Ankara University, Journal of Faculty of Educational Sciences, year: 2005, vol: 38, no: 2, 21-47

The Congeneric Test Theory and The Congeneric Item Analysis:

An Application for Unidimensional Multiple Choice Tests

Halil YURDUGÜL*

Abstract: Congeneric test theory which is developed by Jöreskog is constructed on congeneric measurements. The congeneric test theory is a subtheory of classical test theory and its model includes item parameters such as item-response theory. In this study, firstly, congeneric test theory and congeneric item analysis were introduced. Then, the results of classical item analysis and congeneric item analysis for multiple choice data were compared.

Key Words: Classical test theory, congeneric test theory, item-response theory, item analysis

SUMMARY

The congeneric measures and congeneric test model were described by Jöreskog (1971). However, "congeneric test theory" as an expression, was first used by Alwin (1976) and Lucke (2005a). The congeneric test theory is a subtheory of classical test theory.

Classical test theory based on relation between the observed score (X_i), and true score (T_i) and error score (E_i).

$$X_i=T_i+E_i$$

* Dr. Hacettepe University, Faculty of Education, yurdugul@hacettepe.edu.tr

On other hand, the congeneric test model includes item parameters such as item difficulties and item discriminations.

$$X_i=b_i+a_iT_i+E_i$$

In this equation, a_i denotes level of item discrimination of *i*th item and b_i denotes item difficulty. The analysis of the model and obtaining of model's parameters is named "congeneric item analysis".

In this study, congeneric item analysis and classical item analysis were used in an aptitude test to determine the qualification of items. The test is called *The 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, and natural sciences. Each subtest consists of 25 items. In this study, only the data of the math subtest were used. The data gathered from 553108 pupils and SSPE-SE carried out in 2001 were utilized. The psychometric properties of items in the math test mentioned above were obtained through classical and congeneric item analysis. These properties are item difficulties, item discriminations, and item reliabilities. In addition, the reliability of whole test was calculated by using KR_{20} and McDonald's ω.

The item discrimination values and item reliabilities obtained from congeneric analysis were higher than the results of classical item analysis. The reason for this result is that unstandardized solutions are used in congeneric item analysis. On the other hand, the correlation between discrimination values in congeneric and classical analysis was found as 0.99. This correlation points out the relation between discrimination values.

The reliability of test was found as KR_{20}=0.81 and ω=0.82. These values are found to be very close. Because the deviation of variance-covariance found as 0.02, the items in test are approximately essentially tau equivalent measures. As known, the KR_{20} and ω coefficients are equal when the items are essentially tau equivalent measures.

According to results, the item parameters obtained from classical item analysis were similar to the results from congeneric item analysis. The matrix of variance-covariance was investigated and found that the items are approximately essentially tau equivalent. In addition to this, item reliabilities obtained differently according to both of models. The item reliabilities and item discrimination values obtained from classical item analysis was found to be higher more than the item reliabilities and the item discrimination values obtained from congeneric item analysis.

Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, yıl: 2005, cilt: 38, sayı: 2, 21-48

Konjenerik Test Kuramı ve Konjenerik Madde Analizi: Tekboyutlu Çoktan Seçmeli Testler Üzerine Bir Uygulama

Halil Yurdugül*

Öz: Konjenerik test kuramı, Jöreskog tarafından geliştirilen ve konjenerik ölçmeler üzerine kurulu olan bir yaklaşımdır. Klasik test kuramının bir alt kuramı olan konjenerik test kuramı, testi oluşturan maddelerin karakteristik özelliklerini de içerdiğinden dolayı madde-yanıt kuramı ile benzerlik göstermektedir. Bu çalışmada konjenerik test kuramı ve madde analizi tanıtılmış aynı zamanda çoktan seçmeli bir test üzerinden klasik madde analizi ve konjenerik madde analizi sonuçları karşılaştırılarak tartışılmıştır.

Anahtar Kelimeler: Klasik test kuramı, konjenerik test kuramı, madde-yanıt kuramı, madde analizi

GİRİŞ

Eğitim ve psikolojide yer alan deneysel araştırmalarda gözlenen ile gözlenemeyen özellikler arasındaki bağıntılar temel olarak doğrusal ve doğrusal olmayan modeller ile açıklanmaktadır. Bunlardan *doğrusal model*, klasik test kuramı (classical testing theory) tarafından kullanılan doğrusal regresif fonksiyonu içermektedir (Lord ve Novick, 1968).

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

$$X_i=T_i+E_i \qquad (1)$$

Burada X_i, i. madde puanı, T_i, maddeye ilişkin gerçek puan ve E_i ise modelin hata puanıdır. Gözlenen ile gözlenemeyen puanlar arasındaki *doğrusal olmayan model* ise madde-yanıt kuramı (item-response theory) tarafından kullanılan lojistik fonksiyon ile ifade edilir.

$$P(\theta)=\frac{e^{a_i(\theta-b_i)}}{1+e^{a_i(\theta-b_i)}} \qquad (2)$$

Burada θ, ölçülmesi amaçlanmış özelliğin düzeyini, a_i maddenin ayırıcılık düzeyini, b_i maddenin güçlük düzeyini ve $P(\theta)$ ise maddeyi doğru yanıtlama olasılığını göstermektedir (Lord, 1980).

Klasik test kuramının (KTK) ve madde-yanıt kuramının (MYK) modellerindeki terimler karşılaştırıldığı zaman MYK modelinde bireyin maddeyi doğru yanıtlama olasılığının bireyin özelliklerine (θ) ve maddenin karakteristik özelliklerine (a_i, b_i) bağlı olduğu görülmektedir. KTK modelinde ise madde özelliklerine ilişkin terimler yer almaz. McDonald (1999) bu durumu KTK modelinin birleşik testlerdeki (composite test) madde puanlarının toplamları üzerinden elde edilmesi ve testi oluşturan maddelerin paralel ve/veya eşdeğer madde yapısına uygun olmasından kaynaklandığını belirtmektedir. Bu nedenle maddelerin yapısı, ölçme kuramında önemli bir yere sahiptir. Birleşik testte yer alan maddeler ölçülmek istenilen yapıyı eşit büyüklükte ve eşit duyarlıkta (hata varyansları eşit) ölçüyor ise *paralel* maddeler olarak tanımlanır. Eğer maddeler ölçülmek istenilen yapıyı eşit büyüklükte ancak eşit olmayan duyarlıkta ölçüyor ise bu tür maddeler için *eşdeğer* maddeler tanımlaması yapılır (Adamson vd., 2000; Alwin, 1976; Novick ve Lewis, 1967; Raykov, 1997).

Jöreskog (1971) ve Alwin (1976) maddelerin ölçmeye yöneldiği yapıyı eşit olmayan büyüklük ve duyarlıkta ölçmesi durumunda bu tür madde için *konjenerik* madde ifadesini kullanmıştır. Konjenerik maddeler için KTK modeli yeterli olamayacağından dolayı KTK modelinin bir alt modeli olan konjenerik test modeli geliştirilmiştir (Lucke, 2005a, 2005b; Raykov, 1997, 2001). Konjenerik test kuramının (KonTK) modeli;

$$X_i=b_i+a_iT+E_i \qquad (3)$$

şeklindedir. Burada b_i; i. madde ortalamasını/güçlüğünü ve a_i, maddenin ayırıcılık gücünü göstermektedir (Alwin, 1976; Jöreskog, 1971, Jöreskog ve Sörbom, 1996; Lucke, 2005b, McDonald, 1999).

KTK, paralel ya da eşdeğer ölçmeler üzerine kurulu bir kuramdır. KonTK ise doğrulayıcı faktör analizi üzerine kuruludur ve konjenerik ölçmelerin çözümlenmesinde kullanılmaktadır (DeVellis, 2003).

Bu çalışmada KTK ve KonTK'nın özellikleri tanıtılmış ve bu kuramlarda yer alan madde analizlerinin yapıları ile çalışma ilkeleri karşılaştırılmıştır. Çalışmanın uygulama bölümünde ise KTK ve KonTK madde analizleri 2001 Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı'nda yer alan matematik alt testindeki maddelere uygulanmış ve sonuçlar tartışılmıştır.

Klasik Test Kuramı

Eğitimsel araştırmalarda kullanılan ölçme sonuçlarının çözümlenmesi için yaygın bir şekilde klasik test kuramından yararlanılmaktadır. KTK modeli, i. madde ile ölçülmek istenilen yapı (T_i) ve ilgili maddeden elde edilen ölçme/gözlem sonuçları (X_i) arasındaki bağıntıyı Eşitlik 1 ile verilen fonksiyon ile açıklamaktadır. Bu modelin varsayımları;

$$\varepsilon(E_i)=0$$

$$Kov(T,E_i)=0$$

$$Kov(E_i,E_j)=0$$

şeklinde sıralanmaktadır. Burada $\varepsilon(\ )$ beklenen değeri ve Kov(.,.) ise belirtilen terimler arasındaki kovaryansı göstermektedir (Lord ve Novick, 1968). Bu varsayımlar içerisinde hata terimlerinin ilişkisiz olduğuna ilişkin varsayımın [$Kov(E_i,E_j)=0$] bozulumu modelin geçerliğini düşüren bir etkendir (Lord ve Novick, 1968; Novick ve Lewis, 1967; Raykov, 2001). Bu varsayım, teste yer alan maddelerin paralel, eşdeğer ya da eşbiçimli yapıya sahip olduklarında sağlanır. Ancak maddeler konjenerik ölçme yapısına sahip olduğunda ise bu varsayım bozulmaktadır [$Kov(E_i,E_j)\neq 0$] (Alwin, 1976; Jöreskog, 1971; Lucke 2005a; Raykov, 1997).

Daha önce tanımları verilen paralel, eşdeğer, eşbiçimli ve konjenerik madde yapılarının özellikleri istatistiksel olarak KTK modeli üzerinde de ifade edilebilir. Buna göre; herhangi iki maddeye ilişkin KTK modelleri,

$$X_i=T_i+E_i$$

$$X_j=T_j+E_j$$

şeklinde verildiğinde, *paralel maddeler* $T_i=T_j$ ve $Var(E_i)=Var(E_j)$ özellikleri gösterir. Bu özelliklerinden dolayı iki madde aynı özelliği ölçmeye yönelmiş ve hata dağılımları eşittir. Ancak *eşdeğer* maddeler (tau-equivalent measurement) için hata terimleri varyansı eşit değildir [$T_i=T_j$,

$Var(E_i) \neq Var(E_j)$] ve tanım gereği paralel ve eşdeğer maddeler aynı yapıyı eşit büyüklükte ölçerler (Adamson vd, 2000; Alwin, 1976; Novick ve Lewis, 1967; Traub, 1994). Eşdeğer ölçmelerin özel bir biçimi olan *eşbiçimli* maddeler (essentially tau-equivalent measurement) ise her iki maddenin aynı yapıyı eşit büyüklükte ölçmeye yönelmiş olmasına karşılık maddelerin gerçek puanları arasında *b* gibi bir sabit olduğu varsayılır (McDonald, 1999; Raykov, 1997; Traub; 1994).

$$T_i=b_{ij}+T_j$$

Bununla birlikte her üç madde yapısının ortak özelliği ise maddeler arası kovaryans değerlerinin birbirine eşit olmasıdır ve bu özellik sayesinde KTK modelindeki "hataların ilişkisiz olma" varsayımını sağlamaktadır (McDonald, 1999). Ancak maddelerin ölçmeye çalıştıkları gerçek puanlar arasında a_i gibi bir çarpan var ise bu durumda maddeler için konjenerik (congeneric) tanımlaması yapılmaktadır (Alwin, 1976; Jöreskog, 1971; Traub, 1994).

$$T_i=b_{ij}+a_{ij}T_j$$

Konjenerik maddeler için KTK modelinin "hataların ilişkisiz olma" varsayımı bozulur çünkü maddeler arası kovaryans terimleri eşit değildir (Alwin, 1976; Lucke, 2005b; McDonald, 1999). Maddelerin yapısına ilişkin istatistiksel özellikler Ek 1'de verilmiştir.

Klasik Test Kuramında Gerçek Puanların Kestirilmesi

KTK uygulamalarda yaygın bir şekilde kullanılmasına karşın en büyük olumsuzluğu gözlenen puanlardan gerçek puanları kestirirken, testi oluşturan maddelerin karakteristik özelliklerinden (madde güçlüğü ve madde ayırıcılık gücü) bağımsız olarak ele alınmasıdır. MYK'nın lojistik modelinde yer alan *maddenin karakteristik özellikleri* aynı zamanda öğrencilerin yetenek ya da bilgi düzeylerinin kestirilmesinde birer etken olarak görev yapmaktadır (Hambleton ve Swaminathan, 1985; Lord,1980).

Eşitlik 1 ile verilen KTK modeli tek bir maddeye/ölçmeye ilişkin bir modeldir ve birleşik ölçmelerdeki karşılığı ise, k (k>1) madde sayısını, X madde puanlarının toplamını ($X=X_1+X_2+...+X_k$) göstermek üzere;

$$X_1=T_1+E_1$$

$$X_2=T_2+E_2$$

ve $T_1=T_2=...=T_k$ olmak üzere,

$$X=T+E \qquad (4)$$

şeklindedir. McDonald (1999; Sayfa:131)'a göre, ölçme aracındaki maddeler paralel ise bu durumda gözlenen puanlar aynı zamanda gerçek puanların kestiricisi olarak kullanılabilir.

$$X=E(T)$$

Klasik Güvenirlik

Gerçek puanların hipotetik bir yapıda olması ve doğrudan elde edilememesine karşın gerçek puanlar varyansı kolaylıkla elde edilebilir. Çünkü gerçek puanlar varyansının kestiricisi olarak maddeler arası kovaryans terimleri kullanılmaktadır (McDonald, 1999).

$$\text{Var(T)=Kov}(X_i,X_j) \qquad (5)$$

KTK'nda gerçek puanlar varyansının maddeler arası kovaryanslar ile açıklanması (McDonald, 1999, sayfa: 79) aynı zamanda güvenirlik kavramı için bir zemin oluşturmaktadır. Bilindiği gibi güvenirlik kavramı gerçek puanlar varyansının gözlenen puanlar varyansına oranı olarak ifade edilmektedir.

$$\text{Güvenirlik}=\frac{Var(T)}{Var(X)} \qquad (6)$$

KTK'nda maddelerin paralel, eşdeğer, eşbiçimli ve konjenerik olmalarına göre içtutarlık anlamındaki güvenirlik katsayılarının elde edilmesi de farklılık göstermektedir. Bu durumda eğer birleşik testi oluşturan maddeler paralel ise maddeler arası korelasyon aynı zamanda güvenirliğe karşılık gelmektedir (Lord ve Novick; 1968, Traub, 1994).

$$\rho(X_i,X_j)=\frac{\mathrm{Kov}(X_i,X_j)}{\sqrt{\mathrm{Var}(X_i)}\sqrt{\mathrm{Var}(X_j)}}=\frac{\mathrm{Var(T)}}{\mathrm{Var(X)}}$$

Eğer birleşik testi oluşturan maddeler eşdeğer ya da eşbiçimli ise bu durumda maddelerin hata puanları varyansı eşit olmayacağından dolayı güvenirlik katsayılarını elde edebilmek için korelasyon terimleri yerine maddeler arası kovaryans terimlerinden yararlanılır. Buna göre birleşik testin gerçek puanlar varyansını elde edebilmek için k madde için;

$$\binom{k}{2}=\frac{k(k-1)}{2}$$

adet ikili maddenin kovaryanslarının ortalaması alınır (Cronbach, 1951).

$$\mathrm{Var(T)} = \frac{1}{\mathrm{k(k-1)/2}} \sum_{\mathrm{i<j}} \mathrm{Kov}(\mathrm{X_i}, \mathrm{X_j})$$

Var(T) ifadesi Eşitlik 6'da yerine konulursa;

$$\text{Güvenirlik}=\frac{k}{k-1}\left[1-\frac{\sum_{i=1}^{k} Var(X_i)}{Var(X)}\right] \qquad (7)$$

elde edilir. Bu ifade aynı zamanda Cronbach (1951) tarafından geliştirilen α katsayına karşılık gelmektedir. Çoktan seçmeli testlerde maddeler (0 ve 1) şeklinde puanlandığından dolayı madde varyansları binominal dağılımın parametresi [$Var(X_i)=p_i(1-p_i)$] olarak alınmaktadır. Bu durumda ise güvenirlik katsayısı için α yerine KR_{20} kullanılır (Lord ve Novick; 1968).

$$KR_{20}=\frac{k}{k-1}\left[1-\frac{\sum_{i=1}^{k} p_i(1-p_i)}{Var(X)}\right] \qquad (8)$$

Eşitlik 7 ve Eşitlik 8 ile verilen güvenirlik katsayıları paralel, eşdeğer ve eşbiçimli ölçmeler için gerçek güvenirliği, aksi halde gerçek güvenirliğin alt sınırını vermektedir. Bu ölçmelerin ortak özellikleri ise maddeler arası kovaryansların eşit olmasıdır. Konjenerik ölçmelerde ise maddeler arası kovaryanslar eşit değildir. Bu tür ölçmeler için McDonald tarafından geliştirlen omega katsayısı daha uygundur (Lucke, 2005a; McDonald, 1985; Zinbarg vd., 2005).

KTK'nda *maddelerin güvenirlikleri* ise maddenin ayırıcılık düzeyi ve maddenin standart sapmasının (σ_i) bir fonksiyonudur.

$$\rho_i = r_{çs}\sigma_i$$

Burada $r_{çs}$, madde ayırıcılık gücünün kestiricisi olarak çift serili korelasyonu göstermektedir (Baykul, 2000).

Klasik Madde Analizi

Klasik madde analizi, ölçme aracını oluşturan maddelerin istatistiksel özelliklerini ortaya koymak için yapılan çalışmalardır. Bu anlamda madde analizinde ele alınan iki özellik ön plana çıkmaktadır. Bunlar sırasıyla; madde güçlük düzeyi (p) ve madde ayırıcılık düzeyidir (r). Madde güçlük düzeyi aynı zamanda *n* adet öğrencinin maddeye verdikleri yanıtların ortalamasını gösterir. Madde ayırıcılık düzeyi ise maddenin ölçmeye yöneldiği özelliğe sahip olan ve olmayan öğrencileri ayırtedebilmek amacıyla kullanılan madde-test korelasyonudur. Bu korelasyonları elde edebilmek için özellikle çoktan seçmeli testlerde çift-serili ya da nokta çift-serili korelasyon ifadeleri kullanılır (Baykul, 2000).

Klasik test kuramına dayalı çözümlemelerde, maddelerin istatistiksel özellikleri ya da bu özellikleri belirlemeye yönelik madde analizleri KTK modelini kestirmek amaçlı değil, test geliştirme amacıyla kullanılmaktadır.

Konjenerik Test Kuramı

Konjenerik test kuramı, klasik test kuramının bir alternatifi olmadığı gibi aksine KonTK, KTK'nın bir alt kuramıdır (Lucke, 2005a). Konjenerik kavramına ilişkin ilk tanımlamalar Jöreskog (1971) ve Alwin (1976) tarafından yapılmıştır. Buna göre birleşik testi oluşturan maddelerin her birinin ilgili yapıyı aynı büyüklükte ve aynı duyarlıkta (madde varyanslarının eşit olma durumu) ölçmesi beklenemez. Bu durum maddeler arası kovaryans terimlerinin eşit olmamasıyla açıklanır.

KonTK'ndaki yaklaşıma göre; her bir maddenin sahip olduğu gerçek puanlar (T_i) ve testin tamamı ile ölçülmek istenilen gerçek puan (T) arasında Eşitlik 9'daki gibi doğrusal bir bağıntı vardır (Alwin, 1976; Lucke, 2005b; McDonald, 1999; Traub, 1994).

$$T_i=b_i+a_iT \qquad (9)$$

Eşitlik 9 ile verilen ifade, Eşitlik 1 ile verilen KTK modelinde yerine konulursa KonTK modeline ulaşılır.

$$X_i=b_i+a_iT_i+E_i \qquad (10)$$

Bu eşitlikte b_i ile ifade edilen madde güçlük düzeyi, madde ortalamasını gösterdiğinden dolayı Eşitlik 10 aynı zamanda standartlaştırılmamış faktör analitik modeline

$$X_i=\mu_i+\lambda_i F+E_i \qquad (11)$$

karşılık gelmektedir (Jöreskog, 1971; McDonald, 1985; 1999). Ancak SPSS ve Statistica gibi paket programlar tarafından kullanılan açıklayıcı (exploratory) faktör analizi standartlaştırılmış faktör çözümleri üretmektedir.

$$Z_i=\lambda_i F+E_i$$

Burada Z_i, i. maddeye ilişkin standartlaştırılmış gözlenen puanı göstermektedir. Z ifadesi maddeye ilişkin gözlenen puanlardan madde ortalamalarının çıkartılması ile elde edilmektedir (Lord ve Novick, 1968; Jöreskog, 1971; McDonald, 1985, 1999).

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

Madde ortalamaları çözümlenecek modele dahil edilirse Eşitlik 11'e ulaşılmış olur.

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

Eşitlik 10 ve Eşitlik 11 karşılaştırıldığında madde ortalamaları madde güçlük düzeyine karşılık gelmektedir. F sembolü ile gösterilen ortak faktör (common factor) değerleri gerçek puanların bir kestiricisi ve faktör yükleri (λ) ise madde ayırıcılık düzeyini gösterir (McDonald, 1985, 1999; Raykov ve Penev, 1997).

Eşitlik 11'deki ifadenin çözümlenebilmesi için, bir diğer ifade ile standartlaştırılmamış faktör yüklerini elde edebilmek için doğrulayıcı (confirmatory) faktör çözümlemesine (Jöreskog ve Sörbom, 1996; McDonald, 1999; Zinbarg vd, 2005) ve bu çözümlemeleri yapmak için ise LISREL, AMOS gibi bilgisayar programlarına ihtiyaç vardır. Jöreskog (1971), açıklayıcı faktör analizinin ençok olabilirlik (maximum likelihood) faktör yöntemini geliştirerek doğrulayıcı faktör analizini önermiştir.

$$\tau=\Lambda\xi+\psi^2 \qquad (12)$$

Burada, Λ standartlaştırılmamış faktör yüklerini, ξ ölçülmeye çalışılan özelliği (attribute) bir başka ifade ile ortak faktörleri ve ψ^2 ise modelin hata varyansını (tekil faktörleri-unique factors) göstermektedir. Eşitlik 12 ile verilen bağıntıdaki faktör modeli, kovaryans terimlerinden üretilen standartlaştırılmamış faktör terimlerini içermektedir ve bu ifadenin açılımı ise k madde sayısı ve m ise ölçeğin alt boyut sayısını göstermek üzere,

$$\tau = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ : \\ \lambda_k \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ : \\ \xi_m \end{pmatrix} + \psi^2 \qquad (13)$$

şeklindedir. Bu durumda m=1 olduğunda bu bağıntı tek boyutlu yapıları ifade etmektedir (Jöreskog, 1971; Lucke, 2005a, 2005b; McDonald, 1985). Bu ifadeye ilişkin faktör analitik çözüm şeması Çizim 1'de verilmiştir.

Çizim 1 ve Eşitlik 12 ile verilen ifadelerde, doğrulayıcı faktör analitik modeline göre birleşik testlerde yer alan k adet madde için $\lambda_1=\lambda_2=...=\lambda_k$ ve $\psi_1=\psi_2=...=\psi_k$ eşitlikleri söz konusu olduğunda *paralel ölçme*, $\lambda_1=\lambda_2=...=\lambda_k$ ve $\psi_1\neq\psi_2\neq...\neq\psi_k$ ise k madde için eşdeğer ya da eşbiçimli ölçme[2] ve $\lambda_1\neq\lambda_2\neq...\neq\lambda_k$ olduğunda ise maddeler için konjenerik ölçme olduğu ifade edilir (Luke, 2005a, 2005b; McDonald, 1999).

Konjenerik Test Kuramında Gerçek Puanların Kestirilmesi

Ölçme sonuçlarının çözümlenmesine ilişkin kurulan modellerin temel amacı, gözlenen puanlardan yararlanılarak gözlenemeyen özelliklerin ya da teste yönelik gerçek puanların kestirilmesidir. MYK'nda gözlenemeyen özelliklerin kestirimi, Eşitlik 2 ile verilen ifadenin ençok olabilirlik (maximum likelihood) ya da Bayesci kestirim (Bayesian estimation) yöntemleri ile olanaklıdır (Hambleton ve Swaminathan, 1985; Lord, 1980).

MYK'na dayalı çözümleme yapan bilgisayar paket programlarında ölçmeye konu olan özelliğin kestirilmesinde genellikle ençok olabilirlik kestirimi kullanılmaktadır. Buna göre öğrencilerin k adet maddeye verdikleri yanıtlardan $[u=(u_1,...,u_k)]$ türetilen olabilirlik fonksiyonunu

$$L(\theta)=P_1(\theta)...P_k(\theta)$$

enbüyükleyen (maximized) θ düzeyini bulma algoritmasını içeren Newton-Raphson iterasyon yöntemi ile kestirilir. Diğer taraftan öğrencilerin uygulanan teste ilişkin gerçek puanları, her bir maddeye ilişkin maddeyi doğru yanıtlama olasılıklarının toplamı olarak ifade edilir (Hambleton ve Swaminathan, 1985).

[2] Faktör modelindeki eşdeğer ve eşbiçimli ayrımı için özvektörlere göre karar verilmektedir (Lucke, 2005a).

T=$\sum P_i(\theta)$

KTK'nda ise testi oluşturan maddelerin paralel olması durumunda gerçek puanların kestiricisi olarak gözlenen puanlar [X=E(T)] kullanılmaktadır. Ancak bu durum, maddelerin konjenerik ölçmeler olduğunda gerçek puanlar yanlı (bias) olarak kestirilecektir. Konjenerik maddeler için gerçek puanların kestiricisi olarak

$$\hat{T}_j = \sigma_E{}^2 \left[\sum_{i=1}^{k} \frac{\lambda_i}{\psi_i{}^2} (X_{ij} - \mu_i) \right] + \sum_{i=1}^{k} \mu_i$$

$$\sigma_E{}^2 = \left[\sum_{i=1}^{k} \frac{\lambda_i{}^2}{\psi_i{}^2} \right]^{-1}$$

eşitliği önerilmektedir. Bu eşitlikte X_{ij}, j. öğrencinin i. maddeye ilişkin gözlenen puanı göstermektedir. Bu yaklaşımla konjenerik maddelerin oluşturduğu ölçme aracına yönelik gerçek puanların yansız kestirimleri olanaklıdır (McDonald, 1999).

Konjenerik Güvenirlik

Birleşik testi oluşturan maddeler konjenerik bir yapıya sahip ise bu durumda klasik güvenirlik katsayıları (örneğin Cronbach α katsayısı) gerçek güvenirliğin altında değerler üretir (Lucke, 2005a, 2005b; Raykov, 1977, Zinbarg vd., 2005). Raykov (1997) ise yaptığı çalışmada hata terimleri arasındaki ilişkiler pozitif [$Kov(E_i,E_j)>0$] ise birleşik testin güvenirliğinin gerçek güvenirliğin üstünde, hata terimleri arasındaki ilişki negatif [$Kov(E_i,E_j)<0$] ise klasik güvenirlik katsayılarının gerçek güvenirliğin altında çıkabileceğini belirtmiştir. Her iki durumda da konjenerik ölçmelerde Cronbach α katsayısı yanlı davranmaktadır.

Klasik güvenirlikte, Eşitlik 5'te verildiği gibi gerçek puanlar varyansının kestiricisi olarak madde kovaryansları kullanılmaktadır. Konjenerik güvenirlikte ise standartlaştırılmamış faktör yükleri (λ_i) aynı zamanda kovaryans terimlerine karşılık geldiğinden dolayı i. ve j. madde yüklerinin çarpımı her iki maddeye ilişkin kovaryans terimini vermektedir (McDonald, 1999).

$$Kov(X_i,X_j)= \lambda_i\lambda_j \qquad (14)$$

Bu durumda i. maddeye ilişkin varyans ise

$$Var(X_i)=\lambda_i^2+\psi_i^2 \qquad (15)$$

ve i. maddeye yönelik gerçek puan varyansı ise;

$$Var(T_i)=\lambda_i^2 \qquad (16)$$

şeklindedir. Buna göre k maddeden oluşan birleşik testin güvenirliği madde toplamları üzerinden elde edilebilmektedir.

$$Var(X)=\left(\sum_{i=1}^{k}\lambda_i\right)^2+\sum_{i=1}^{k}\psi_i^2$$

$$Var(T)=\left(\sum_{i=1}^{k}\lambda_i\right)^2$$

Verilen bu eşitliklere göre konjenerik ölçmeler için McDonald tarafından geliştirilen omega (ω) güvenirlik katsayısı;

$$\omega=\frac{Var(T)}{Var(X)}=\frac{\left(\sum_{i=1}^{k}\lambda_i\right)^2}{\left(\sum_{i=1}^{k}\lambda_i\right)^2+\sum_{i=1}^{k}\psi_i^2}$$

şeklinde ifade edilmektedir (McDonald, 1985, 1999). Diğer taraftan, standartlaştırılmamış faktör analitik modelinden elde edilen terimler ile Cronbach α katsayısını elde etmek olanaklıdır (McDonald, 1985).

$$\alpha=\frac{k}{k-1}\left[\frac{k(\bar{\lambda})^2-(\overline{\lambda^2})}{k(\bar{\lambda})^2+\overline{\psi}^2}\right]$$

KonTK'na göre *madde güvenirlikleri* ise Eşitlik 17 ifadesi ile elde edilebilir.

$$\omega_i=\frac{Var(T_i)}{Var(X_i)}=\frac{\lambda_i^2}{\lambda_i^2+\psi_i^2} \qquad (17)$$

Konjenerik Madde Analizi

KonTK'ında madde analizi aynı zamanda KonTK modellerinin kestirilmesi için bir süreçtir. KonTK modellerinin kestirilmesi ise standartlaştırılmamış faktör çözümlerini içeren doğrulayıcı (confirmatory) faktör analizi ile olanaklıdır. Buna göre KonTK modelindeki b_i parametreleri madde ortalamalarını (madde güçlük düzeyleri) ve a_i parametreleri ise maddenin ayırıcılık gücüne (faktör yükleri) karşılık gelen standartlaştırılmamış faktör yüklerini gösteren madde özellikleridir (Lucke, 2005b).

$X_1=b_1+a_1T_1+E_1$

$\vdots$

$X_i=b_i+a_iT_i+E_i$

$X_j=b_j+a_jT_j+E_j$

$\vdots$

$X_k=b_k+a_kT_k+E_k$

KonTK'na dayalı model çözümlemesinin KTK modeline olan bir diğer üstünlüğü ise model uyumunun test edilebilirliğidir. Buna göre, veri kümesinden elde edilen madde kovaryansları σ_{ij} ve model kestiriminden elde edilen madde kovaryansları s_{ij} ile gösterilmek üzere

$\sigma_{ij}=\pi_{ij}-\pi_i\pi_j$

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

kullanılarak elde edilen s_{ij}-σ_{ij} farkı modelin geçerliği için kullanılan bir girdidir. Burada π_{ij}, i. ve j. maddeyi birlikte doğru yanıtlayanların oranını göstermektedir.

$$\delta=\frac{1}{k^2}\sum_{i}^{k}\sum_{j}^{k}\left(s_{ij}-\sigma_{ij}\right)^2 \qquad (18)$$

Tek boyutlu çalışmalar için Eşitlik 18 ile verilen ifadede Eşitlik 14, 15 ve 16 ile verilen ifadelerden yararlanarak *k*x*k* boyutlu kovaryans matrisinin köşegen ve köşegen dışı değerleri ayrıştırılabilir.

$$\delta = \frac{1}{k^2}\left[\sum_{i \neq j}\sum (s_{ij} - \lambda_i \lambda_j)^2 + \sum_i (s_{ii} - \lambda_i^{\ 2} - \psi_i^{\ 2})^2\right] \quad (19)$$

Eşitlik 19 ile verilen kovaryanslar arası fark değerleri KonTK modelinin geçerliği için kullanılmaktadır. Uyum iyiliği indeksi (goodness of fit index, GFI), Uİİ, tüm maddeler üzerinden uyum iyiliğini vermektedir.

$$c = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k} s_{ij}^{\ 2} \quad (20)$$

$$Uİİ = 1 - \frac{\delta}{c} \quad (21)$$

Elde edilen Uİİ değeri, 0.95 değerinden büyük ise ölçme modellerinin veri-model uyumunun mükemmel olduğu, 0.94 ve 0.90 değerleri arasında ise “kabul edilebilir” bir uyum olduğunu ifade etmektedir (McDonald, 1999). KonTK madde analizine ilişkin uygulama örneği bu çalışmanın yöntem bölümünde açıklanmıştır.

YÖNTEM

Bu araştırmanın uygulama bölümünde, çoktan seçmeli test maddelerinin özelliklerini belirlemek için klasik madde analizi ve konjenerik madde analizleri karşılaştırılmıştır. Bu amaçla 2001 yılında uygulanan Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı’nın (OK-ÖSYS) matematik alt testinde yer alan 25 madde ve bu maddelere ilişkin 553108 öğrenci yanıtları kullanılmıştır.

Ancak KonTK’nda madde analizi kapsamının açıklanması için OK-ÖSYS sınavında “sözel problem çözme” becerilerini ölçmek üzere hazırlanmış 5 madde (matematik alt testindeki - orijinal test formunda numaraları ile- 2., 3., 4., 5. ve 6. maddeler) ve basit rasgele örneklem ile seçilen 5000 öğrencinin [0-1] şeklinde puanlanmış yanıtları örnek olarak seçilmiş ve yöntem bölümünde kovaryans değerleri ile KonTK’na dayalı madde analizlerini açıklamalarla ele alınmıştır.

Bu maddelere ilişkin öğrenci yanıtlarından elde edilen varyans-kovaryans matrisi [$\sigma_{ij}=\pi_{ij}-\pi_i\pi_j$] ve madde ortalamaları Tablo 1’de verilmiştir. Varyans-kovaryans matrisinin köşegen elemanları ise [$\sigma_{ii}^{\ 2}=\pi_i(1-\pi_j)$] yaklaşımı ile elde edilmiştir.

Tablo 1: Maddelere ilişkin istatistikler

		Varyans-Kovaryans Matrisi (Σ)				
Madde	**p**	**1**	**2**	**3**	**4**	**5**
1	0,32	**0,22**				
2	0,32	0,04	**0,22**			
3	0,30	0,04	0,04	**0,21**		
4	0,40	0,04	0,05	0,05	**0,24**	
5	0,62	0,02	0,04	0,04	0,06	**0,24**

Tablo 1 ile verilen varyans-kovaryans matrisindeki[3] (Σ) kovaryans terimlerinin 0,02 ve 0,06 arasında değerler aldığı görülmektedir. Eğer tüm kovaryans terimleri eşit olsaydı 5 adet madde eşbiçimli ölçmeler olarak ele alınacaktı. Ancak kovaryans değerlerinin birbirine çok yakın olması nedeniyle 5 madde için "yaklaşık" eşbiçimli ölçmeler ifadesi kullanılabilir. Tablo 1'de verilen ve öğrenci yanıtlarından hesaplanan varyans-kovaryans değerleri, Eşitlik 14 ifadesindeki $s_{ij}=\lambda_i\lambda_j$ yaklaşımı ile yeniden kestirilerek (S) modellerin uygunluğu test edilebilir. Buna göre doğrulayıcı faktör modelinin çözümlenmesi ile elde edilen model terimleri ve bu terimlerden kestirilen kovaryans matrisi Tablo 2'de verilmiştir.

Tablo 2: 5 Maddeye İlişkin Kestirim Değerleri

				Varyans-Kovaryans Matrisi (**S**)				
Madde	**p**	λ_i	ψ_i^2	**1**	**2**	**3**	**4**	**5**
1	0,32	0,17	0,97	**0,03**				
2	0,32	0,19	0,96	0,03	**0,04**			
3	0,30	0,21	0,96	0,04	0,04	**0,04**		
4	0,40	0,26	0,93	0,04	0,05	0,05	**0,07**	
5	0,62	0,19	0,96	0,03	0,04	0,04	0,05	**0,04**

[3] Öğrenci yanıtlarından elde edilen maddeler arası varyans-kovaryans matrisi **Σ** sembolu ile, Kon TK modelinin kestirilmesi ile elde edilen varyans-kovaryans matrisi ise S sembolü ile verilmiştir.

Gözlenen puanlardan elde edilen kovaryans terimleri (Σ) ile kestirim sonucu (Eşitlik 14 ile) elde edilen kovaryans terimleri (S) arasındaki farklılık Tablo 3'te verilmiştir.

Tablo 3: Kovaryans farklar matrisi (**Σ–S**)

0,19	0,01	0,00	0,00	-0,01
0,01	**0,18**	0,00	0,00	0,00
0,00	0,00	**0,17**	0,00	0,01
0,00	0,00	0,00	**0,17**	0,01
-0,01	0,00	0,00	0,01	**0,20**

Tablo 3'te yer alan kovaryans farklar matrisinin köşegen elemanları aynı zamanda KonTK modelinin hatasını da içermektedir (Çizim 2). Bu hata terimleri $\psi_i^2=\sigma_{ii}-\lambda_i^2$ eşitliği ile de elde edilebilir. Buna göre maddelere ilişkin KonTK modelleri tekrar düzenlenirse;

$X_1=0,32+0,17T_1+0,19$

$X_2=0,32+0,19T_2+0,18$

$X_3=0,30+0,21T_3+0,17$

$X_4=0,40+0,26T_4+0,17$

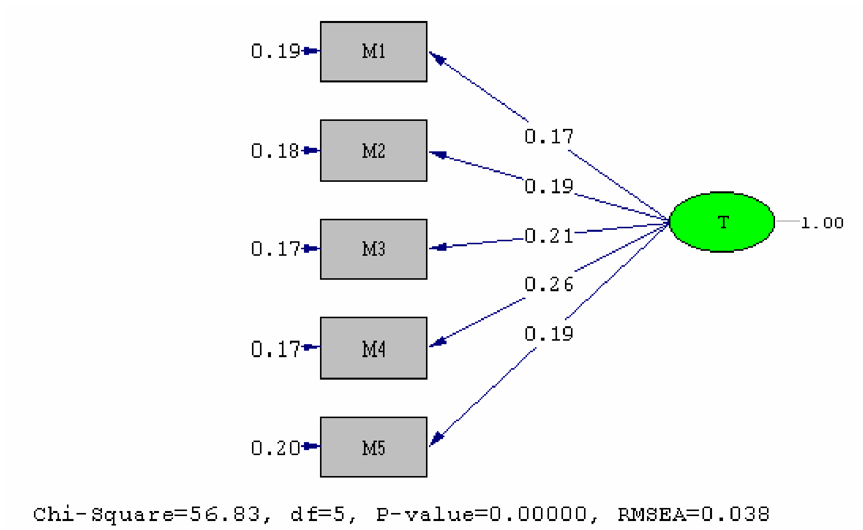
$X_5=0,62+0,19T_5+0,20$

eşitlikleri elde edilir. Burada yer alan modellerin uyum iyiliği indeksi ise Eşitlik 19 ve 20'ye göre;

$\delta=(1/5^2)*[0,01^2+0,00^2+0,00^2+(-0,01^2)+...+0,01^2]=0.00$

$c=(1/5^2)*[0,04^2+0,04^2+0,04^2+0,04^2)+...+0,06^2]=0.037$

Buna göre tüm maddelere ilişkin kurulan modellerin uyum iyiliği indeksi, Uİİ (GFI) Eşitlik 21'e göre 0,999 olarak elde edilmiştir. Bu değere göre 5 madde üzerinden elde edilen KonTK modelleri istatistiksel olarak anlamlıdır ve geçerlidir. Doğrulayıcı faktör yapısının geçerliliğinin test edilebilmesi için "uyum" ve "hata" üzerine kurulu farklı ölçütlerde geliştirilmiştir. Bunlardan 1980 yılında Steiger ve Lind tarafından önerilen "yaklaşıklık hata kareler ortalaması karekökü, YHKOK" (root mean square error of approximation, RMSEA) en yaygın kullanılan ölçüttür. Maddelere ilişkin kovaryans terimlerindeki (Σ-S) uyumunda bu ölçütün 0,05 değerinden küçük olması önerilir (Steiger, 2000).



Çizim 2: "Sözel problem" maddelerine ilişkin KonTK modellerinin elde edilmesi

Aynı zamanda 5 maddenin oluşturduğu veri kümesi için KR_{20} ve McDonald'ın ω katsayıları elde edilmiştir. KR_{20} katsayısı hesaplanırken; gözlenen puanlar varyansını elde edebilmek için madde varyansları toplamı ve madde kovaryansları toplamından yararlanılır.

$$\mathrm{Var}(X) = \sum_{i=1}^{k} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Kov}(X_i, X_j)$$

Bu ifadenin sağındaki ilk terim varyans-kovaryans matrisinin köşegen elemanlarının toplamlarına, ikinci terim ise varyans-kovaryans matrisinin köşegen dışı elemanlarının toplamına karşılık gelmektedir. Diğer bir ifade ile varyans-kovaryans matrisinin elemanları toplamı, gözlenen puanlar varyansına karşılık gelmektedir. Tablo 1'den yararlanarak varyans-kovaryans terimleri üzerinden 5 maddeye ilişkin gözlenen puanlar varyansı elde edilebilir.

Var(X)=1,13+0,84=1,97

KR_{20}=(5/4) [1-(0,22+0,22+0,21+0,24+0,24)/1,97]=<u>0,53</u>

Konjenerik güvenirliği (ω) elde edebilmek için ise Tablo 2'de verilen standartlaştırılmamış faktör yükleri kullanılırsa;

$$\omega=[(0,17+0,19+0,21+0,26+0,19)^2/1,97]=0,53$$

değeri elde edilir. Buna göre KR_{20} ve ω güvenirlik katsayıları eşit olarak elde edilmiştir. Tablo 2 göz önüne alındığında 5 adet maddenin yaklaşık eşbiçimli oldukları görülmektedir. Klasik güvenirlik katsayıları, ölçme aracındaki maddelerin paralel, eşdeğer ya da eşbiçimli olduğu durumlarda gerçek güvenirliği verdikleri, konjenerik ölçmelerde ise gerçek güvenirliğin altında değer ürettikleri ölçme literatüründe sıkça vurgulanmaktadır. Aynı zamanda paralel, eşdeğer ve eşbiçimli maddeler için konjenerik güvenirlik katsayısı (ω) klasik güvenirlik katsayıları ile aynı değerleri üretirken konjenerik maddeler için ω katsayısı klasik güvenirlik katsayılarından daha büyük değerler üretmektedir (Lucke, 2005a, Raykov, 1977, 2001, Traub, 1994).

Bu çalışmada OK-ÖSYS 2001 matematik alt testinde yer alan 25 madde için KonTK ve KTK modelleri elde edilmiş ve karşılaştırılmıştır.

BULGULAR

2001 yılında uygulanan OK-ÖSYS matematik alt testinde yer alan 25 maddeye 553108 öğrencinin verdikleri yanıtlara ilişkin KonTK ve KTK dayalı madde analizleri yapılmış ve KonTK modelleri elde edilmiştir. KonTK'na göre standartlaştırılmamış faktör analitik çözümlemelerden elde edilen model kestirimleri ile KonTK ve KTK'na dayalı madde istatistikleri verilmiştir.

Tablo 4'de madde istatistikleri olarak madde güçlük düzeyleri ve madde ayırıcılık düzeyleri verilmiştir. Her iki yaklaşım için madde güçlük düzeyleri ortak istatistiklerdir. Maddelerin ayırıcılık düzeyleri olarak KTK için çift serili ve nokta çift serili korelasyonlar KonTK için ise standartlaştırılmamış madde faktör yükleri verilmiştir. Çift serili ya da nokta çift serili korelasyonlar nicel olarak standartlaştırılmamış faktör yüklerine göre büyük değerler vermiştir. Bunun nedeni standartlaştırılmamış faktör yükleri ile korelasyonların aynı ölçekte yer almamasıdır. Her iki kuramsal yaklaşımdaki madde ayırıcılık düzeylerinin birlikte değişimini gözlemleyebilmek için değerler arasındaki korelasyonlar elde edilmiştir.

$$\rho(\lambda, r_{ÇS}) = 0,97$$

$$\rho(\lambda, r_{NÇS}) = 0,99$$

Tablo 4: 25 Maddeye ilişkin madde istatistikleri

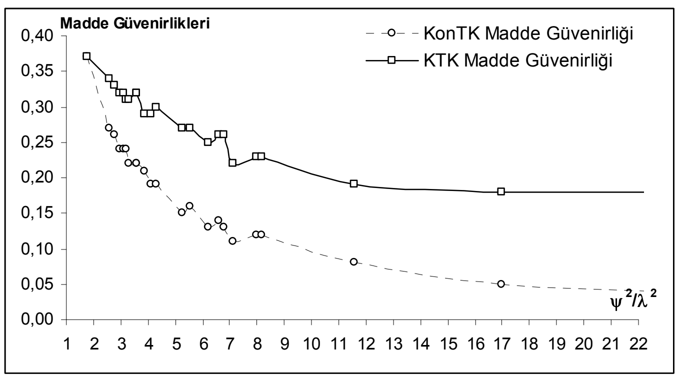
	KonTK Modelleri		**KonTK Madde İstatistikleri**			**KTK Madde İstatistikleri**			
Madde	**Model**	ψ^2	**b (μ)**	**a (λ)**	ρ_i (ω_i)	**p**	r^*	r^{**}	ρ_i
1	$X_1=0,34+0,23x\mathbf{T_1}$	0,17	0,34	0,23	0,24	0,34	0,50	0,65	0,31
2	$X_2=0,31+0,17x\mathbf{T_2}$	0,18	0,31	0,17	0,13	0,31	0,42	0,55	0,25
3	$X_3=0,32+0,17x\mathbf{T_3}$	0,19	0,32	0,17	0,14	0,32	0,42	0,55	0,26
4	$X_4=0,29+0,18x\mathbf{T_4}$	0,18	0,29	0,18	0,16	0,29	0,44	0,58	0,27
5	$X_5=0,39+0,24x\mathbf{T_5}$	0,18	0,39	0,24	0,24	0,39	0,51	0,65	0,32
6	$X_6=0,63+0,21x\mathbf{T_6}$	0,19	0,63	0,21	0,19	0,63	0,46	0,59	0,30
7	$X_7=0,34+0,28x\mathbf{T_7}$	0,14	0,34	0,28	0,37	0,34	0,59	0,76	0,37
8	$X_8=0,43+0,23x\mathbf{T_8}$	0,19	0,43	0,23	0,22	0,43	0,50	0,65	0,32
9	$X_9=0,34+0,19x\mathbf{T_9}$	0,19	0,34	0,19	0,15	0,34	0,44	0,57	0,27
10	$X_{10}=0,31+0,24x\mathbf{T_{10}}$	0,16	0,31	0,24	0,26	0,31	0,53	0,70	0,33
11	$X_{11}=0,46+0,22x\mathbf{T_{11}}$	0,20	0,46	0,22	0,19	0,46	0,47	0,59	0,29
12	$X_{12}=0,24+0,19x\mathbf{T_{12}}$	0,14	0,24	0,19	0,21	0,24	0,48	0,66	0,29
13	$X_{13}=0,20+0,14x\mathbf{T_{13}}$	0,14	0,20	0,14	0,11	0,20	0,38	0,54	0,22
14	$X_{14}=0,51+0,24x\mathbf{T_{14}}$	0,19	0,51	0,24	0,22	0,51	0,50	0,63	0,31
15	$X_{15}=0,23+0,14x\mathbf{T_{15}}$	0,16	0,23	0,14	0,12	0,23	0,38	0,54	0,23
16	$X_{16}=0,32+0,25x\mathbf{T_{16}}$	0,16	0,32	0,25	0,27	0,32	0,55	0,72	0,34
17	$X_{17}=0,38+0,24x\mathbf{T_{17}}$	0,17	0,38	0,24	0,24	0,38	0,51	0,65	0,32
18	$X_{18}=0,24+0,10x\mathbf{T_{18}}$	0,17	0,24	0,10	0,05	0,24	0,31	0,43	0,18
19	$X_{19}=0,41+0,23x\mathbf{T_{19}}$	0,19	0,41	0,23	0,22	0,41	0,51	0,64	0,32
20	$X_{20}=0,25+0,09x\mathbf{T_{20}}$	0,18	0,25	0,09	0,04	0,25	0,29	0,40	0,18
21	$X_{21}=0,16+0,01x\mathbf{T_{21}}$	0,13	0,16	0,01	0,00	0,16	0,13	0,19	0,07
22	$X_{22}=0,19+0,11x\mathbf{T_{22}}$	0,14	0,19	0,11	0,08	0,19	0,32	0,46	0,19
23	$X_{23}=0,23-0,03x\mathbf{T_{23}}$	0,18	0,23	-0,03	0,00	0,23	0,03	0,05	0,00
24	$X_{24}=0,73+0,15x\mathbf{T_{24}}$	0,18	0,73	0,15	0,12	0,73	0,39	0,53	0,23
25	$X_{25}=0,50+0,18x\mathbf{T_{25}}$	0,22	0,50	0,18	0,13	0,50	0,41	0,52	0,26

r^*; Nokta çift serili korelasyon ile elde edilen madde ayırıcılık değerleri,
r^{**}; Çift serili korelasyon ile elde edilen madde ayırıcılık değerleri,
ρ_i; Madde güvenirliklerini göstermektedir.

553108 öğrencinin maddelere verdikleri yanıtlar çarpık bir dağılım gösterdiğinden dolayı (çarpıklık katsayısı=1,0; basıklık katsayısı=0,62) nokta çift serili korelasyon faktör yükleri ile daha uyumlu sonuçlar vermiştir (Yurdugül ve Aşkar, 2004). Bu nedenle, çalışmada klasik ayırıcılık gücü indeksi olarak nokta çift serili korelasyon değerleri temel alınmıştır.

Her iki kuramsal modelden (KTK ve KonTK)elde edilen madde güçlük düzeyleri eşit ve madde ayırıcılık düzeyleri ise bağıntılı sonuçlar vermektedir. Bilindiği gibi ölçme aracındaki maddeler paralel, eşdeğer ya da eşbiçimli olduğunda her iki model aynı sonuçları verirken maddeler konjenerik olduğunda sonuçlar farklılaşmaktadır (McDonald, 1999).

Madde analizi sonucunda, maddelere ilişkin istatistiklerde en önemli farklılık madde güvenirliklerinde ve madde ayırıcılıklarında ortaya çıkmıştır. Madde güvenirliklerini karşılaştırmak için modellerin “standart hata” değerleri elde edilmiştir. Bu değerler, her bir regresif modelin hatası (ψ^2) modelin gerçek puanlar varyansı niteliği taşıyan faktör yükünün karesine (λ^2) bölünerek elde edilmiştir. Buna göre Çizim 3'te konjenerik modelden elde edilen madde güvenirliği ile klasik madde analizinden elde edilen madde güvenirliklerinin modelin standart hatalarına göre değişimleri verilmiştir. Modelin standart hatası düştükçe her iki kuramsal modelden elde edilen madde güvenirlikleri yaklaşık olarak eşit değerler üretmektedir. Ancak modellerin standart hataları arttıkça madde güvenirlikleri düşmekte ve klasik madde analizinden elde edilen madde güvenirlikleri konjenerik madde analizinden elde edilen sonuçlara göre daha büyük değerler üretmektedir. Bir diğer ifade ile konjenerik madde analizinde madde güvenirlikleri "hata" teriminden daha çok etkilenmektedir. Diğer taratan, konjenerik ve klasik madde analizinden elde edilen madde güvenirlikleri arasındaki Pearson korelasyon katsayısı 0,92 olarak elde edilmiştir.



Çizim 3: Standart hatalara göre madde güvenirliklerinin

Diğer taraftan; madde ayırıcılık gücü aynı zamanda madde güvenirliklerinin bir fonksiyonu olduğu için Eşitlik 17'ye göre elde edilen madde güvenirlikleri ile madde ayırıcılık gücü değerlerinin grafiği Ek 2'de verilmiştir. Ek 2'deki grafiksel gösterime ve Tablo 4'de göre madde ayırıcılık gücü olarak en düşük değerler konjenerik madde analizinden elde edilmiştir. Klasik madde analizinden elde edilen çift serili korelasyon değerleri ise en yüksek ayırıcılık gücü değerleri olarak elde edilmiştir. Bununla birlikte konjenerik ve klasik madde analizinden elde edilen madde ayırıcılıkları arasındaki Pearson korelasyon katsayısı 0,99 olarak elde edilmiştir. Bu durum her iki yaklaşımdan elde edilen madde ayırıcılık gücü düzeyleri arasında fonksiyonel bir bağıntı olduğunu göstermektedir.

Ölçme aracının geneline ilişkin güvenirlikler elde edilirken klasik güvenirlik olarak KR_{20} ve konjenerik güvenirlik olarak McDonald'ın ω katsayısı kullanılmıştır. Buna göre KR_{20}=0,81 ve ω=0,82 olarak elde edilmiştir. Ölçme aracındaki maddeler paralel, eşdeğer ya da eşbiçimli olduğu sürece klasik güvenirlik ve konjenerik güvenirlik katsayıları aynı sonuçları üretirken konjenerik ölçme araçlarında konjenerik güvenirlikler klasik güvenirliklere göre daha yüksek değerlere sahip olmakta ve gerçek güvenirliği daha yansız kestirebilmektedirler (Lucke, 2005a, 2005b; Raykov, 1997, 2001). Bu amaçla 25 maddeye ilişkin ikili kovaryans terimlerinin standart sapması hesaplanmış ve 0,02 elde edilmiştir. Kovaryans terimlerinin eşit olmama durumu konjenerik ölçmelerin varlığını gösterdiği için konjenerik güvenirlik (ω) ve klasik güvenirlik (KR_{20}) arasındaki 0,01 değerindeki farkın bu nedenle oluştuğu düşünülmüştür. Ancak kovaryans terimlerindeki 0,02 düzeyindeki sapmanın nicel büyüklük olarak oldukça küçük olması aynı zamanda konjenerik ve klasik test güvenirliği arasındaki farkında çok küçük çıkmasına neden olduğu düşünülebilir.

TARTIŞMA

Bu çalışmada bir ölçme aracında yer alan maddelerin yapısı (paralel, eşdeğer, eşbiçimli ve konjenerik) ve bu yapılar üzerine kurulu olan ölçme kuramlarından klasik test kuramı ve konjenerik test kuramına yer verilmiştir. Klasik test kuramı, gözlenen puanların bileşenleri olan gerçek puanlar ve hata puanlarının kestirilmesi ilkesi üzerine kuruludur ve bu bağıntı Eşitlik 1 ile verilen KTK modeli ile ifade edilir. KTK modeli enküçük kareler (least square estimation) yöntemi ile kestirilen bir doğrusal bağıntıdır (Lord ve Novick, 1968) ancak KTK modelinde maddelerin karakteristik özelliklerine ilişkin terimler yer almamaktadır. KTK modelinin varsayımlarına göre hataların ilişkisiz olmaları söz konusudur. Bu varsayım maddeler arasındaki kovaryansların eşit olduğu (paralel, eşdeğer ve eşbiçimli maddeler) durumunda sağlanmaktadır. Konjenerik test kuramı, klasik test kuramının bir alt kuramı olmakla beraber KonTK modelinde madde özelliklerine ilişkin terimlere de yer verilmektedir. KonTK modelinin KTK modeline göre en önemli avantajı, maddelere ilişkin kovaryans terimlerinin eşit olmadığı (ilişkili hatalar-correlated errors) durumlarda da kullanılabilmesidir (Alwin, 1976; Lucke 2005a; McDonald, 1999).

Bu çalışmada klasik test kuramına dayalı madde analizinden ve konjenerik modelin kestiriminden elde edilen madde parametreleri karşılaştırılmıştır. Her iki yaklaşımdan elde edilen parametrelere göre; madde güçlük indeksleri aynı yöntemle elde edildiği için eşit çıkmıştır.

Ancak klasik madde analizinden elde edilen madde güvenirliklerinin ve madde ayırıcılık gücü değerlerinin belirgin bir şekilde konjenerik madde analizinden elde edilen değerlere göre yüksek olduğu gözlenmiştir. Bu farklılığın, konjenerik madde analizinde standartlaştırılmamış değerler kullanıldığından dolayı olabileceği düşünülmektedir.

Bu çalışmanın kapsamına alınmamakla birlikte, ölçme modellerinin temel amaçlarından birisi ölçülmek istenilen öğrenme ürünleri ya da psikolojik yapıya ilişkin bireylerin gerçek puanlarını kestirmektir. Klasik test kuramında ölçme aracı paralel maddelerden oluştuğu durumlarda gözlenen madde puanlarının toplamı gerçek puanların bir kestiricisidir. Ancak maddeler konjenerik bir yapıya sahip olduğunda gerçek puanları, gözlenmiş puanlar üzerinden elde etmek yanlı sonuçlara neden olmaktadır.

ÖNERİLER

Bu çalışmada çoktan seçmeli maddelerin çözümlenmesine yönelik konjenerik kurama dayalı çözümleme ile klasik test kuramına dayalı çözümler karşılaştırılmıştır. Ancak maddelerin kalitesini ortaya çıkarmaya yönelik elde edilen sonuçlar birbirine yakın çıkmıştır. Bunun olası nedenlerinden birisi üzerinde çalışılan maddelerin yaklaşık olarak eşbiçimli yapıya sahip olmaları gösterilebilir. Bu durum konjenerik maddelerin yer aldığı ölçmeler ile tekrarlanıp sonuçlar karşılaştırılabilir. Diğer taraftan 2001 OK-ÖSYS matematik puanları çarpık bir dağılım göstermektedir (Yurdugül ve Aşkar, 2004). Bilindiği gibi doğrusal modeller için veri kümelerinin normal dağılım varsayımını sağlamaları gerekmektedir. KTK ve KonTK modelleri de istatistiksel doğrusal modeller olduğu için; benzer çalışmalar farklı dağılıma sahip ölçmeler için yapılarak sonuçlar karşılaştırılabilir.

Jöreskog (1971) tarafından geliştirilen konjenerik ölçme modelinde maddeler arası kovaryans terimleri doğrulayıcı faktör analizinin çözümlenmesi ile elde edilmektedir. Bu kovaryans terimleri aynı zamanda gerçek puanlar varyansının bir kestirici olduğundan dolayı güvenirlik kavramı ile doğrudan ilişkilidir. Ölçme literatüründe son zamanlarda oldukça yaygınlık kazanan konjenerik ölçme kuramı üzerine yapılan çalışmalarda "konjenerik güvenirlik" kavramı daha çok ele alınmaktadır. Klasik ve konjenerik güvenirliklerin paralel, eşdeğer, eşbiçimli ve konjenerik özellik taşıyan testlerden elde edilerek sonuçlar karşılaştırılabilir.

Eğitim alanında yapılan ölçmeler göz önüne alındığında; genellikle bir ölçme aracından elde edilen gözlenen puanlarda doğrudan değerlendirme kapsamına alınmaktadır. Eğer ölçme aracındaki maddelerin tamamı paralel madde ise bu uygulama doğaldır. Çünkü paralel maddelere ilişkin gözlenen puanlar aynı zamanda gerçek puanlara eşittir. Ancak ölçme aracındaki maddeler konjenerik yapıya sahip ise bu durumda öğrencilerin gözlenen puanlardan hareketle gerçek puanların kestirilmesi gerekmektedir. Bu konuya ilişkin bir diğer nokta ise; klasik test kuramının modelinin madde istatistiklerinden bağımsız olmasıdır. Bu durum KTK'nın en önemi sınırlılığıdır. Çünkü bireylere uygulanan ölçme aracında yer alan maddeler, çok kolay ya da çok zor olduğunda bireylerin gerçek puanları farklılık gösterecektir. Bu nedenle gerçek puanlar kestirilirken madde istatistiklerinin göz önüne alınması gerekir. Konjenerik ölçme modelinin, klasik test modeline yönelik önemli üstünlüklerinden birisi de madde istatistiklerinin modelde yer alması, yani gerçek puanlar kestirilirken madde özelliklerinden yararlanılmasıdır. Bu çalışmada ele alınmamış olmakla birlikte konjenerik ölçme kuramında "gerçek puanların kestirimi"ne ilişkin incelemeler sonraki çalışmalara bırakılmıştır.

KAYNAKLAR

Adamson, G., Shevlin, M., Lloyd, N.S.V. & Lewis, C. A. (2000). An integrated approach for assessing reliability and validity: an application of structural equation modeling to the measurement of religiosity. *Personality and Individual Differences* 29 (2000) 971-979

Alwin, D. F. (1976). Attitude scales as congeneric tests: A re-examination of an attitude-behavior model. *Sociometry, 39,* 377-383.

Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: klasik test teorisi ve uygulaması*. ÖSYM Yayınları, Ankara.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16(3)*, 297-334.

DeVellis, R. F. (2003). *Scale development: Theory and application*, Sage Publications, California.

Hambleton, R. K. & Swaminathan H. (1985). *Item response theory: principles and application.* Kluwer-Nijhoff Publising, Boston.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109–133.

Jöreskog, K. G. & Sörbom, D. (1996). *LISREL 8: user's guide*. Chicago: Scientific Software.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Lawrence Erlbaum. Associates, Inc., Hillsdale,New Jersey.

Lord, F. M., & Novick, R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley.

Lucke, J. F. (2005a). The α and the ω of congeneric test theory: an extension of reliability and internal consistency to heterogeneous tests. *Applied Psychological Measurement, 29(1)*, 65–81.

Lucke, J. F. (2005b). “Rassling the hog”: the influence of correlated item error on internal consistency, classical reliability, and congeneric reliability. *Applied Psychological Measurement, 29(1)*, 106–125.

McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Erlbaum.

McDonald, R. P. (1999). *Test theory: a unified treatment*. Mahwah NJ: Erlbaum.

Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32, 1–13.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21(2)*, 173-184.

Raykov, T. (2001). Bias of coefficient alfa for fixed congeneric measures with correlated errors. *Applied Psychological Measurement*, 25(1), 69–76

Raykov, T. & Penev, S. (1997). Structural equation modeling and the latent linearity hypothesis in social and behavioral research. *Quality & Quantity*, 31, 57-78.

Steiger, J. H. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduk and Glasser. *Structural Equation Modelling, 7(2),* 149-162.

Traub, E. R. (1994). *Reliability for the social sciences: theory and applications*. Measurement methods for the social sciences. Sage Publications.

Yurdugül, H. ve Aşkar, P. (2004). Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı’nın cinsiyete göre madde yanlılığı açısından incelenmesi. *Eğitim Bilimleri ve Uygulama*, 3(5), 3-20.

Zinbarg, R. E., Revelle, W., Yovel, I. & Li W. (2005). Cronbach’s α, Revelle’s β, and Mcdonald’s ω_H: their relations with each other and two alternative conceptualizatıons of reliability. *Psychometrika*, 70(1), 1–11.

EK 1:
Ölçmelerin/Maddelerin Yapısı ve Özellikleri

Ölçmeler	**Varsayımlar**	**Eşitlikler**	**Betimsel Özellikleri**	**Faktör Analitik Özellikleri**
Paralel	$T_1=T_2=T_3$ $\sigma^2_{E1} = \sigma^2_{E2} = \sigma^2_{E3}$	$\overline{X}_1 = \overline{X}_2 = \overline{X}_3$ $\sigma_{X_1} = \sigma_{X_2} = \sigma_{X3}$ $\sigma_{X_1X_2} = \sigma_{X_2X_3}$	Her üç madenin ortalamaları, standart sapmaları ve kovaryansları eşit.	Faktör yükleri eşit
Eşdeğer	$T_1=T_2=T_3$ $\sigma^2_{E1} \neq \sigma^2_{E2} \neq \sigma^2_{E3}$	$\overline{X}_1 = \overline{X}_2 = \overline{X}_3$ $\sigma_{X_1} \neq \sigma_{X_2} \neq \sigma_{X3}$ $\sigma_{X_1X_2} = \sigma_{X_2X_3}$	Maddelere ilişkin ortalamalar eşit ve standart sapmalar farklı, kovaryanslar eşit.	Faktör yükleri eşit
Eşbiçimli	$T_1= a_{12}+T_2$ $T_1= a_{13}+T_3$ $T_2= a_{23}+T_3$ $\sigma^2_{E1} \neq \sigma^2_{E2} \neq \sigma^2_{E3}$	$\overline{X}_1 \neq \overline{X}_2 \neq \overline{X}_3$ $\sigma_{X_1} \neq \sigma_{X_2} \neq \sigma_{X3}$ $\sigma_{X_1X_2} = \sigma_{X_2X_3}$	Maddelere ilişkin ortalamalar ve standart sapmalar farklı, kovaryanslar eşit.	Faktör yükleri eşit
Konjenerik	$T_1=b_{12}T_2+a_{12}$ $T_1= b_{13}T_3+a_{13}$ $T_2= b_{23}T_3+a_{23}$ $\sigma^2_{E1} \neq \sigma^2_{E2} \neq \sigma^2_{E3}$	$\overline{X}_1 \neq \overline{X}_2 \neq \overline{X}_3$ $\sigma_{X_1} \neq \sigma_{X_2} \neq \sigma_{X3}$ $\sigma_{X_1X_2} \neq \sigma_{X_2X_3}$	Maddelere ilişkin ortalamalar, standart sapmalar ve kovaryanslar farklı.	Faktör yükleri farklı

EK 2:
Madde Güvenirlikleri ve Madde Ayırıcılıkları Arasındaki İlişki