puanlar arasındaki ilişki pozitif, oldukça yüksek ve manidar bulunmuştur. Test puanları ile geçme ve kalma kararı verildiği düşünüldüğünde ise iki kuram 231 öğrencinin sadece %12.6'sında kalma kararında farklılaşmış, %87.4'ünde ise aynı kararı vermiştir.

Anahtar Sözcükler: *madde tepki kuramı, klasik test kuramı, öğretmen yapımı test*

Abstract

The aim of this research is examine the correlation between student achievements and the items that construct the test according to the Classical Test Theory and Item Response Theory. The survey method has been chosen for this research. Study group includes 231 students which studying in Eskişehir Osmangazi University and attending to Measurement and Evaluation class. A 40 multiple-choice items test has been developed and administered. The test aims to measure the academic achievement on measurement and evaluation lesson. Scoring has been made according to Classical Test Theory and Item Response Theory, and then scores have been compared. Analyses show that the correlation between the scores is positive, very high and significant. After the item analyses, 5 items have been opt out of the test by classical test theory. And 14 items have been opt out of the test by item response theory. After opting out, scoring has been made again and the correlation between the scores is positive, very high and significant again. When we consider the test scores to pass – fail decision; two theories only differ 12.6% of students and the same decision has been made with 87.4% of students.

Keywords: *item response theory, classical test theory, teacher made tests*

Öğretmen, öğrencilerinin dersle kazandırılmak istenen hedeflere (kazanımlara) ulaşip ulaşmadıklarını belirlemek amacıyla çeşitli ölçme yöntemlerine başvurur. Sınavlar bu amaçla

* Yard. Doç. Dr., Eskişehir Osmangazi Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Ölçme ve Değerlendirme Anabilim Dalı.

** Araştırma Görevlisi, Eskişehir Osmangazi Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Ölçme ve Değerlendirme Anabilim Dalı.

*** Bu çalışma III. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde bildiri olarak sunulmuştur.

kullanılan ölçme yöntemlerinden biridir. Öğretmenin bu amaca ulaşabilmesi için yaptığı sınavlarla geçerli puanlar elde edebilmesi gerekmektedir. Ölçme sonuçlarının geçerli olabilmesi için de ölçülmek istenen değişken olan ders başarısının, başka bir değişkenle karıştırılmadan tam ve doğru olarak ölçülmesi gereklidir. Ölçülen bireyden, ölçmeyi yapan kişiden, ölçülen özellikten, kullanılan ölçme aracından vb. kaynaklanan çeşitli hatalar yüzünden öğrenci başarısının gerçek değerine ulaşmak çoğu zaman mümkün olmamaktadır. Bu hataların ortaya çıkmasındaki en önemli neden psikolojik ölçmelerde gözlemin doğrudan değil, dolaylı olarak yapılmasıdır. Bir başka ifadeyle öğrencinin testte yer alan maddelere verdiği yanıtlara bakılarak gerçek başarısının kestirilmeye çalışılması bir miktar hatayı hep göze almayı gerektirmektedir (Erkuş, 2003).

Ölçme tarihi incelendiğinde testlerin *Klasik Test Kuramı* (KTK) ve *Madde Tepki Kuramı* (MTK) olarak adlandırılan iki temel test kuramı bağlamında geliştirildiği görülmektedir. Psikolojik ölçme tarihinin başlangıcından itibaren test geliştirmede, analizde ve psikolojik ölçeklerin puanlanmasında daha yaygın olarak kullanılan kuram KTK'dır. *Örtük Özellikler Kuramı* olarak da adlandırılan MTK ise KTK'nın bazı sınırlılıklarına alternatif olarak XX. yüzyılın ortalarında ortaya çıkmıştır (Crocker ve Algina, 1986). Günümüzde de KTK yaygın olarak kullanılmakla birlikte, MTK giderek daha popüler ve tercih edilir olmaya başlamıştır (Hambleton, 1994; Reise, Ainsworth ve Haviland, 2005).

Başarının ölçülmesinde KTK'ya göre puan, öğrencinin maddelerden almış olduğu puanın toplanmasıyla bulunmaktadır. Test ve madde istatistikleri de bu şekilde elde edilmiş toplam puanlar ve madde puanları üzerinden yapılmaktadır. Dolayısıyla KTK'ya göre hesaplanan puanlara göre bireylerin aldıkları puanlar o testin güçlük düzeyine göre değişmektedir (Lord ve Novick, 1968). Maddelerin güçlük düzeyleri ise genellikle dikkate alınmamakta ve -maddeler ağırlıklandırılmış olarak puanlanmıyorsa- her maddenin toplam puana katkısı aynı miktardadır. MTK'da ise test maddelerine verilen yanıtlar ile testin

ölçtüğü, doğrudan gözlenemeyen özellik arasında var olduğu düşünülen ilişkiyi matematiksel olarak ifade etmeye çalışan modeller söz konusudur. Varsayımları sağlandığında, bu kuramla uygulanan testten bağımsız olarak kestirimler (madde örnekleminde bağımsız yetenek kestirimleri, yanıtlayıcı örneklemlerinden bağımsız madde kestirimleri) ve değişmez madde parametreleri kestirimi yapılabildiği ifade edilmektedir. (Hambleton ve Swaminathan, 1985; Hambleton, 1990; Hambleton, Swaminathan ve Rogers, 1991; Kelecioğlu, 2001). Ayrıca MTK'ya göre hesaplanan puanlara her bir maddenin katkısı aynı değildir; kullanılan modele göre maddenin güçlüğünden, ayırt edicilik gücünden ve şansla doğru yanıtlanma olasılığından etkilenen madde puanları söz konusudur.

İki kuramın ortak yönleri vardır: Her iki kuramda da test maddelerine verilen tepkilerin doğruluğu veya yanlışlığı üzerine odaklanılmaktadır. Normallik varsayımı her iki kuram için de söz konusudur. DeMars (2010), normallik varsayımının MTK modelleri için gerekli olmadığını belirtse de, normal dağılım sağlanamamasının yetenek kestiriminde sorunlara yol açabileceğini de ifade etmiştir. Bunun yanında, Linden ve Hambleton (1997), modern MTK'nın normallik varsayımını gerektiren faktör analizi tekniğinden çokça etkilendiğini belirtmiştir. Bununla birlikte test puanlarının normalliği sağlandığında, iki kurama göre elde edilen ayırt edicilik güçleri ve madde güçlük indeksleri arasında geçiş sağlamak mümkün olabilmektedir (Lord ve Novick, 1968; Crocker ve Algina, 1986). Belirli bir özelliği ölçmek için geliştirilmiş testin sadece o özelliği ölçmesi ve testteki bir maddeye verilen yanıtın diğerini etkilememesi gerekliliği de her iki kuram için karşılanması gereken varsayımlardır (Demirtaşlı, 1998; Linn, 1998).

Yukarıda bahsedilen iki test kuramından birinin diğerine üstün olup olmadığı hakkında yapılan birçok çalışma vardır. Türkiye'de yapılan çalışmalar incelendiğinde bu çalışmaların genellikle geniş ölçekli bir sınav verisi kullanılarak o sınavın MTK modellerine uyumunun incelendiği veya bu veri seti üzerinden iki kurama göre elde edilen parametrelerin

karşılaştırıldığı çalışmalar (Kılıç, 1999; Yalçın, 1999; Çalışkan, 2000; Çelik, 2001; Karataş, 2001; Özkurt, 2002; Can, 2003; Yapar, 2003; Yeğin, 2003) olduğu görülmektedir. Yurtdışında da benzer olarak testlerin iki kurama göre psikometrik özelliklerinin karşılaştırıldığı çalışmalar bulunmaktadır (Lawson, 1991; Stage, 1998, 2003). Yurtiçi ve yurtdışındaki bu çalışmalarda öğretmen yapımı bir testin KTK ve MTK'ya göre puanlamasının testin psikometrik özelliklerinde ve öğrenci performanslarında ne gibi bir farklılığa yol açabileceğine yönelik çalışmaların oldukça sınırlı olması bu araştırmayı gerekli kılmaktadır.

Çalışmanın amacı, sınıf içi ders başarısını ölçmeyi amaçlayan bir ölçme aracından, KTK'ya dayalı olarak hesaplanan ve MTK'ya dayalı olarak kestirilen öğrenci başarılarını incelemektir. Bu genel amaç doğrultusunda aşağıdaki alt amaçlara yanıt aranmıştır:

1. KTK'ya göre hesaplanan ve MTK'ya dayalı olarak kestirilen öğrenci başarıları arasında manidar bir ilişki var mıdır?
2. KTK ve MTK'ya dayalı olarak yapılan madde analizleri sonucunda, düzeltilmeden tekrar kullanılmaması gereken madde sayıları kaçtır?
3. Madde analizleri sonucunda testte kalan maddeler üzerinden KTK ve MTK'ya dayalı olarak elde edilen öğrenci başarıları arasında manidar bir ilişki var mıdır?
4. Test puanlarına dayalı olarak geçme – kalma kararı verilirse, KTK ve MTK'ya göre geçen ve kalan öğrenci sayıları kaçtır?

Yöntem

Bu çalışma, bir testin KTK ve MTK'ya dayalı olarak puanlanmasına bağlı olarak öğrencilerin sıralamalarını ve bir değerlendirme yapıldığında, değerlendirme sonucunun nasıl değiştiğini incelemeyi amaçlayan betimsel türde bir araştırmadır.

Çalışma Grubu

Çalışma için uygun örnekleme yapılmıştır. Eskişehir Osmangazi Üniversitesi Eğitim Fakültesi'nde öğrenim gören ve 2011-2012 Öğretim Yılı Bahar Döneminde Ölçme ve Değerlendirme dersini alan, İlköğretim Fen Bilgisi Öğretmenliği (85 öğrenci), İlköğretim Matematik Öğretmenliği (98 öğrenci) ve Bilgisayar ve Öğretim Teknolojileri (48 öğrenci) programlarında öğrenim gören 231 öğrenci ile çalışma grubu oluşturulmuştur.

Veri Toplama Araçları

Veriler, çoktan seçmeli 40 maddeden oluşan bir başarı testi ile toplanmıştır. Maddeler, lisans düzeyinde yürütülmekte olan Ölçme ve Değerlendirme dersinin amaçları temel alınarak hazırlanmıştır. Dönem sonu sınavı olarak uygulanan testte dersin öğretim üyesi tarafından tüm kapsamın örneklenmesine dikkat edilmiştir. Testin uygulanma süresi 60 dakikadır. Test puanlarına ait iç tutarlılığın araştırılması için KR-20 güvenirlik katsayısı hesaplanmış ve bu katsayı .90 olarak elde edilmiştir. BILOG-MG 3 tarafından, MTK'ya dayalı Bayesian yöntemi ile hesaplanan görgül güvenirlik (empirical reliability) katsayısı ise .80 olarak bulunmuştur (BILOG MG 3 için kullanılan syntax ek olarak sunulmuştur).

Verilerin Toplanması

Uygulama tek oturumda, birden çok sınıfta yapılmış ve her sınıfta bir gözetmen bulunmuştur. Testte yer alan maddelerin sıraları değiştirilerek A ve B formu oluşturulmuş ve uygulama bu şekilde gerçekleşmiştir. Madde sıraları değiştirilirken konu bütünlüğünün sağlanmasına dikkat edilmiştir.

Verilerin Analizi

Araştırma kapsamında KTK ve MTK'ya dayalı analizler yapılmıştır. Birinci alt amaç için, testte yer alan 40 madde üzerinden KTK'ya dayalı olarak öğrenci başarıları hesaplanmış ve MTK'ya dayalı olarak öğrenci yetenekleri kestirilmiştir. KTK için, öğrencilerin teste

verdikleri doğru yanıt sayıları toplanarak öğrencilerin ham puanı elde edilmiş, daha sonra bu puanlar T puanına dönüştürülmüştür. MTK içinse, her öğrenci için yetenek kestirimleri dikkate alınarak, bu yetenek kestirimleri T puanına dönüştürülmüştür (araştırmanın dördüncü alt amacı için MTK'dan elde edilen yetenek kestirimlerinin T puanına dönüştürülmesi gerektiğinden, araştırmanın bütünlüğü gözetilerek tüm alt amaçlar için yetenek kestirimlerinin T puanına dönüştürülmesi uygun görülmüştür). Her iki kuramdan elde edilen yetenek ölçüleri T puanlarına dönüştürüldüğünde yetenek ölçüleri arasındaki ilişki Spearman Sıra Farkları Korelasyon Katsayısı ile incelenmiştir. İkinci alt amaç için, KTK'ya dayalı madde ayırt edicilik indeksleri, nokta çift serili korelasyon katsayısı olarak hesaplanmıştır. Ayırt edicilik indeksi .30'un altında kalan maddeler, öğrenci puanları hesaplanırken test dışında tutulmuştur. MTK'ya dayalı olarak maddeleri incelemek içinse, öncelikle tek-boyutluluk varsayımının karşılanması amacıyla tetra-korik korelasyon matrisine dayalı faktör analizi yapılmıştır. Faktör analizi sonuçlarına göre tek-boyutluluğu sağlamayan maddeler *uyumsuz* olarak kabul edilerek puanlamadan çıkarılmıştır. Daha sonra kalan maddeler 1 parametrelili ve 2 parametrelili lojistik modele göre daha iyi uyum sağlaması nedeniyle 3 parametrelili lojistik modele (3PL) göre analiz edilmiştir. Madde - model uyumu incelenmiş ve madde - model uyumunu sağlamayan ($\chi^2 < .05$) maddeler analizden çıkarılmıştır. Üçüncü alt amaca yönelik olarak, öğrenci başarıları hem KTK hem de MTK'ya dayalı olarak elde edilmiştir. Bunun için KTK'da öğrenci puanları önce Z sonra T puanına; MTK'da ise kestirilen yetenekler T puanına dönüştürülmüştür. Ardından iki yöntemle elde edilen T puanları arasındaki ilişki Spearman Sıra Farkları Korelasyon Katsayısı ile elde edilmiştir. Dördüncü alt amaç için T puanları dikkate alınmış, ölçüt olarak 50 puan belirlenmiş ve 50 puan ve üzerindeki öğrencilerin dersi geçtiği; 50 puanın altındaki öğrencilerin ise kaldığı düşünülerek, her iki kurama göre geçen ve kalan öğrenci sayıları

hesaplanmıştır. Buna göre bir kurama göre geçip, diğer kurama göre kalan öğrenci sayıları çapraz tablo ile belirlenmiştir.

Bulgular

Birinci alt amaca yönelik olarak öğrencilerin testte yer alan 40 maddeye verdikleri yanıtlar doğrultusunda KTK ve MTK'ya dayalı olarak elde edilen T puanları arasındaki ilişki Spearman Sıra Farkları Korelasyon Katsayısı ile incelenmiş ve sonuçlar Tablo 1'de verilmiştir.

Tablo 1. KTK ve MTK'ya Göre Belirlenen Öğrenci Başarı Puanları Arasında Hesaplanan Spearman Sıra Farkları Korelasyon Katsayısı

	KTK	MTK
KTK	1.000	.951**
MTK	.951**	1.000

**p < .01

Yukarıda yer alan tablodan da görüleceği gibi, test KTK ya da MTK'ya göre puanlandığında iki puan arasındaki ilişki pozitif yönde, yüksek ve manidardır.

İkinci alt amaca yönelik olarak KTK'ya dayalı olarak madde analizleri; MTK'ya dayalı olarak ise madde-model uyumu incelenmiştir. Buna göre; $r_{jx} = .30$ 'un altında yer alan madde sayıları ve MTK'ya dayalı olarak yapılan analizler için tek-boyutluluğu sağlamayan maddeler ile madde-model veri uyumu sağlamayan madde sayıları belirlenmiştir.

KTK'ya dayalı olarak yapılan madde analizi sonucunda $r_{jx} < .30$ olan yalnızca 5 madde (7, 21, 28, 33, 37) bulunmuştur. MTK'nın bir varsayımı olan tek-boyutluluğun sağlanması için yapılan faktör analizi sonucunda 13 (1, 7, 12, 13, 15, 16, 20, 25, 28, 31, 32, 33, 37) maddenin test dışına çıkarılması gerektiği görülmüştür. Madde-model uyumu test edildiğinde ise madde-model uyumu sağlamayan ($\chi^2 < .05$), bir maddenin (24) test dışına çıkarılması gerektiği bulunmuştur. Buna göre, MTK'ya dayalı olarak yapılan analizler için toplam 14 maddenin test kapsamı dışına çıkarılması gerektiği gözlemlenmiştir.

Üçüncü alt amaca yönelik olarak, öncelikle kalan maddeler üzerinden (KTK, 35; MTK, 26) KTK'ya göre öğrenci başarıları belirlenmiş ve MTK'ya göre yetenek kestirimleri yapılmıştır. Daha sonra bu puan ve kestirim değerleri T puanına dönüştürülerek KTK ile MTK'dan elde edilen puanlar arasındaki ilişki Spearman Sıra Farkları Korelasyon Katsayısı ile incelenmiştir. Analiz sonuçları Tablo 2'de yer almaktadır:

Tablo 2. KTK ve MTK'ya Göre Belirlenen Öğrenci Başarı Puanları Arasında Hesaplanan Spearman Sıra Farkları Korelasyon Katsayısı

	KTK	MTK
KTK	1.000	.953**
MTK	.953**	1.000

**p < .01

Analiz sonucu göstermektedir ki, maddelerin çıkartılmasının ardından her ne kadar iki test maddeleri farklı testler haline dönüştüyse de, KTK ile 35 madde üzerinden elde edilen öğrenci başarı puanları ile, MTK ile 26 madde üzerinden kestirilen öğrenci başarı puanları arasında .95 düzeyinde, yüksek manidar bir ilişki mevcuttur. Başka bir ifadeyle, testin KTK ya da MTK'ya göre geliştirilmiş olması öğrenci başarı sıralamalarında manidar bir değişikliğe yol açmamaktadır.

Dördüncü alt amaca yönelik olarak, T = 50 puan ölçütü alınarak bir değerlendirme yapıldığında, KTK ve MTK'ya göre geçen ve kalan öğrenci sayıları çapraz tablo olarak Tablo 3'te verilmiştir.

Tablo 3. Ölçüt Olarak T = 50 Puan Alındığında KTK ve MTK'ya Göre Değerlendirme Sonuçları

	MTK Geçti	MTK Kaldı	Toplam
KTK Geçti	107	24	131
KTK Kaldı	5	95	100
Toplam	112	119	231

Tablo 3'te görüldüğü üzere, puanlama MTK'ya göre yapıldığında; KTK'ya göre geçmesi gereken 24 öğrenci için kaldı kararı verilmektedir. KTK'ya göre yapılan puanlamada ise, MTK'ya göre geçti kararı verilen 5 öğrenci için kaldı kararı verilmiştir.

Tartışma ve Yorum

Sınıf içi ders başarısını ölçmeye yönelik olarak hazırlanan bir test KTK ve MTK'ya göre puanlandığında öğrenci sıralamaları arasındaki ilişki olduğu görülmüştür. Bu bulgu, iki ayrı kurama göre hesaplanan puanların ve yetenek ölçülerinin bireyleri benzer şekilde sıraya koyduğunu, biri yerine diğeri kullanılması durumunda sonuçlarda çok değişiklik meydana gelmeyeceğini göstermektedir. Bulunan katsayı, yapılan birçok çalışmada (Başarır-Erden, 1997; Baykul, 1979; Fan, 1998; Hwang, 2002; Nartgün, 2002; Çelen 2008) KTK'ya dayalı olarak hesaplanan puanlar ile MTK'ya dayalı olarak yapılan yetenek kestirimi korelasyonları ile örtüşmektedir.

Yapılan uygulama bir deneme uygulaması kabul edilip KTK ve MTK ölçütlerine göre testten çıkarılması gereken maddeler incelendiğinde, KTK'ya göre beş; MTK'ya göre ise 14 maddenin testten çıkartılması gerektiği görülmüştür. KTK'ya göre testten çıkarılması gereken maddelerin 21. madde hariç tümü aynı zamanda MTK'ya göre de testten çıkarılması gereken maddelerdir. Bu durum bir kurama göre testten çıkarılması gereken bir maddenin diğerkurama göre testte yer alması ihtimalinin düşük olduğu görüşünü desteklemektedir (Çelen, 2008). Sonuçlar Berberoğlu (1989)'nun çalışmasıyla ve Stage (1998)'in aynı grup üzerinde KTK ve MTK'ya göre elde ettiği madde istatistikleri arasında yüksek ilişki bulunduğu çalışmasıyla paralellik göstermektedir. Lawson (1991) da 1PL ile KTK karşılaştırmasında, iki kuramdan elde edilen sonuçlar arasında büyük benzerlikler bulunduğunu ifade etmiştir. Stage (2003), SweSAT testi için yaptığı benzer bir çalışmada ise KTK'ya göre yapılan puanlamanın daha iyi sonuçlar verdiğini belirtse de, bu durumun test verisi ile model uyumunun sağlanamamasından kaynaklı olabileceğini not etmiştir.

Maddeler testten çıkartılıp kalan maddeler üzerinden test puanları hesaplandığında, her iki kurama göre elde edilen puanlar arasındaki ilişki hala kuvvetlidir. Belirli bir kesme puanıyla öğrencilerin bu test sonuçlarına göre geçme kalma kararının verilmesi durumunda, MTK ile elde edilen puanlar sonucunda KTK 'ya göre geçmesi gereken 24 öğrencinin kaldığı; KTK'ya göre puanlama yapıldığında ise, MTK'ya göre geçmesi gereken beş öğrencinin kaldığı bulunmuştur. Bir başka ifadeyle iki kuram 231 öğrencinin sadece %12.6'sında kalma kararında farklılaşmış, %87.4'ünde ise aynı kararı vermiştir. Geçme ve kalma kararlarındaki bu farklılığa test maddelerinin farklılaşmasına bağlı olarak test kapsamlarının da farklılaşmasının yol açmış olabileceği unutulmamalıdır.

Bu sonuçlar doğrultusunda, MTK ve KTK'nın tercih edilmesi gerektiğinde, dersi yürüten öğretmenin kullanışlılık doğrultusunda karar vermesi doğru olacağını söylemek mümkündür. Ders kapsamında, bireye uyarlanmış test kullanımı, test eşitleme vb. amaçlanıyorsa MTK; geçti – kaldı kararı verileceğinde ise KTK'nın tercih edilmesi daha kullanışlı olacaktır.

Kaynaklar

- Başarır-Erden, D.B. (1997). *Örtük Özellikler ve Klasik Test Teorisi Yaklaşımına Dayalı Olarak Geliştirilen Likert Tipi Tutum Ölçeğinin Psikometrik Özelliklerinin Karşılaştırılması*. Yayımlanmamış doktora tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Baykul, Y. (1979). *Örtük Özellikler ve Klasik Test Kuramları Üzerine Bir Araştırma*. Yayımlanmamış doktora tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Berberoğlu, G. (1989). Eriş Testlerine Madde Seçiminde Klasik Test Kuramı ve Rasch Modelinin Karşılaştırılması. *Eğitim ve Bilim*, Cilt:13 S:74, 61-67.
- Can, S. (2003). *Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı sözel bölümünün Madde Tepki Kuramı modellerine göre analizi*. Yayımlanmamış doktora tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Crocker, L., Algina, J. (1986). *Introduction Classical and Modern Test Theory*. USA: CBS College Publishing Company.
- Çalışkan, M. (2000). *Madde Tepki Kuramı'nın (MTK) bir, iki ve üç parametrelili modellerinin Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Dairesi'nin (MEB-EARGED) fen bilgisi başarı testi verilerine uygunluğu*. Yayımlanmamış doktora tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Çelen, Ü. (2008). Klasik Test Kuramı ve Madde Tepki Kuramı Yöntemleriyle Geliştirilen İki Testin Geçerlilik ve Güvenilirliğinin Karşılaştırılması, *İlköğretim Online*, 7(3), 758-768.
- Çelik, D. (2001). *Madde Tepki Kuramının (MTK) bir-, iki-, ve üç parametrelili modellerinin Milli Eğitim Bakanlığı ortaöğretim kurumları öğrenci seçme ve yerleştirme sınavı testi verilerine uygunluğu*. Yayımlanmamış doktora tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Çıkrıkçı-Demirtaşlı N. (1998). Test Geliştirmede Yeni Yaklaşımlar: Örtük Özellikler Kuramı - Temel Özellikleri, Varsayımları, Model ve Sınırlılıkları. *Ankara Üniversitesi Eğitim Fakültesi Dergisi*, 2 (28), 161-173.

- DeMars, C. (2010). *Item Response Theory*. Oxford University Press, Oxford.
- Erkuş, A. (2003). *Psikometri Üzerine Yazılar*. Türk Psikologlar Derneği Yayınları, Ankara.
- Fan, X. (1998). Item response theory and classical theory: an empirical comparison of their item-person statistic. *Educational and Psychological Measurement*, v:58 n:3, 357-381.
- Hambelton, R.K. (1990). Item response theory: introduction and bibliography. *Psicothema*, 2, 1, 97-107.
- Hambelton, R.K. (1994). Item Response theory: a broad psychometric framework for measurement advances. *Psicothema*, 6, 3, 535-556.
- Hambelton, R.K., Swaminathan H., and Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. California: Sage Publications Inc.
- Hambleton, R.K., Swaminathan H. (1985). *Item Response Theory: Principles and Application*. Kluwer, Nijhoff Publishing a Member of the Kluwer Academic Publisher Group.
- Hwang, D.Y. (2002). Classical Test Theory and Item Response Theory: Analytical and Empirical Comparison. Speeches/meeting paper, presented at the Annual Meeting of the Southwest Educational Research Association (Austin, Feb. 14-16, 2002).
- Karataş, A. G. (2001). *Madde Tepki Kuramı (MTK) modellerini kullanarak bir İngilizce yeterlilik sınavının ölçeklendirilmesi*. Yayınlanmamış doktora tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Kelecioğlu, H. (2001). Örtük Özellikler Teorisindeki b ve a Parametreleri ile Klasik Test Teorisindeki p ve r İstatistikleri Arasındaki İlişki. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 20: 104-110.
- Kılıç, İ. (1999). *Madde tepki kuramının (MTK) bir, iki ve üç parametrelili modellerinin Öğrenci Seçme ve Yerleştirme Merkezi'nin (ÖSYM) Öğrenci Seçme Sınavına (ÖSS) uygunluğu*. Yayınlanmamış doktora tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Lawson, S. (1991). One-Parameter Latent Trait Measurement: Do the Results justify the Effort? *The Annual Series of the Southwest Educational Research Association*, Vol. 1, 159-168.
- Linden, W. J. Van Der, & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. Springer, New York.
- Linn, R.L. (1998). *Educational Measurement*. New York: Macmillan Publishers.
- Lord, F.M., Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. Addison Wesley Publishing Company, Educational Testing Service.
- 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 İncelenmesi*. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Özkurt, S. (2002). *Madde Tepki Kuramı'nın (MTK) bir-, iki-, ve üç- parametrelili modellerinin bir İngilizce yeterlilik başarı testi verilerine uygunluğu*. Yayınlanmamış doktora tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Reise, S.P., Ainsworth, A.T., and Haviland, M.G. (2005). Item Response theory. Fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science*, 14, 2, 95-101.
- Stage, C. (1998). A comparison between item analysis based on Item Response Theory and Classical Test theory. A study of the SweSAT Subtest WORD. *Educational Measurement*, 29.
- Stage, C. (2003). *Classical test theory or item response theory: The Swedish experience* (Vol. 42). Univ.
- Yalçın, M. (1999). *Eğitimi araştırma ve geliştirme dairesi başarı testlerinin madde-tepki kuramının bir, iki, üç parametreye uygunluğu*. Yayınlanmamış doktora tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Yeğin, O. P. (2003). *Başkent Üniversitesi İngilizce yeterlik sınavının (Büiys) madde Madde Tepki Kuramı'nın (MTK) üç parametrelili modelinin kullanımıyla elde edilen yetenek kestirimlerinin yordama geçerliği*. Yayınlanmamış doktora tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Yapar, T. (2003). *İki parametrelili tepki kuramı (MTK) modelinin yetenek kestirimleriyle Başkent Üniversitesi İngilizce yeterlik sınavının yordama geçerliğini inceleme çalışması*. Yayınlanmamış doktora tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.

EK-1: BILOG MG3 için Syntax

```
>GLOBAL DFNAME='olcme.dat', NPARM=3, SAVE;  
>SAVE PARM='olcme.par', SCORE='olcme.sco';  
>LENGTH NITEMS=40;  
>INPUT NTOTAL=40, NALT=5, NIDCHAR=3;  
>ITEMS INAMES=(OLC01(1)OLC40);  
>TEST1 TNAME='PRETEST', INUMBER=(1(1)40);  
(3A1,1X,40A1)  
>CALIB NQPT=31, CYCLES=25, NEWTON=10, CRIT=0.001, ACCEL=0.0, CHI=15,  
PLOT=1;  
>SCORE RSCTYPE=4, INFO=2, POP;
```