

# **Farklı Madde Puanlama Yöntemlerinin ve Farklı Test Puanlama Yöntemlerinin Karşılaştırılması**

Halil YURDUGÜL\*  
Hacettepe Üniversitesi

## **Özet**

Bu araştırmada klasik test kuramına dayalı ağırlıksız, ağırlıklandırılmış madde puanlama yöntemi, madde yanıt kuramına dayalı bir ve iki parametrelî modelden kestirilen olasılıksal madde puanlama yöntemi ve konjenerik test kuramına dayalı madde puanlama yöntemleri karşılaştırılmıştır. Karşılaştırmalar, veri kümesinin güvenilirliği, tekboyutluluk ve veri-model uyumu açısından ele alınmıştır. Elde edilen sonuçlara göre madde-yanıt kuramına dayalı puanlama yöntemi ile konjenerik test modeline dayalı puanlama yöntemlerinin klasik test kuramına dayalı madde puanlama yöntemlerine göre daha iyi sonuçlar verdiği gözlenmiştir. Diğer taraftan madde puanlarının toplamı şeklindeki test puanlarına göre sıralama karşılaştırması yapılmış ve en uyumsuz yöntemin şans başarısından arındırılmış test puanlama yöntemi olduğu görülmüştür.

**Anahtar Sözcükler:** Test puanlama yöntemleri, madde puanlama yöntemleri, klasik test kuramı, madde yanıt kuramı, konjenerik test kuramı

## **Abstract**

In this study, it was compared different item scoring methods which consist of unweighted item scoring method and weighted item scoring method, based on classical testing theory and probability item scoring methods, based on item response theory, and congeneric item scoring method based on congeneric test theory. It was investigated the effects of item scoring methods on reliability, unidimensionality, and model-data fitting of data set. As a result, the item scoring methods based on item response model and based on congeneric test theory were more satisfactory than the others. In addition to this, in this study, it was compared different test scoring. As known, the students who took an administered achievement tests is ranked according to their test scores. The obtained findings was reported and discussed in this context.

**Keywords:** Test scoring methods, item scoring methods, classical testing theory, item response theory, congeneric test theory

Ülkemizin gündemini çok sık bir şekilde büyük ölçekli sınavlar oluşturmaktadır. Bu sınavlarda öğrencilerin genellikle ilköğretim ya da orta öğretimdeki öğrenme alanlarına ilişkin bilişsel yeterlikleri ya da bilgi düzeyleri kestirilmeye çalışılmaktadır ve bu sınavlardan elde edilen bilgiler öğrencilerin bir üst eğitim kurumlarına yerleştirilmesinde kullanılmaktadır. Bu nedenle öğrencilerin test puanları, öğrencilerin ailelerinden öğretmenlere, eğitim alanındaki karar vericilerden eğitim araştırmacılarına kadar çok geniş bir kesimi ilgilendirmektedir. Çünkü öğrencilerin bir üst eğitim kurumlarına yerleştirilmesini amaçlayan sınavlarda öğrenciler test puanlarına göre sıralanmaktadır<sup>1</sup> ve seçme ve/veya yerleştirmeler bu başarı sırasına göre yapılmaktadır. Bu durum başarı sıralamasının doğru yapılıp yapılmadığı tartışmasını da beraberinde getirmektedir. Bu tartışma ise çoktan seçmeli testlerin puanlamasını ve test puanlarını oluşturan madde puanlama yöntemlerinin özelliklerini ön plana çıkarmaktadır. Eğitimde ölçme ve değerlendirme açısından baktığımızda; öğrencilerin maddelere verdikleri yanıtlar birer gözlenen puandır ve bu puan ölçmeye konu olan alandaki gerçek

\* Doç. Dr., Hacettepe Üniversitesi Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, [yurdugul@hacettepe.edu.tr](mailto:yurdugul@hacettepe.edu.tr)

<sup>1</sup> Ülkemizdeki öğrencilerin başarı sıralaması test puanlarına göre değil, sınav puanlarına göre yapılmaktadır. Ancak sınav puanları içerisinde en önemli payı test puanları almaktadır.

puan ile hata puanının bir birleşimidir. Oysaki ölçme modellerinin temel amacı, gözlenen puanlardan gerçek puanların kestirilmesidir (Lee, Brennan & Kolen,; 2006).

Bu çalışmanın amacı klasik test kuramına, madde yanıt kuramına ve konjenerik test kuramına dayalı test puanlama yöntemlerinin ve benzer şekilde madde puanlama yöntemlerinin karşılaştırılmasını olarak belirlenmiştir. Özellikle madde puanlama yöntemleri tekboyutluluk, model-veri uyumu ve testin güvenilirliği üzerindeki etkileri; bununla birlikte test puanlarının ise öğrencilerin sıralanmasındaki etkileri (sıra değerleri korelasyonu) araştırılacaktır. Araştırmada ele alınan veri kümesi olarak, 2001 yılında yapılan Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı kapsamında yer verilen fen bilgisi alt testindeki 25 maddeye basit rasgele örnekleme yöntemi ile seçilen 10000 öğrencinin madde yanıtları kullanılmıştır.

### Klasik Test Kuramına Dayalı Madde ve Test Puanlama Yöntemleri

Çoktan seçmeli testlerden elde edilecek test puanları madde puanlarının toplamından oluşur. Örneğin, k adet maddeden oluşan bir testten j. öğrencinin test puanı;

$$TP = X_1 + X_2 + \dots + X_k \quad (1)$$

şeklinde. Bu durumda maddelerin puanlanması bir diğer ifade ile ağırlıklandırılması önem kazanmaktadır. Madde puanlarının ağırlıklandırılmasının ön plana çıkan iki amacı; a) ölçülmeye çalışılan gerçek puanın daha iyi kestirilmesi ya da b) çoktan seçmeli testlerde seçeneklerin ağırlıklandırılması ile kısmi bilgilerin ölçülmesidir (Croker & Algina, 1986; Downing, & Haladyna, 2006). Bu çalışmada, çoktan seçmeli testlerde gerçek puanı kestirme amaçlı madde puanlama yöntemleri ele alınmıştır.

Klasik test kuramına dayalı geleneksel madde puanlama yöntemi, Bernoulli ağırlıklandırması ya da ağırlıksız madde puanlaması (unweighted item scoring) olarak bilinir (Rotou, Headrick, & Elmore, 2002; Stocking, 1996) ve bu yöntemde öğrencinin doğru seçeneği işaretlemesi "1", çeldiriciyi işaretlemesi ise "0" olarak nicelendirilir (Downing, & Haladyna, 2006; Rotou, Headrick, & Elmore, 2002; Stocking, 1996). Buna göre bir öğrencinin k adet maddeden oluşan çoktan seçmeli bir testten aldığı test puanı,

$$TP_{KTK} = X_1 + X_2 + \dots + X_k = \sum_{i=1}^k X_i \quad (2)$$

olmaktadır.

Maddelerin ağırlıksız puanlanmasına, bir diğer ifade ile madde parametrelerinden bağımsız olarak elde edilen bu madde puanlarına alternatif olarak Birnbaum (1968), her bir maddenin madde ayıricılık değerlerine göre ağırlıklandırılmasını önermiştir. Bu ağırlıklandırma için ise genellikle nokta çift serili korelasyon kullanılmaktadır (Birnbaum, 1968).

$$TP_{KTKa} = w_1 X_1 + w_2 X_2 + \dots + w_k X_k = \sum_{i=1}^k w_i X_i \quad (3)$$

Burada  $w_i$ , i. maddenin nokta çift serili korelasyonunu göstermektedir. Klasik test kuramına dayalı ağırlıklandırılmış madde puanlama yöntemi, maddenin karakteristik özelliklerini (madde ayırıcılık düzeyi) hesaba katmaktadır.

Klasik test kuramında ele alınabilecek bir başka test puanı yaklaşımı ise şans başarısından arındırılmak üzere test puanlarıdır. Özellikle çoktan seçmeli testlerde, testi alan bir öğrenci, öğrenmeye konu olan özellikten yoksun olsa bile maddeyi doğru yanıtlama düzeyi (şans başarısı) seçenek sayısı ile doğru orantılıdır (Baykul, 2000). Buna göre, öğrencinin şans başarısından arındırılmış test puanı;

$$TP_{\text{ŞBA}} = TP_{\text{KTK}} - \frac{Y}{C} \quad (4)$$

Burada k madde sayısını, Y öğrencinin yanlış yanıtladığı madde sayısını ve C ise çeldirici sayısını göstermektedir<sup>2</sup>. Croker ve Algina (1986),  $TP_{\text{ŞBA}}$  puanların gözlenen puan olmadığını, aksine gerçek puanların bir kestiricisi olduğunu ifade etmişlerdir.

#### *Klasik Test Kuramında Gözlenen Puan ve Gerçek Puan İlişkisi*

Klasik test kuramında öğrencinin bir maddeye yönelik performansı, bir diğer ifade ile gözlenen puanı (X), öğrencinin ölçmeye konu olan öğrenme alanındaki gerçek puanı (T) ve hata puanı (E) toplamından oluşmaktadır.

$$X = T + E \quad (5)$$

Klasik test kuramı, madde yanıt kuramı ve konjenerik test kuramı gibi ölçme kuramlarının temel amacı, gözlenen puanlardan gerçek puanları kestirme amacı taşımaktadır. Klasik test kuramında, gözlenen puanlardan gerçek puanların kestirilmesi için Kelley (1947) bir regresyon modeli ileri sürmüştür:

$$T_{\text{Kelley}} = \rho_X X + (1 - \rho_X) \bar{X} \quad (6)$$

Burada  $\rho_X$ , ölçmelerin güvenilirlik katsayısını göstermektedir (Downing, & Haladyna, 2006; Kelley, 1947).

#### **Madde Yanıt Kuramına Dayalı Madde ve Test Puanlama Yöntemleri**

Klasik test kuramından farklı olarak, madde yanıt kuramının (MYK) önermesi: “öğrencilerin maddeye yönelik performansı maddenin özellikleri (item characteristics) ve öğrencinin özelliklerine (person characteristics) bağlıdır.” şeklindedir. Bu nedenle, öğrencinin bir maddeyi doğru yanıtlama olasılığı  $[P(\theta)]$ , madde istatistikleri [madde güçlük düzeyi (b), madde ayırıcılık düzeyi (a)] ve

<sup>2</sup> Şans başarısının madde istatistikleri üzerine etkisi için bakınız: Baykul, 2000.

öğrencinin ölçmeye konu olan öğrenme alanındaki yeterlik düzeyi ( $\theta$ ) ile bağıntılıdır. Madde karakteristiklerinden yalnızca b parametresini ele alan bir parametrelili MYK ölçme modeli;

$$P_i(\theta_j; b) = \frac{1}{1 + e^{-(\theta - b)}} \quad (7)$$

ve madde karakteristiklerinden a ve b parametrelerini ele alan iki parametrelili MYK ölçme modeli;

$$P_i(\theta_j; a, b) = \frac{1}{1 + e^{-a(\theta - b)}} \quad (8)$$

şeklinde. Burada  $P_i(\theta_j)$ , j. yeterlik düzeyindeki öğrencinin i. maddeyi doğru yanıtlama olasılığını göstermektedir. Bu olasılıkların k madde üzerinde toplamı ise, j. öğrencinin test puanını vermektedir:

$$\begin{aligned} TP_{MYK\_1P} &= P_1(\theta_j; b) + P_2(\theta_j; b) + \dots + P_k(\theta_j; b) \\ &= \sum_{i=1}^k P(\theta_j; b) \end{aligned} \quad (9)$$

$$\begin{aligned} TP_{MYK\_2P} &= P_1(\theta_j; b, a) + P_2(\theta_j; b, a) + \dots + P_k(\theta_j; b, a) \\ &= \sum_{i=1}^k P(\theta_j; b, a) \end{aligned} \quad (10)$$

### Madde Yanıt Kuramında Gözlenen Puan ve Gerçek Puan İlişkisi

KTK'nda gözlenen puan ile test puanı arasındaki bağıntı Kelley'in regresyon denklemi ile ortaya konmaktadır. Ancak MYK'nda Eşitlik 9 ve Eşitlik 10 ile ulaşılan değerler aynı zamanda öğrencilerin lojistik model ya da normal ogive modellerden kestirilen gerçek puanlardır. MYK'nın modeli kestirilirken öğrencilerin k adet maddeye verdiği 0 ya da 1 şeklinde puanlanmış (number correct scoring) yanıt vektörü<sup>3</sup> ( $u$ ) kullanılmaktadır.

$$u_j = [X_1 \ X_2 \ \dots \ X_k]$$

Burada  $u_j$ , j. öğrencinin k adet maddeye verdiği ağırlıksız yanıtlardan elde edilen yanıt vektörünü (gözlenen puanları) göstermektedir. Bu vektör Eşitlik 7 ya da Eşitlik 8 ile verilen modellerin kestiriminde bir girdi olarak kullanılarak öğrencinin yeterlik düzeyleri (person characteristic) ve madde karakteristikleri elde edilmekte ve daha sonra ise Eşitlik 9-10 gerçek puanlara ulaşılmaktadır. Kısacası, MYK'nda (KTK'ndan farklı olarak) gözlenen puanlar birer girdi olarak kullanılmaktadır ve model çıktısı ise gerçek puanlardır.

<sup>3</sup> Ağırlıksız madde puanlarına alternatif olarak, MYK modellerinin kestiriminde madde yanıt örüntülerinin (item pattern) kullanımları da söz konusudur.

### Konjenerik Test Kuramına Dayalı Madde ve Test Puanlama Yöntemleri

Konjenerik test kuramı KTK'nın bir alt boyutudur (Lucke, 2005) ve Eşitlik 1 ile verilen doğrusal modelin madde karakteristiklerini kapsayan geliştirilmiş bir biçimdir. Konjenerik ölçme modeli;

$$X_i = \mu_i + \lambda_i F + E_i$$

şeklinde (Lucke, 2005; McDonald, 1999; Yurdugül, 2005). Burada  $\mu_i$  ifadesi; i. ölçmenin/maddenin güçlük indeksini,  $\lambda_i$  ise i. maddenin ayırıcılık değerini ve F ise tüm ölçmelerin ortak faktör değerini göstermektedir. McDonald (1999) bu modele dayalı madde puanını;

$$T_i = \mu_i + \lambda_i F \quad (11)$$

olarak tanımlamıştır. Yine McDonald (1999) testi alan bir öğrencinin test puanını elde etmek üzere Kelley'in yaklaşımından yola çıkarak yeni bir eşitlik önermiştir.

$$TP_{KonTK} = \left[ \sum_{i=1}^k \frac{\lambda_i^2}{\psi_i^2} \right]^{-1} \sum_{i=1}^k \left[ \frac{\lambda_i}{\psi_j^2} \right] (x_{ij} - \mu_i) + \sum_{i=1}^k \mu_i \quad (12)$$

Bu araştırmada, farklı ölçme modellerini temel alan madde puanlama yöntemleri ve test puanlama yöntemleri karşılaştırılmıştır.

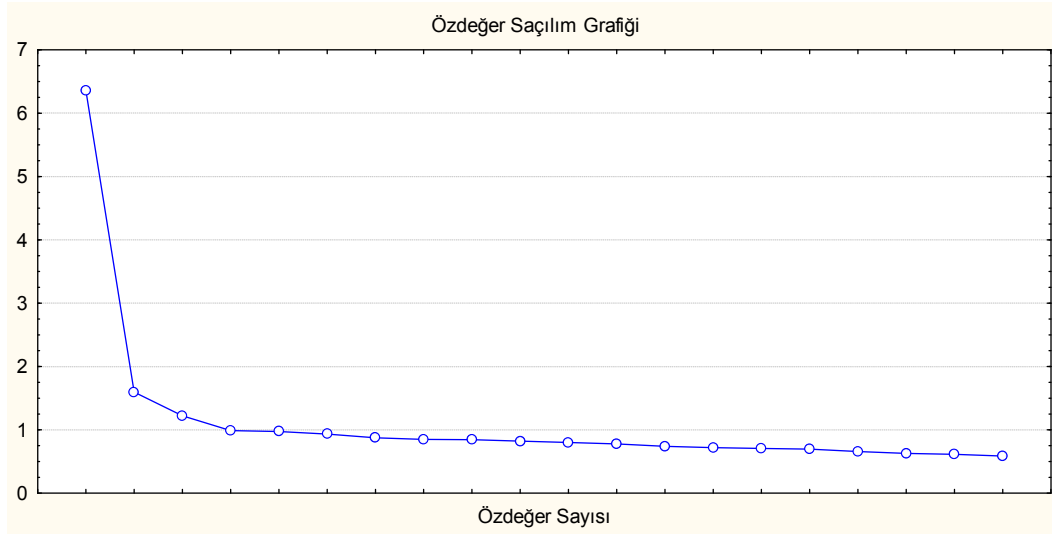
### Yöntem

Bu araştırmada veri kümesi olarak, 2001 yılında yapılan Ortaöğretim Kurumları Seçme ve Yerleştirme Sınavı'nda (ÖÖK-ÖSYS) yer alan 25 maddelik fen bilgisi alt testine 10000 öğrencinin verdiği yanıtlar kullanılmıştır.

Bu çalışmada KTK, MYK ve KonTK üzerine kurulu madde puanlama ve test puanlama yöntemleri incelenmiştir. KTK kapsamındaki *madde puanlama yöntemleri* olarak 1) ağırlıksız puanlama yöntemi ve 2) ağırlıklandırılmış puanlama yöntemi, MYK kapsamında 3) bir parametrelili MYK ve 4) iki parametrelili MYK modelinden üretilen madde puanlama yöntemleri ile KonTK kapsamında ise 5) Eşitlik 11 ile verilen yöntem ele alınmıştır. Karşılaştırmalar, veri kümesinin güvenilirliği, tekboyutluluk değerleri ve model-veri uyumu değerlerine göre yapılmıştır. Özellikle ağırlıksız madde puanlama yönteminde veri kümesi 0 ve 1 olduğu için  $25 \times 10000 = 250000$  adet ölçmenin yer aldığı veri kümesinin güvenilirliğini elde etmek amacıyla  $KR_{20}$  ve Cronbach alfa güvenilirlik katsayıları hesaplanmış ve her iki katsayı değeri 0,782 olarak bulunmuştur. Diğer madde puanlama yöntemlerinden elde edilen veri kümeleri için ise alfa güvenilirlik katsayı kullanılmıştır. Her bir madde puanlama yönteminden elde edilen veri kümelerinin tekboyutluluk değerleri hakkında bilgi edinmek amacıyla birinci ve ikinci özdeğerler hesaplanmıştır. Son karşılaştırma değeri ise model-veri uyumunu göstermek üzere uyum iyiliği indeksi üzerine yapılandırılmıştır. Burada uyum iyiliği indisi Jöreskog tarafından geliştirilen doğrulayıcı faktör analizinden elde edilen GFI indisi kullanılmıştır. (Yurdugül, 2005)

MYK modellerinin çözümlenebilmesi için gerekli olan varsayımlar tekboyutluluk ve yerel bağımsızlıklar test edilmiş ve bunların sağlandığı gözlenmiştir (Şekil 1). Bununla birlikte bir ve iki

parametrelili MYK modellerinin veri-uyum değerleri istenilen düzeydeki iken üç parametrelili MYK modelinin kestiriminde veri-model uyumu sağlanamadığı için çalışma kapsamına alınmamıştır.



**Şekil 1:** Veri kümesinin tetrakorik korelasyon üzerine kurulu faktör çözümü

Test puanları olarak ise; KTK kapsamındaki  $TP_{KTK}$  (Eşitlik 2),  $TP_{KTKa}$  (Eşitlik 3),  $TP_{SBA}$  (Eşitlik 4) ve  $T_{Kelley}$  (Eşitlik 6) test puanlama yöntemleri, MYK kapsamındaki test puanlama yöntemi olarak  $TP_{MYK\_1P}$  (Eşitlik 9) ve  $TP_{MYK\_2P}$  (Eşitlik 10) ve KonTK kapsamındaki test puanlama yöntemi olan  $TP_{KonTK}$  (Eşitlik 12) ele alınmıştır. Basit rastgele örnekleme yöntemi ile seçilen 10000 öğrenci için inceleme kapsamındaki her bir test puanlama yöntemine göre test puanları hesaplanmıştır. Bu puanlamalara göre öğrencilerin başarı düzeylerine göre sıralama farklılıklarını belirlemek amacıyla Sperman'ın sıralama korelasyon katsayısı hesaplanmış ve yöntemler arası puan sıralaması uyumu bu korelasyon değerlerine göre yapılandırılmıştır.

### Bulgular

Madde puanlama yöntemlerinin karşılaştırılması için 10000 öğrencinin 25 maddeye ilişkin yanıtlarını içeren 5 farklı veri kümesi ile çalışılmıştır. Bu veri kümeleri sırasıyla; ağırlıksız madde puanları, ağırlıklandırılmış madde puanları, bir parametrelili MYK, iki parametrelili MYK ve son olarak Konjenerik madde puanlarından üretilmiştir. Bu veri kümelerine ilişkin elde edilen değerler Tablo 1'de verilmiştir.

**Tablo 1:** Madde puanlama yöntemlerinin karşılaştırılması

Veri Kümesi (Madde Puanları)	Model	Puan Türü	Güvenirlilik	Özdeğer Oranları	GFI
Ağırlıksız	KTK	Gözlenen	0,782	4,137	0,980
Ağırlıklı	KTK	Gerçek	0,790	4,137	0,960
1 Parametrelili MYK	MYK	Gerçek	0,996	22,888	1,000
2 Parametrelili MYK	MYK	Gerçek	0,995	22,845	0,992
Konjenerik	KonTK	Gerçek	0,992	24,975	1,000

Elde edilen bulgular doğrultusunda, MYK ve KonTK'na dayalı madde puanlama yöntemlerinden oluşturulan veri kümesinin güvenirlilik değerleri KTK kapsamındaki madde puanlama yöntemlerinden daha yüksek elde edilmiştir. Yine, birinci özdeğerin ikinci özdeğere oranı olarak en yüksek oran konjenerik madde puanlama yönteminde gözlenmiştir. Daha sonra bu değerleri MYK

modelinden elde edilen madde puanları takip etmiştir. En düşük özdeğer oranı ise KTK kapsamındaki madde puanlarında gözlenmiştir<sup>4</sup>.

Farklı madde puanlamalarının tekboyutlu model-veri uyumu sınavında ise en büyük uyum değerleri bir parametrelili MYK madde puanları ve konjenerik madde puanlarında gözlenmiştir. Test puanlarının karşılaştırılması için ise 7 farklı test puanlama yönteminden elde edilen örneklemdeki 10000 adet öğrencinin başarı düzeyleri elde edilmiş bunlar Spearman sıralama korelasyonu ile öğrencilerin başarı sıralamasındaki tutarlılıkları incelenmiştir.

Tablo 2’den görüleceği gibi, ağırlıksız madde puanlarının toplamından elde edilen test puanları ( $TP_{KTK}$ ), öğrencilerin başarı sıralaması açısından  $T_{Kelley}$  ve  $TP_{MYK\_1P}$  ile tam bir uyum söz konusudur.  $TP_{KTK}$ , açısından mükemmel uyum olmasa bile en yüksek korelasyonu  $TP_{KonTK}$  arasında gözlenmiştir.

Tablo 2: Test Puanlarının Spearman Sıralama Korelasyon Değerler

	$TP_{KTK}$	$TP_{SBA}$	$T_{Kelley}$	$TP_{KTKa}$	$TP_{MYK\_1P}$	$TP_{MYK\_2P}$	$TP_{KonTK}$
<b>Açıklama</b>	Eşitlik 2	Eşitlik 4	Eşitlik 6	Eşitlik 3	Eşitlik 9	Eşitlik 10	Eşitlik 12
<b>Puan Türü</b>	Gözlenen	Gerçek	Gerçek	Gerçek	Gerçek	Gerçek	Gerçek
$TP_{KTK}$	1,000						
$TP_{SBA}$	0,963	1,000					
$T_{Kelley}$	1,000	0,963	1,000				
$TP_{KTKa}$	0,992	0,961	0,992	1,000			
$TP_{MYK\_1P}$	1,000	0,963	1,000	0,992	1,000		
$TP_{MYK\_2P}$	0,915	0,893	0,915	0,925	0,915	1,000	
$TP_{KonTK}$	0,987	0,958	0,987	0,999	0,987	0,926	1,000

Bununla birlikte  $TP_{KTK}$ , en düşük korelasyonu şans başarısından arındırılmış test puanları ( $TP_{SBA}$ ) ve  $TP_{MYK\_2P}$  ile göstermiştir. Tüm karşılaştırmalar içerisinde en düşük korelasyon ise  $TP_{SBA}$  ve  $TP_{MYK\_2P}$  arasında gözlenmiştir.

### Sonuç ve Öneriler

Bu çalışmada farklı *madde puanlama türleri* ve *test puanlama türleri* ele alınarak karşılaştırılmıştır. *Madde puanları* üzerine olan karşılaştırmalar, eğitimde ölçme ve değerlendirme alanında çalışan araştırmacılara kullanacakları veri kümesi hakkında bilgiye ulaşma amacı taşımaktadır. Bununla birlikte, *test puanları* kapsamındaki karşılaştırmalar ise eğitim alanındaki karar vericilere yönelik olması amaçlanmıştır.

Madde puanlama türleri arasında madde puanlarının oluşturduğu veri kümesinin güvenilirliği, tekboyutluluğu ve veri-model uyumu açısından MYK ve KonTK kapsamındaki puanlama türleri en iyi sonuçları vermiştir.

Diğer taraftan test puanları olarak ise ülkemizde sıkça ele alınan şans başarısından arındırılmış test puanları kullanılmakta ve öğrencilerin başarı sıralaması bu puanlara göre yapılmaktadır. Ancak bu çalışmada göstermiştir ki, şans başarısından arındırılmış test puanlarına göre başarı sıralamaları diğer test puanı türleri ile en düşük uyumu göstermektedir. Bu konuda ortaya çıkabilecek kuşular seçme ve yerleştirme sistemlerinin adaletli olup olmadığı konusundaki tartışmaya yol açabilecektir. Bu konuda “en iyi test puanı” kavramını tartışabilmek için her bir test puanına ilişkin yordama ve/veya uygunluk geçerliği gücünün araştırılması gerekir.

<sup>4</sup> Ancak Şekil 1’de gösterildiği gibi, 0 ve 1 puanları (birinci özdeğer=4,137) yerine tetrakorik korelasyon üzerine kurulu faktör çözümlemesinde bu değer 6,360’a yükselmiştir.

Son olarak vurgulanması gereken bir diğer konu ise;  $TP_{KTK}$  doğrudan “gözlenen puan (X)” ve  $TP_{KTK}$  dışındaki tüm test puanlama yöntemleri ise gerçek puanı (T) temsil etmektedir. Aynı şekilde ağırlıksız madde puanları dışındaki tüm madde puanları da gerçek puanlardır ( $T_i$ ). Bilindiği gibi, eğitim alanında yapılan ölçmeler bir ölçme modeli üzerinde gerçekleştirilir ve ölçme modelleri ise gözlenen puanlardan gerçek puanların kestirilmesine olanak veren istatistiksel bağıntılardır. Gözlenen puanları gerçek puan gibi ele alındığı durumlar ( $X=T$ ) hatanın olmadığı, mükemmel güvenilirlik (perfect reliability) düzeyine sahip (paralel ya da eşdeğer) ölçme kümeleri için olanaklıdır. Bu tür ölçme kümelerinin ortak özelliği ise varyansları ve maddeler arası kovaryans değerlerinin eşit olduğu ölçmelerden oluşmasıdır. Örneğin; bir ölçme kümesindeki tüm maddelerin güçlük düzeylerinin  $p_i=0,5$  düzeyinde olması durumunda tüm madde puanlarının varyansları ve kovaryansları eşit olacaktır, bunun doğal sonucu olarak da güvenilirlik düzeyi 1 olarak elde edilecektir. Ancak bu durumda  $X=T$  bağıntısı kullanılabilir. Test puanlarına göre öğrenci başarılarının sıralanması söz konusu olduğunda ölçmelerin paralel ya da eşdeğer ölçmeler olmadığı durumlarda gerçek puan kestirimlerini kullanılması önerilir.

### Kaynaklar

- Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara: ÖSYM Yayınları.
- Birnbaum, A. (1968). Some latent models and their use in inferring an examinee's ability. In F.M. Lord & M. R.Novick (Eds.), *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Downing, S.M., & Haladyna, T.M. (2006). *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum.
- Hambleton, R., Swaminathan, H.,&Rogers, J. (1991). *Fundamentals of item response theory*. London: Sage.
- Lee, W., Brennan, R. L., & Kolen, M. J. (2006). Interval estimation for true raw and scale scores under the binomial error model. *Journal of Educational and Behavioral Statistics*, 31 ( 3), 261–281.
- Lucke, J. F. (2005). The  $\alpha$  and the  $\omega$  of congeneric test theory: an extension of reliability and internal consistency to heterogeneous tests. *Applied Psychological Measurement*, 29(1), 65–81.
- McDonald, R. P. (1999). *Test theory: a unified treatment*. Mahwah NJ: Erlbaum.
- Rotou, O., Headrick, T. C., & Elmore P. B. (2002). A Proposed Number Correct Scoring Procedure Based on Classical True-Score Theory and Multidimensional Item Response Theory , *International Journal of Testing*, 2(2), 131 – 141.
- Stocking, L. M. (1996). An alternative method for scoring adaptive tests. *Journal of Educational and Behavioral Statistics*, 21(4), 365–389.
- Yurdugül, H. (2005). Konjenerik test kuramı ve konjenerik madde analizi: Tek boyutlu çoktan seçmeli testler üzerine bir uygulama. *A.Ü. Eğitim Bilimleri Fakültesi Dergisi*, 38(2), 21-47