



Obtaining Classical Reliability Terms from Item Response Theory in Multiple Choice Tests

Halil Yurdugül*

ABSTRACT: The classical testing theory and item response theory are commonly used in the analysis of educational measurements. On the other hand, classical reliability is only obtained from the terms of classical testing theory. In item response theory, however, the reliability is derived from information functions. Dimitrov has revealed that classical reliability terms could also be obtained from item response theory by equations which he developed. This study firstly investigates item and test reliability obtained through item response theory by using Dimitrov's equations. Then, these reliabilities would be compared with classical reliability.

Keywords: Item response theory, classical testing theory, reliability, item reliability, test reliability

SUMMARY

Purpose and significance: Dimitrov has shown how to get the variances of terms in the classical test model from the item response model. In this study, the variances of terms of classical test model were obtained from item response model with the Dimitrov's equations on a multiple choice data. As known, the variances also are parameters of classical reliabilities. In this

*Halil Yurdugül, Ph. D. Hacettepe University, Faculty of Education, yurdugul@hacettepe.edu.tr

way, the purpose of this research is to compare the classical item and test reliabilities based on the Dimitrov's equations, and classical test model.

Methods: In this study, it was used the data of the multiple choice test, SSPE-SE (Student Selection and Placement Examination for Secondary Education) in Turkey. The test consists of 100 multiple-choice items, divided into four subtests: Turkish, mathematics, sciences, social sciences, and each subtest consist of 25 items. In the present study, the data gathered from 553108 pupils and SSPE-SE carried out in 2001 was utilized. In this study, it was investigated the data of Turkish subtest in SSPE-SE and given the variance-covariance matrix of the subtest in Appendix.

In this study, the item reliabilities were calculated with three methods are: a) classical item analysis, b) classical item reliabilities according to the Dimitrov's (2003a) equalities, and c) item reliabilities based on terms of confirmatory factor analysis as the determination coefficient of measurement models. The results were compared.

Also, the test reliabilities were calculated according to Kuder Richardson (KR_{20}) formulas and the Dimitrov's equalities.

Results: The ρ_{Di} , the item reliability based on the Dimitrov's equalities were high correlation with classical item reliability (ρ_i) and item reliabilities (R^2) in terms of confirmatory factor analysis. However, in different samples, the test reliabilities obtained from KR_{20} were higher than classical test reliabilities based on the Dimitrov's equalities as average 0.03 differently.

Discussion and Conclusions: In this research, it was found that the classical item and test reliabilities can be obtained from the terms in item response model by the Dimitrov's equalities.



Çoktan Seçmeli Testlerde Klasik Güvenirlik Terimlerinin Madde-Yanıt Kuramından Elde Edilmesi

Halil Yurdugül*

ÖZ: Klasik test kuramı ve madde-yanıt kuramı eğitimde ölçme sonuçlarının çözümlenmesinde yaygın olarak kullanılan iki farklı yaklaşımdır. Öte yandan klasik güvenilirlikler klasik test kuramının terimlerinden elde edilmektedir. Madde-yanıt kuramında ise güvenilirlikler bilgi fonksiyonlarından elde edilmektedir. Dimitrov, geliştirmiş olduğu eşitlikler ile klasik güvenilirlik terimlerinin madde-yanıt kuramından elde edilebileceğini ortaya koymuştur. Bu çalışmada, çoktan seçmeli testlerde Dimitrov eşitlikleri kullanılarak madde-yanıt kuramına dayalı olarak madde ve test güvenilirlikleri ele alınmıştır. Daha sonra bu güvenilirlikler klasik güvenilirlikler ile karşılaştırılmıştır.

Anahtar Sözcükler: Madde-yanıt kuramı, klasik test kuramı, güvenilirlik, madde güvenirligi, test güvenirligi

GİRİŞ

Eğitim alanındaki ölçme sonuçlarının çözümlenmesinde yaygın olarak klasik test kuramı (classical test theory) ve madde-yanıt kuramı (item response theory) kullanılmaktadır. Klasik test kuramı modeli, tek bir ölçmeye/maddeye ilişkin olarak gözlenen puanlar (X_i), ölçülmek istenilen gerçek puanlar (T_i) ve ölçmeye karışan hata puanları (E_i) arasındaki *doğrusal bağıntı* ile ifade edilir. Madde-yanıt kuramı modelinde ise bireylerin bir maddeyi doğru yanıtlama olasılığı, ölçmeye konu olan alandaki yeterlik

* Dr. Halil Yurdugül, Hacettepe Üniversitesi, Eğitim Fakültesi, yurdugul@hacettepe.edu.tr

düzeyi ve maddenin karakteristik özellikleri arasındaki *doğrusal olmayan bağıntı* ile açıklanmaktadır (Lord, 1980).

Her iki kuramsal yaklaşım, fonksiyonel bağıntı olarak birbirinden farklı olsa da ölçme kümeleri normal dağılım gösterdiği durumlarda klasik test kuramı (KTK) modelinde yer alan madde parametrelerini, madde-yanıt kuramı (MYK) modelinden elde etmek olanaklıdır (Hambleton ve Swaminathan, 1985; Kelecioğlu, 2001). Ancak normal dağılım koşulu altında MYK terimlerinden KTK terimlerinin elde edilmesine karşın MYK'ndan KTK modelindeki terimlerin varyanslarını elde etmek için geliştirilmiş bir yöntem yakın zamana kadar söz konusu değildi. Dimiter Dimitrov (2003a, 2003b) yapmış olduğu simülatif bir çözümlenme ile KTK modelindeki terimlere ilişkin varyansları çeşitli fonksiyonlar yardımıyla MYK modelinden elde edilebileceğini göstermiştir.

Bilindiği gibi; KTK modelinde yer alan gözlenen puan, gerçek puan ve hata puanları terimlerinin varyansları güvenilirlik ile doğrudan ilişkilidir. Ölçme araçlarının hatasızlığının bir ölçütü olan güvenilirlik kavramı, KTK modelindeki terimlerin varyansları ile ifade edilir (güvenirlik, gerçek puanlar varyansının gözlenen puanlar varyansına oranıdır). Dimitrov (2003a, 2003b), bu terimlerin MYK modelinden de elde edilebileceğini ortaya koymuştur. Bu durum Dimitrov'un (2003a, 2003b) geliştirdiği eşitlikler yardımı ile madde-yanıt kuramından klasik güvenilirliklerin elde edilebileceği anlamına gelmektedir.

MYK'nda (KTK'nın aksine) doğrudan ölçme sonuçlarına ilişkin bir güvenilirlik kestirimi söz konusu değildir (Doran, 2005). Çünkü MYK'nda her bir yetenek düzeyinden elde edilen bilgi fonksiyonlarıyla (information function) ifade edilebilen bir güvenilirlikten bahsedilebilir. Diğer taraftan Doran (2005), bilgi fonksiyonlarına dayalı olarak elde edilen güvenilirliklerin klasik güvenilirlikten farklı olduğunu ortaya koymuştur. Dimitrov'un (2003a, 2003b) geliştirdiği eşitlikler yardımı ile MYK'ndan elde edilen değerler ile klasik güvenilirliklerin elde edilmesi olanaklı duruma gelmiştir.

Bu çalışmada çoktan seçmeli testlerde, Dimitrov'un (2003a) eşitlikleri kullanılarak MYK modelinden elde edilen terimler yardımıyla üretilen klasik güvenilirlikler ile KTK modelinden elde edilen madde ve test güvenilirlik katsayıları karşılaştırılmıştır¹.

¹ Bu çalışmada farklı ölçme kuramlarından elde edilen terimlerden yararlanılarak *klasik güvenilirlikler* karşılaştırılmıştır. Bu nedenle çalışma kapsamına yalnızca klasik güvenilirlikler alınmış, MYK'na dayalı güvenilirlikler çalışma kapsamı dışında tutulmuştur.

Klasik Test Kuramı ve Güvenirlik

Eğitim alanında kullanılan ölçme araçlarının güvenilirliği aynı zamanda ölçme sonuçlarının kesinliğinin bir ölçüsüdür. Güvenirlik kavramına ilişkin operasyonel tanım genellikle klasik test kuramının ($X=T+E$) doğrusal modelinin terimleriyle ifade edilir.

$$X=T+E \quad (1)$$

$$\text{Var}(X)=\text{Var}(T)+\text{Var}(E) \quad (2)$$

Buna göre güvenirlik; gerçek puanlar (T) varyansının gözlenen puanlar (X) varyansına oranı ya da hata puanları (E) varyansının bir fonksiyonu olarak elde edilir.

$$\rho^2_{XT} = \frac{\text{Var}(T)}{\text{Var}(X)} = 1 - \frac{\text{Var}(E)}{\text{Var}(X)} \quad (3)$$

KTK modelindeki gözlenen puanların bir bileşeni olan *gerçek puanlar* doğrudan gözlenemediğinden dolayı Lord ve Novick (1968) tarafından kullanılan ifadeye göre “platonik” bir yapısı vardır. Bu nedenle Eşitlik 3 ile verilen güvenirlik indeksini elde edebilmek için paralel, eşdeğer (tau-equivalent) ya da eşbiçimli (essentially tau-equivalent) ölçmelere ihtiyaç duyulur (Traub, 1994). Eğitimde paralel ya da eşdeğer ölçmeler olarak genellikle k adet maddeden oluşan tek boyutlu birleşik testler (composite test) kullanılır ($X=X_1+X_2+\dots+X_k$). Birleşik test maddeleri arasındaki korelasyon aynı zamanda güvenirlik indeksinin kestiricisi olan güvenirlik katsayısını (ρ_{X_1,X_2}) verir (Lord ve Novick, 1968).

$$\rho^2_{XT} \leq \rho_{X_1,X_2} \quad (4)$$

Çoktan seçmeli testlerden elde edilen puanların güvenirliklerin kestirilmesinde en yaygın kullanılan güvenirlik katsayısı, Kuder-Richardson tarafından geliştirilen KR_{20} ve KR_{21} katsayılarıdır.

$$KR_{20} = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \pi_i (1 - \pi_i)}{\text{Var}(X)} \right) \quad (5)$$

Burada π_i ; i. maddenin güçlük indeksini, $[\pi_i(1-\pi_i)]$; i. Madde puanları varyansını ve $\text{Var}(X)$ ise birleşik test puanlarının varyansını göstermektedir (Lord ve Novick, 1968).

Klasik Güvenirlik Terimlerinin Madde-Yanıt Modelinden Elde Edilmesi

MYK'na göre; öğrencilerin herhangi bir maddeyi doğru yanıtlama olasılıkları, öğrencilerin sahip olduğu özellik düzeyi ve maddenin karakteristik özellikleri ile açıklanır. KTK, ölçülmek istenilen özellik ile gözlem sonuçları arasındaki bağıntıyı doğrusal bir model ile açıklarken; MYK ise bu bağıntıyı doğrusal olmayan model ile açıklamaktadır (Hambleton ve Swaminathan, 1985). Buna göre 2 parametrelî model;

$$P_i(\theta) = \frac{e^{a_i(\theta-b_i)}}{1 + e^{a_i(\theta-b_i)}} \quad (6)$$

burada $P_i(\theta)$, θ yeteneğine sahip öğrencinin i . maddeyi doğru yanıtlama olasılığını, a_i ; i . maddenin ayırıcılık gücünü ve b_i ise i . maddenin güçlük düzeyini göstermektedir. Buna göre i . maddeye ilişkin doğru yanıtların marjinal olasılığı:

$$\pi_i = \int_{-\infty}^{\infty} P_i(\theta)\varphi(\theta)d\theta \quad (7)$$

burada $\varphi(\theta)$; ölçülmeye çalışılan özelliğin dağılımına ilişkin fonksiyondur. Dimitrov (2003a, 2003b), bu fonksiyonun simülatif çözümünü yapmış ve basit bir fonksiyona indirgemıştır.

$$\pi_i = \frac{1 - hf(\eta_i)}{2} \quad (8)$$

Burada π_i ; klasik test kuramındaki madde güçlük (aslında madde kolaylık) indeksine karşılık gelmektedir ve $hf(\eta_i)$ ise i . maddeye ilişkin hata fonksiyonudur. Bu fonksiyonun parametresi Eşitlik 9'da ve fonksiyonu Eşitlik 10'da verilmiştir.

$$\eta_i = \frac{a_i b_i}{\sqrt{2(1 + a_i^2)}} \quad (9)$$

$$hf(\eta) = 1 - \left(\frac{1}{1 + 0.278393 \eta + 0.230389 \eta^2 + 0.000972 \eta^3 + 0.078108 \eta^4} \right)^4$$

Madde güçlüklerine ilişkin hesaplamalarda, $b_i < 0$ olduğunda $\eta_i < 0$ olacaktır ve bu durumda hata fonksiyonunun mutlak değeri, $|\eta_i|$ ve hata fonksiyonunun negatifi alınır, $hf(-\eta) = -hf(\eta)$ (Dimitrov, 2003a, 2003b).

Madde Varyansları

Dimitrov'un (2003a) geliştirdiği fonksiyonlar kullanılarak, KTK modelinde yer alan terimlerin varyansları, MYK modelinin çözümlenmesi ile elde edilebilmektedir. Bu terimler gözlenen puanlar varyansı $[\text{Var}(X_i)]$, gerçek puanlar varyansı $[\text{Var}(T_i)]$ ve hata puanları varyansıdır $[\text{Var}(E_i)]$.

Çoktan seçmeli birleşik ölçmelerde gözlenen puanlarda varyansı, multinominal dağılımın parametresi olarak Eşitlik 10 ile verilen ifade ile elde edilebilir.

$$\text{Var}(X_i) = \pi_i(1 - \pi_i) \quad (10)$$

Burada π_i , Eşitlik 8 ile çözümlenen madde güçlük düzeylerini göstermektedir. MYK'na göre; herhangi bir maddeye ilişkin hata varyansı ise Eşitlik 11'de verilmiştir.

$$\text{Var}(E_i) = \int_{-\infty}^{\infty} \frac{e^{a_i(\theta - b_i)}}{[1 + e^{a_i(\theta - b_i)}]^2} \left(\frac{e^{-0.5\theta^2}}{\sqrt{2\pi}} \right) d\theta \quad (11)$$

Eşitlik 11'deki ifadenin çözümlenmesi ile i. maddeye ilişkin hata varyansı elde edilmektedir. Dimitrov (2003a, 2003b), bu ifadenin fonksiyonel çözümünü a_i ve b_i 'nin fonksiyonu olacak şekilde basit bir fonksiyona indirgeyerek önermiştir.

$$\text{Var}(E_i) = 0.2646 - 0.118 a_i + 0.0187 a_i^2 e^{-0.5(b_i/d_i)^2} \quad (12)$$

burada $d_i = (0.7427) + (0.7081/a_i) + (0.0074/a_i^2)$.

Tek bir maddeye ilişkin gerçek puanlar varyansı, Eşitlik 1 ve Eşitlik 2 ile verilen KTK modelindeki doğrusal ilişkilerden elde edilebilmektedir.

Buna göre i. maddenin gerçek puan varyansı, Eşitlik 10 ile verilen maddeye ilişkin gözlenen puanlar varyansı ile Eşitlik 12 ile verilen madde hata puanları varyansının farkına eşittir.

$$\text{Var}(T_i) = \pi_i(1-\pi_i) - \text{Var}(E_i) \quad (13)$$

Eşitlik 10, Eşitlik 12 ve Eşitlik 13 yardımı ile maddelere ilişkin güvenilirlikler;

$$\rho_{Di} = \frac{\text{Var}(T)}{\text{Var}(X)}$$

$$\rho_{Di} = \frac{\pi_i(1-\pi_i) - \text{Var}(E)}{\pi_i(1-\pi_i)} \quad (14)$$

şeklinde elde edilebilmektedir.

Madde güvenilirliklerinin klasik test kuramındaki karşılığı ise;

$$\rho_i = \rho_{iX} \sqrt{\text{Var}(X_i)} \quad (15)$$

şeklinde dir. Burada ρ_{iX} , maddenin ayırtıcılık gücünü göstermektedir (Baykul, 2000).

Test Varyansları ve Güvenirlik

Tek boyutlu birleşik testlerde yer alan madde puanlarının toplamı, tüm maddelerin ölçmeye yöneldiği tek bir özelliğe ilişkin toplam gözlenen puanları verir. Benzer şekilde madde varyanslarının toplamı da testin tümüne ilişkin varyanslara eşittir. Buna göre testin hata varyansı;

$$\text{Var}(E) = \sum_{i=1}^k \text{Var}(E_i) \quad (16)$$

Testin gerçek puanlar varyansı;

$$\text{Var}(T) = \sum_{i=1}^k \sum_{j=1}^k \sqrt{[\pi_i(1-\pi_i) - \text{Var}(E_i)][\pi_j(1-\pi_j) - \text{Var}(E_j)]} \quad (17)$$

şeklinde elde edilmektedir (Dimitrov, 2003a, 2003b).

Buna göre; Eşitlik 1 ve Eşitlik 3 ile verilen bağıntılardan çoktan seçmeli test puanlarının güvenilirliği, Dimitrov (2003a) yaklaşımına göre; Eşitlik 18 ile verilen ifade ile elde edilebilir.

$$\rho_D = \frac{\sum_{i=1}^k \sum_{j=1}^k \sqrt{[\pi_i(1-\pi_i) - \text{Var}(E_i)][\pi_j(1-\pi_j) - \text{Var}(E_j)]}}{\sum_{i=1}^k \pi_i(1-\pi_i)} \quad (18)$$

Bu çalışmada çoktan seçmeli test sonuçları üzerinde klasik madde güvenirlikleri olarak ρ_{Di} ve ρ_i , klasik test güvenirlikleri olarak da KR_{20} ve ρ_D katsayıları karşılaştırılmıştır.

YÖNTEM

Bu çalışmanın uygulama bölümü iki aşamalı olarak ele alınmıştır. İlk aşamada KTK'na dayalı klasik madde güvenirlikleri ve Dimitrov tekniği ile MYK'ndan elde edilen madde güvenirliklerinin karşılaştırılmalarına yer verilmiştir. Bu karşılaştırmalara dayanak oluşturması açısından aynı zamanda kovaryans terimleri ile elde edilen doğrulayıcı faktör analizinde (DFA) maddelere ilişkin model determinasyon (R^2) değerleri madde güvenirlikleri olarak kullanılmıştır. Buna göre, üç farklı yöntemden elde edilen madde güvenirlikleri karşılaştırılmıştır. Karşılaştırmalarda 2001 yılında uygulanan Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı (OK-ÖSYS) Türkçe alt testinde yer alan 25 maddeye ilişkin 553108 adet öğrencinin yanıtlarından oluşan veri kümesi kullanılmıştır.

Madde Yanıt Kuramı Modelinin Çözümlemesi

Madde puanlarından oluşan ölçme kümesi Eşitlik 6 ile verilen 2 parametrelili lojistik MYK modelinden yararlanılarak çözümlenmiştir. Bu çözümlemelerde BILOG paket programından yararlanılmıştır. Modelin çözümlenmesiyle elde edilen a_i ve b_i parametreleri Dimitrov tekniğine girdi olarak kullanılarak KTK modelindeki terimlerin varyans karşılıkları bulunmuştur. Eşitlik 8 ile verilen ifade kullanılarak madde güçlük düzeyleri (π_i), Eşitlik 10 ile madde puanları varyansları, Eşitlik 13 ile maddelerin gerçek puanlar varyansları, Eşitlik 12 ile hata varyansları ve Eşitlik 14 ile verilen madde güvenirlikleri kestirilmiştir. Aynı şekilde birleşik teste ilişkin test varyansları da elde edilerek test güvenirlikleri kestirilmiştir (Eşitlik 18).

Klasik Test Kuramı Modelinin Çözümlemesi

Bu kapsamda ilgili maddelerin KTK ilkelerine dayalı olarak madde analizi yapılmıştır. Madde analizi sonucu madde parametreleri (güçlük düzeyleri ve madde ayırıcılık düzeyleri) ve aynı zamanda Eşitlik 15 ile verilen ifade kullanılarak madde güvenirlikleri elde edilmiştir. Ancak, KTK

modelindeki gerçek ve hata puanları terimlerinin varyanslarını elde edebilmek için doğrulayıcı faktör analitik çözümlemesine gidilmiştir.

Doğrulayıcı Faktör Analitik Modelinin Çözülmesi

Doğrulayıcı faktör analitik (DFA) modelinin çözümlenmesinde maddelere ilişkin varyans-kovaryans matrisi girdi olarak kullanılmıştır. Bu yöntem ile standartlaştırılmamış faktör yükleri elde edilmiştir (McDonald, 1999). Varyans-kovaryans matrisinin köşegen elemanları ilgili maddelerin $V(X_i)$ terimlerinden oluşmaktadır (Ek 2). Gerçek puanlar varyansları ise maddeler arası kovaryans terimlerinden üretilmektedir.

$$V(T_i) = V(T_j) = \text{Kov}(X_i, X_j)$$

Doğrulayıcı faktör analizinde elde edilen faktör yükleri (λ_i) aynı zamanda kovaryans teriminin fonksiyonu olduğu için;

$$\text{Kov}(X_i, X_j) = \lambda_i \lambda_j$$

$$\text{Var}(T_i) = \lambda_i^2$$

yaklaşımlarıyla maddelere ilişkin gerçek puanlar varyansı elde edilmiştir (McDonald, 1999). Doğrulayıcı faktör analitik modelindeki her bir maddeye ilişkin doğrusal modellerin

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

hata terimlerinin varyansları, modellerin DFA yöntemiyle çözümlenmesiyle elde edilmiştir. Bu yaklaşımla $\text{Var}(X_i)$, $\text{Var}(T_i)$ ve $\text{Var}(E_i)$ terimleri elde edilmiştir ve bu modellerin determinasyon katsayıları (R^2) madde güvenilirlikleri olarak ele alınmıştır. DFA çözümlemesine ilişkin sonuçlar Ek 1'de verilmiştir.

Uygulamanın buraya kadar olan bölümünde KTK ve MYK modellerinden elde edilen madde güvenilirlikleri karşılaştırılmıştır. Uygulamanın 2. aşamasında ise test güvenilirlikleri ele alınmış ve KR_{20} ile MYK modelinden elde edilen güvenilirlik katsayıları karşılaştırılmıştır. Karşılaştırmalar; bir önceki aşamada ele alınan veri kümesinden 10 adet örneklem çekilerek elde edilen örneklem kümeleri üzerinden yapılmıştır. Örneklem kümeleri, basit rasgele örneklem yöntemi ile her biri 1000 öğrencinin madde yanıtlarından oluşturulmuştur.

BULGULAR

Bu çalışmada 2001 yılındaki OK-ÖSYS Türkçe alt testine ilişkin ölçme sonuçları kullanılmıştır. Bunun temel nedeni; 553108 öğrencinin Türkçe alt testinde yer alan 25 madde puan toplamlarından oluşan verilerin yaklaşık

normal dağılım göstermesidir (Yurdugül ve Aşkar, 2004). Dolayısıyla bu çalışmada normal dağılımlı veri kümesinin kullanımı aynı zamanda çalışmanın bir sınırlılığı olarak ifade edilebilir. Bu veri kümesine ilişkin betimsel bulgular Tablo 1’de verilmiştir.

Tablo 1: Türkçe alt testine ilişkin betimsel bulgular

Madde Sayısı	25
Öğrenci Sayısı	553108
Ortalama	13.62
Varyans	20.79
Çarpıklık	-0,04
Basıklık	-0,67
Güvenirlik (KR ₂₀)	0,77

Tablo 1’e görüldüğü gibi ele alınan ölçme kümesi normal dağılım göstermektedir. Ancak dağılımın yaklaşık simetrik olmasına karşın hafifçe basık bir dağılım olduğu görülmektedir. Öğrencilerin 25 maddeye verdikleri yanıtların KTK, MYK ve DFA modellerine göre çözümlenmesiyle elde edilen parametre değerleri Tablo 2’de verilmiştir. Bu değerler madde parametreleri, madde güçlük düzeyleri (π), maddelerin gözlenen puan varyansları [$V(X_i)$], gerçek puan varyansları [$V(T_i)$] ve hata puanları varyansları [$V(E_i)$]. Çalışmaya konu olan ρ_{Di} katsayısı ise Dimitrov eşitlikleri ile MYK modelinden elde edilen ve klasik madde güvenirliliğini, aynı şekilde ρ_i ise klasik madde analizinden elde edilen klasik madde güvenirliliklerini göstermektedir. R^2 ise DFA sonucunda elde edilen ölçme modelinin determinasyon katsayısını ifade etmektedir.

Tablo 2: Türkçe alt testine ilişkin MYK ve KTK modellerinin kestirimleri

Madde Yanıt Kuramı								Klasik Test Kuramı							
Parametreler				Dimitrov Tekniği				Parametreler			Doğrulayıcı Faktör Analizi				
No	a_j	b_i	π_i	$V(X_i)$	$V(E_i)$	$V(T_i)$	ρ_{Di}	p_i	r_i	ρ_i	λ_i	$V(X_i)$	$V(E_i)$	$V(T_i)$	R_i^2
1	0,41	0,33	0,45	0,25	0,22	0,03	0,12	0,45	0,35	0,17	0,16	0,25	0,23	0,03	0,10
2	0,87	-2,13	0,92	0,07	0,07	0,00	0,06	0,92	0,50	0,14	0,09	0,08	0,07	0,01	0,10
3	0,77	-1,17	0,76	0,18	0,14	0,04	0,20	0,76	0,52	0,22	0,19	0,18	0,14	0,04	0,20
4	0,25	2,55	0,27	0,20	0,19	0,01	0,05	0,26	0,21	0,09	0,07	0,19	0,18	0,00	0,03
5	0,48	0,98	0,34	0,22	0,19	0,03	0,13	0,33	0,40	0,19	0,17	0,22	0,19	0,03	0,13
6	0,33	0,21	0,47	0,25	0,23	0,02	0,09	0,47	0,29	0,14	0,12	0,25	0,23	0,01	0,06
7	0,83	-0,80	0,70	0,21	0,16	0,05	0,25	0,69	0,56	0,26	0,22	0,21	0,16	0,05	0,24
8	0,30	-0,45	0,55	0,25	0,23	0,02	0,08	0,55	0,26	0,13	0,12	0,25	0,24	0,01	0,06
9	0,40	-1,01	0,65	0,23	0,20	0,02	0,11	0,65	0,33	0,16	0,14	0,23	0,21	0,02	0,09
10	0,71	-0,53	0,62	0,24	0,18	0,05	0,23	0,62	0,51	0,25	0,23	0,24	0,19	0,05	0,19
11	0,51	0,61	0,39	0,24	0,20	0,04	0,16	0,39	0,41	0,20	0,18	0,24	0,21	0,03	0,13
12	0,41	0,75	0,39	0,24	0,21	0,03	0,12	0,38	0,35	0,17	0,14	0,24	0,22	0,02	0,09
13	0,70	-1,21	0,76	0,18	0,15	0,03	0,18	0,76	0,49	0,21	0,18	0,18	0,15	0,03	0,18
14	0,46	-0,82	0,63	0,23	0,20	0,03	0,13	0,64	0,37	0,18	0,16	0,23	0,20	0,03	0,11
15	0,31	2,50	0,23	0,18	0,17	0,01	0,06	0,23	0,26	0,11	0,09	0,17	0,16	0,01	0,05
16	0,12	0,87	0,46	0,25	0,25	0,00	0,00	0,46	0,11	0,05	0,05	0,25	0,25	0,00	0,01
17	0,75	0,32	0,42	0,24	0,18	0,06	0,25	0,42	0,54	0,27	0,24	0,24	0,18	0,06	0,18
18	0,66	-0,23	0,55	0,25	0,19	0,05	0,22	0,55	0,50	0,25	0,23	0,25	0,20	0,05	0,22
19	0,54	-1,67	0,79	0,17	0,15	0,02	0,11	0,79	0,40	0,16	0,13	0,17	0,15	0,02	0,10
20	0,43	-0,22	0,53	0,25	0,22	0,03	0,13	0,54	0,36	0,18	0,16	0,25	0,22	0,03	0,10
21	0,84	-0,92	0,72	0,20	0,15	0,05	0,24	0,72	0,56	0,25	0,21	0,20	0,16	0,04	0,22
22	0,63	0,35	0,43	0,24	0,19	0,05	0,21	0,42	0,48	0,24	0,21	0,24	0,20	0,04	0,18
23	0,80	-1,26	0,78	0,17	0,14	0,03	0,20	0,78	0,52	0,21	0,18	0,17	0,14	0,03	0,20
24	0,35	2,14	0,24	0,18	0,17	0,01	0,07	0,24	0,27	0,12	0,09	0,18	0,17	0,01	0,05
25	0,84	-0,46	0,62	0,24	0,17	0,07	0,12	0,61	0,57	0,28	0,25	0,24	0,18	0,06	0,27

Tablo 2'den görüleceği gibi MYK ve KTK modellerinden elde edilen madde güçlük indeksleri eşit çıkmaktadır. Eşitlik 8 ile MYK modelinden elde edilen π_i değerleri ile klasik madde analizinden elde edilen p_i değerleri arasında Tablo 3'te verildiği gibi 1,00 düzeyinde bir korelasyon elde edilmiştir. Buna göre Eşitlik 8 ile verilen π_i 'nin madde güçlük düzeyi olan p_i 'nin iyi bir kestiricisi olduğu ifade edilebilir.

Diğer taraftan Ek 2'de maddelerin varyans-kovaryans matrisi verilmiştir. Maddelerin varyans-kovaryans matrislerinin köşegen değerleri ilgili maddenin gözlenen puanlar varyansını $[V(X_i)]$ vermektedir. Varyans-kovaryans matrisi kullanılarak çözümlenen doğrulayıcı faktör analizi ile maddelerin gerçek puan varyansları $[V(T_i)]$, ve hata puan varyansları

$[V(E_i)]$, elde edilmiştir. KTK ve MYK'dan elde edilen maddelerin gözlenen puanlar varyansları arasındaki korelasyon 1,00 ve hata puanları varyansları arasındaki korelasyon ise 0,99 olarak hesaplanmıştır. Verilen karşılaştırmalar arasında en düşük korelasyon MYK modeline dayalı madde güvenilirlikleri ile KTK modelinden elde edilen madde güvenilirlik değerlerinde gözlenmiştir (0,91). Tablo 3'te, üç adet madde güvenilirlik kestiricileri (ρ_{Di} , ρ_i , R^2) arasındaki korelasyon değerleri verilmiştir.

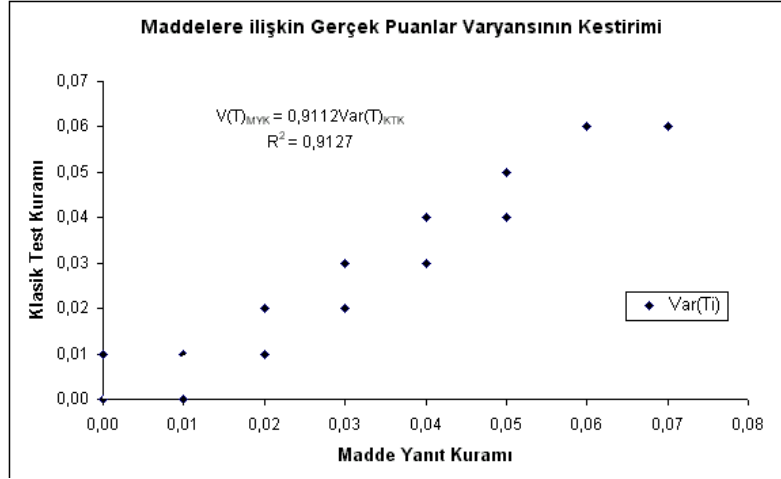
Tablo 3: Kestirim değerleri arasındaki korelasyonlar

Klasik Test Kuramı						
	π	V(X)	V(T)	V(E)	ρ_i	R^2
π	1,00					
V(X)*		1,00				
V(T)*			0,96			
V(E)*				0,99		
ρ_{Di}					0,91	0,95

* işaretli terimler Dimitrov tarafından geliştirilen eşitlikler yardımıyla MYK modelinden üretilmiştir.

Dimitrov eşitlikleri yardımı ile MYK modelinden elde edilen madde güvenilirlikleri ile doğrulayıcı faktör analizinden elde edilen madde güvenilirlikleri arasındaki korelasyon ise Tablo 3'te belirtildiği gibi 0,95 olarak elde edilmiştir.

Çizim 1'de KTK modelinden elde edilen maddelere ilişkin gerçek puanlar varyansları ile MYK modelinden elde edilen gerçek puanlar varyanslarının saçılım grafikleri verilmiştir. Gerçek puanlar varyansının her iki kestirim yöntemine göre elde edilmesinden hesaplanan regresyon denkleminin yuvarlatılmış biçimi $V(T)_{MYK}=0,91 \times V(T)_{KTK}$ şeklindedir. Buna göre MYK modelinden elde edilen gerçek puanlar varyansı, KTK modelinden elde edilen ve DFA ile kestirilen gerçek puanlar varyansının bir yordayıcısı olduğu söylenebilir.



Çizim 1: Maddelere ilişkin gerçek puanlar varyansının saçılım grafiği

Buraya kadar yapılan karşılaştırmalar maddelere ilişkin model terimlerinin varyansları ve madde güvenilirliklerini kapsamakta idi. Bu aşamada ise, testi oluşturan maddeler üzerinden elde edilen test güvenilirlikleri karşılaştırılmıştır. Bu karşılaştırmalar için bir önceki aşamada kullanılan ölçme kümesinden 1000 adet gözlem içeren basit rasgele örnekleme yöntemi ile 10 adet örneklem çekilmiştir. Her bir örneklemde elde edilen KTK modeline dayalı KR_{20} güvenilirlik katsayısı ile MYK modelinden elde edilen (Eşitlik 18) test güvenilirlikleri (ρ_D) karşılaştırılmıştır. Karşılaştırma sonuçlarının da klasik güvenilirlik olan KR_{20} ile ρ_D güvenilirlik kestirimleri yaklaşık sonuçlar vermiştir. Bu sonuçlara ilişkin kestirim değerleri Tablo 4’te verilmiştir. Tablo 4’te yer alan güvenilirlik değerlerine göre ρ_D , KR_{20} güvenilirlik değerlerinden ortalama 0,03 kadar düşük değerde elde edilmektedir. Ancak daha çok tekrar sayılı yüksek düzeyli simülasyon çalışmaları ile daha kesin sonuçlar elde edilebilir.

Tablo 4: Örneklemelere ilişkin test güvenilirlikleri

Örneklem	ρ_D	KR_{20}	Fark
1	0,78	0,80	0,02
2	0,77	0,80	0,03
3	0,77	0,80	0,03
4	0,77	0,80	0,03
5	0,77	0,80	0,03
6	0,75	0,78	0,03
7	0,78	0,80	0,02
8	0,77	0,80	0,03
9	0,78	0,80	0,02
10	0,78	0,80	0,02
		Ortalama	0,03

Tablo 4'teki değerlere göre MYK modelinden elde edilen ρ_D ile KR_{20} güvenilirlik değerleri arasındaki farklılık değerlerinde bir sistematik olduğu görülmektedir. Çalışma kapsamında ele alınan 10 örneklem kümesine göre bu farklılık ortalama 0,03 değeri civarındadır.

Her iki güvenilirlik kestirim değerlerine yönelik bir genelleme yapabilmek için çok sayıda örneklem üzerinde çalışmakla olanaklıdır. Bu çalışmada kuramsal bir genellemeye yapılması yerine güvenilirlik katsayılarının basit karşılaştırmaları yapılmıştır. Bu nedenle, 10 adet örneklem üzerinde yapılan çalışma ile yalnızca ρ_D güvenilirlik kestiriminin KR_{20} değerinden büyük elde edildiği sonucu rapor edilmiştir.

SONUÇ VE ÖNERİLER

Bilindiği gibi güvenilirlik kavramı KTK modelinin terimleri ile ifade edilmektedir ve bu tür güvenilirlikler "klasik güvenilirlik" olarak adlandırılmaktadır (Dimitrov, 2002). MYK'nda ise madde ve test güvenilirliklerini elde etmek için bilgi fonksiyonlarından yararlanılmaktadır. Ancak Doran (2005) bilgi fonksiyonundan elde edilen güvenilirliklerin klasik güvenilirlikten farklı olduğunu ifade etmektedir. Buna karşılık; Dimitrov (2003a, 2003b), MYK modelinin parametrelerini kullanarak KTK terimlerinin varyanslarının elde edilebileceğini çeşitli eşitlikler yardımıyla göstermiştir. Bu çalışmada ilgili terimlerin varyansları kullanılarak MYK modelinden klasik madde ve test güvenilirlikleri elde edilmiştir. Elde edilen bu eşitlikler klasik madde analizine dayalı madde güvenilirlikleri ve DFA'nde yer alan ölçme modellerinin determinasyon katsayısı ile yüksek korelasyon vermektedir. Aynı şekilde Dimitrov eşitlikleri ile elde edilen klasik test güvenilirliği ile KR_{20} güvenilirlik katsayısı yaklaşık eşit değerler üretmektedir. Böylelikle MYK'ndan elde edilen klasik güvenilirlik katsayıları, KTK'ndan elde edilen güvenilirliklerin bir kestirici olduğu ifade edilebilir. Bu yaklaşım, MYK'nda kullanılan bilgi fonksiyonlarına dayalı güvenilirliklerin bir alternatifi olarak görülebilir.

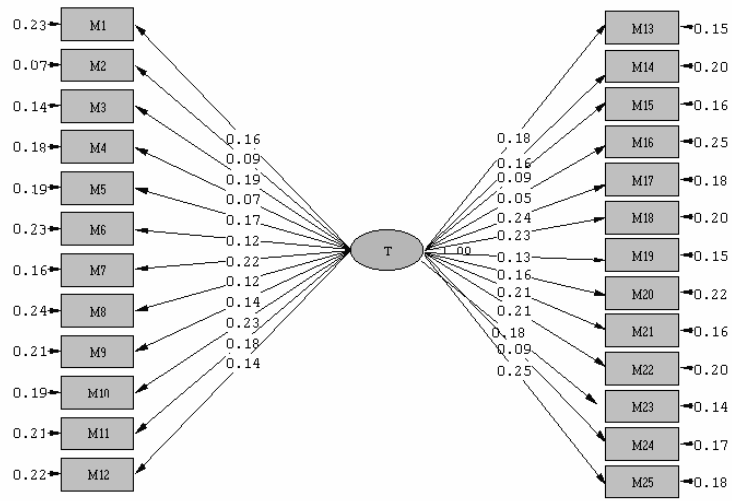
Bu konuda göz önünde tutulması gereken bir diğer konu ise; KTK ve MYK modelleri arasındaki bağıntı, Hambleton ve Swaminathan'ın (1985) belirttiği gibi ölçme kümelerinin normal dağılım gösterdiği durumlarda geçerlidir. Ancak bu durumda KTK modelindeki madde parametrelerinin MYK modelinden elde edilebilirliği söz konusudur. Diğer yandan bu çalışmada kullanılan veri kümesi betimsel bulgularda belirtildiği gibi yaklaşık olarak doğrusal dağılım göstermektedir. Üzerinde çalışılan veri kümesinin normal dağılımdan uzaklaştığı durumlarda MYK modelinden elde edilen klasik güvenilirliklerin davranışı farklı bir çalışma kapsamında ele alınabilir.

KAYNAKÇA

- Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme*, Ankara: ÖSYM Yayınları.
- Dimitrov, D. M. (2002) Reliability: Arguments for multiple perspectives and potential problems with generalization across studies. *Educational and Psychological Measurement* (62), 783-801.
- Dimitrov, D. M. (2003a) Marginal True-Score Measures and Reliability for Binary Items as a Function of Their IRT Parameters. *Applied Psychological Measurement*, 27(6), 440-58.
- Dimitrov, D. M. (2003b). Reliability and true-score measures of binary items as a function of their Rasch difficulty parameter. *Journal of Applied Measurement*, 4(3), 222-233.
- Doran, H. C. (2005). The Information Function for the One-Parameter Logistic Model: Is it Reliability? *Educational and Psychological Measurement*, 65(5), 665-675.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston: Kluwer-Nijhoff Publishing.
- Kelecioğlu, H. (2001). Örtük özellikler teorisindeki b ve a parametreleri ile klâsik test teorisindeki p ve r istatistikleri arasındaki ilişki, *Hacettepe Üniversitesi, Eğitim Fakültesi Dergisi*, 20, 104-110.
- Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. New York: Methuen.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Mahwah, NJ: Erlbaum
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah NJ: Erlbaum.
- Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.
- Traub, R. E. (1994) *Reliability for the Social Sciences: Theory and Applications*. Sage, Thousand Oaks, CA.
- Yurdugül, H ve Aşkar, P. (2004). Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı'nın cinsiyete göre madde yanlılığı açısından incelenmesi. *Eğitim Bilimleri ve Uygulama Dergisi*, 3(5), 3-20

EK 1:

Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı (OK-ÖSYS; 2001) Türkçe alt testinde yer alan 25 maddeye ilişkin 553108 adet öğrencinin yanıtlarından elde edilen doğrulayıcı faktör analitik çözümü.



Chi-Square=105011.13, df=275, P-value=0.00000, RMSEA=0.026

EK 2: Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı (OK-ÖSYS, 2001) Türkçe Alt Testinde Yer Alan 25 Maddeye İlişkin 553108 Adet Öğrencinin Yanıtlarından Elde Edilen Varyans-Kovaryans Matrisi.

MADDELER																									
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
1	0.25																								
2	0.02	0.08																							
3	0.03	0.03	0.18																						
4	0.01	0.00	0.01	0.19																					
5	0.03	0.01	0.03	0.01	0.22																				
6	0.02	0.01	0.02	0.01	0.02	0.25																			
7	0.03	0.02	0.04	0.02	0.04	0.03	0.21																		
8	0.02	0.01	0.02	0.01	0.02	0.02	0.03	0.25																	
9	0.02	0.02	0.03	0.01	0.02	0.01	0.03	0.02	0.23																
10	0.04	0.02	0.04	0.02	0.04	0.03	0.05	0.03	0.04	0.24															
11	0.03	0.01	0.03	0.02	0.03	0.02	0.04	0.02	0.02	0.04	0.24														
12	0.02	0.01	0.02	0.01	0.03	0.02	0.03	0.02	0.02	0.03	0.03	0.24													
13	0.03	0.02	0.04	0.01	0.03	0.02	0.04	0.02	0.03	0.04	0.03	0.02	0.18												
14	0.02	0.02	0.03	0.01	0.02	0.02	0.04	0.02	0.02	0.04	0.03	0.03	0.02	0.23											
15	0.01	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.01	0.02	0.02	0.01	0.02	0.17											
16	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.25										
17	0.04	0.02	0.04	0.02	0.04	0.03	0.05	0.03	0.03	0.06	0.05	0.04	0.04	0.02	0.01	0.24									
18	0.03	0.02	0.04	0.02	0.04	0.03	0.05	0.03	0.03	0.05	0.04	0.04	0.03	0.02	0.01	0.06	0.25								
19	0.02	0.01	0.02	0.01	0.02	0.01	0.03	0.01	0.02	0.03	0.02	0.02	0.01	0.01	0.03	0.03	0.17								
20	0.02	0.01	0.03	0.01	0.03	0.02	0.03	0.02	0.02	0.04	0.03	0.02	0.03	0.01	0.01	0.04	0.03	0.25							
21	0.03	0.02	0.04	0.01	0.03	0.03	0.05	0.03	0.03	0.05	0.03	0.04	0.03	0.02	0.01	0.05	0.06	0.03	0.04	0.20					
22	0.03	0.01	0.04	0.02	0.04	0.03	0.05	0.02	0.03	0.05	0.04	0.03	0.03	0.02	0.01	0.05	0.05	0.03	0.03	0.04	0.24				
23	0.03	0.02	0.04	0.01	0.03	0.02	0.04	0.02	0.03	0.04	0.03	0.02	0.04	0.03	0.01	0.04	0.04	0.03	0.03	0.04	0.04	0.17			
24	0.02	0.00	0.01	0.02	0.02	0.02	0.02	0.01	0.01	0.02	0.03	0.02	0.01	0.02	0.01	0.00	0.03	0.02	0.01	0.01	0.03	0.00	0.18		
25	0.04	0.02	0.05	0.02	0.04	0.03	0.06	0.03	0.03	0.05	0.05	0.04	0.05	0.04	0.02	0.01	0.06	0.06	0.03	0.04	0.05	0.05	0.03	0.24	