

Test Güvenirliği ve Test Bilgi Fonksiyonunun Klasik ve Örtük Özellikler Test Teorilerine Göre Kestirilmesi

The Prediction of Test Reliability and Test Information Function According to Classical Test Theory and Latent Trait Theory

Duygu Anıl
Hacettepe Üniversitesi

Öz

Bu araştırmada, deneme uygulaması yapılamayan durumlarda madde özelliklerinin uzman tahminlerine dayalı kestirimlerinin Klasik Test Teorisi ve Örtük Özellikler Teorisinin iki parametreliliği logistik modeline dayalı olarak hesaplanan test güvenirliliğini ve test bilgi fonksiyonlarını tahmin etme gücünün nasıl olduğu incelenmeye çalışılmıştır. Araştırma, deneme uygulaması yapılamayan sınavlardan biri olan Milli Eğitim Bakanlığı'nın 1999-2000 öğretim yılı Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı'na giren rasgele seçilmiş 9914 öğrencinin matematik alt testlerine vermiş oldukları cevapları üzerinde yürütülmüştür. Araştırmada, uzman tahmin değerleri; 16 matematik öğretmenine uygulanan her iki teoriye göre ayrı ayrı hazırlanmış metrik ölçek uygulamalarından elde edilmiştir.

Araştırmadan elde edilen bulgularda, Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı matematik alt testine ait uzman tahminlerine dayalı olarak elde edilen testin güvenirliliği ile klasik test teorisine göre elde edilen testin güvenirliliği arasında karşılaştırmalar yapılmış, 0,05 düzeyinde anlamlı bir fark bulunmamıştır. Araştırmada, Örtük Özellikler Teorisinde yer alan test bilgi fonksiyonları üzerinde çalışılmış ve uzman tahmin değerlerine ve Örtük Özellikler Teorisine göre farklı yetenek düzeyleri için kestirilen test bilgi fonksiyon değerlerinin -2 ve $+0,5$ yetenek düzeyine sahip olan bireyler için artma gösterdiği, $+0,5$ ve $+2$ yetenek düzeyine sahip olan bireyler içinde test bilgi fonksiyon değerlerinde düşme gösterdiği görülmüştür.

Anahtar sözcükler: Klasik Test Teorisi, Örtük Özellikler Teorisi, madde ve test parametrelerinin tahmini, test bilgi fonksiyonları.

Abstract

In this study, the prediction power of the item characteristics based on the experts' predictions in exams which cannot be piloted was examined for test reliability and test information functions. For this aim, classical test theory and logistic model with two parameters of latent trait theory were used. The study was carried out on the results of the math sub-tests of 9914 randomly selected students who took The Student Selection and Placement Examination for Secondary Schools in the 1999-2000 academic year. The experts' prediction values were obtained from separate metric scales based on the above mentioned theories. These scales were given to 16 math teachers.

The results of the study revealed no significant difference (0,05) between test reliability values obtained through experts' predictions and through classical test theory. Furthermore, test information functions taking place in latent trait theory was examined within the scope of this study. The findings showed that test information function values predicted for different competence levels increase for individuals having -2 ve $+0,5$ competence level, test information function values decrease for the individuals having $+0,5$ and $+2$.

Keywords: Classical test theory, latent trait theory, prediction of item and test parameters, test information functions

Giriş

Günlük yaşamımızda ve bilimlerin gelişmesinde ölçme

Arş. Gör. Dr. Duygu Anıl, Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü, Ankara. aduygu@hacettepe.edu.tr

önemli bir yer tutar. Bugün bilimler ölçme yöntemlerini geliştirebildiği, ölçme araçlarını çok küçük farkları ölçecek şekilde hassas hale getirebildiği oranda hızlı bir gelişme gösterebilmektedir. Ölçmesiz bilimin olamaya-

cağı yargısı bir gerçeğin ifadesidir “Doğada birşey varsa, belirli bir miktar vardır” diyen Thorndike ve “Bir şey, belli bir miktarda varsa, o şey ölçülebilir” diyen McCall bütün bilim dalları için ölçmenin önemi ve yerini tartışma gerektirmeyecek açıklıkta ortaya koymaktadırlar.

Eğitim ve psikolojinin de içinde yer aldığı davranış bilimlerinde, ölçmek istediğimiz nitelikler fiziksel bilimlerde olduğu gibi doğrudan gözleyebildiğimiz değişkenler değildir. Doğrudan gözlenemeyen değişkenleri ölçmede; hatayı en aza indirmek, ölçmek istediğimiz niteliğe başka değişkenleri karıştırmamak ve daha duyarlı ölçme sonuçları elde etmek için diğer bilim dallarında olduğu gibi eğitimdeki ölçmelerde de araç kullanılmaktadır. Eğitim sistemi içerisinde farklı amaçlarla ölçme yapıldığından, eğitimde kullanılan ölçme araçları da ölçülecek özelliğe, aracın kullanılacağı gruba ve ölçme amacına bağlı olarak çeşitlilik gösterir. Eğitimde kullanılan ölçme araçlarının başlıcaları; yazılı yoklamalar, kısa cevaplı testler, doğru-yanlış testleri, çoktan seçmeli testler, sözlü yoklamalar, ödevler ve projelerdir.

Bireyler hakkında kararların isabetlilik derecesi ölçütün uygunluğu yanında, ölçme sonuçlarının veya ölçme araç ve yöntemlerinin güvenilirlik ve geçerliğine bağlıdır. Eğitimdeki ölçmelerde yukarıda sözü edilen araçlardan çoktan seçmeli testler, bu iki özellik yönünden diğerlerine göre üstünlük sağlar. Eğitim ve psikolojideki ölçmelerde çoktan seçmeli test geliştirme çalışmaları ile özellikleri önceden kestirilebilen testler hazırlamak mümkün olmaktadır. Bu bakımdan çoktan seçmeli testler, özellikle büyük gruplar üzerinde yapılan ölçme çalışmaları için halihazırda en geniş uygulama alanı bulan araç durumundadırlar.

Test geliştirmenin amacı, uygun madde seçerek istenilen nitelikte yani güvenilir ve geçerli ölçme aracı geliştirmek olduğundan, madde analiz çalışmaları test geliştirme sürecinde önemli bir yer tutmaktadır. İstenen niteliklere uygun bir test elde etmek, istenen test parametrelerini verecek maddeler seçilmesini gerektirir. Bu da maddelerin yapısı hakkında bilgi edinmeyi gerektirir. Bu bilginin sağlanması, deneme uygulaması yapmakla mümkün olmaktadır. Örneğin; Klasik Test Teorisi'nde kullanılan madde istatistikleri, genellikle, madde güçlük ve madde ayırıcılık gücü indeksleri, madde güvenilirliği ve maddelerarası korelasyonlardır. Belirtilen bu madde istatistikleri yardımıyla testten elde edilecek puanların ortalaması, standart kayması, güvenilirlik ve geçerliği gibi test istatistikleri kestirilebilir. Test geliştirmede en çok başvurulan iki teoriden biri Klasik Test Teorisi, diğeri ise

Örtük Özellikler Teorisidir. Günümüzde madde analiz çalışmalarında en çok uygulama alanı bulan Klasik Test Teorisi' dir. Bunun yanında, özellikle 1950'lerden sonra ölçme araçlarının geliştirilmesinde Örtük Özellikler Teorisi (Latent Trait Theory) de etkili olmuştur. Bu iki teori ana hatlarıyla aşağıda sırasıyla açıklanmıştır.

Klasik Test Teorisi

Klasik Test Teorisi'nde bir bireyin bir teste ilişkin gözlenen puanı (X); gerçek (T) ve hata (E) puanları toplamından oluşur ve $X=T+E$ eşitliği, Klasik Test Teorisi'nin temel denklemini oluşturur (Hambelton ve Swaminathan, 1985, 15). Klasik Test Teorisi'ne göre ölçümlerin tamamen hatadan arınık olması beklenmeyen durum olmasına rağmen, Klasik Test Teorisi (CTT) ile hatayı en aza indirerek gözlenen puanın gerçek puana yakın bir değer alması olası bir durumdur. Bu amaca ulaşabilmek için Klasik Test Teorisi aynı özelliğin birden fazla ölçülmesi yolunu kullanır. Çünkü hata, ortalaması 0 olan tesadüfi bir değişkendir ve ancak sonsuz sayıda ölçme yapıldığı zaman bu değer 0 olur. Bu sebeple sonsuz sayıda yapılan ölçme sonucunda elde edilen kişinin gözlenen puanının ortalaması, o kişinin gerçek puanına eşit olur. Fakat aynı özelliği ölçmek için sonsuz sayıda uygulama yapılamayacağından dolayı, testler kişinin gerçek puanını tam olarak kestiremez.

Örtük Özellikler Teorisi

1916 yılında Binet ve Simon tarafından, bir maddeye verilen doğru cevabın yaşın bir fonksiyonu olarak ifade edilmesinden bu yana madde analizi, eğitimdeki ölçme çalışmalarında önemli bir konu olarak günümüze kadar gelmiştir. 1950'lerden sonra Klasik Test Teorisi'nin sınırlılıklarına alternatif olarak Örtük Özellikler Teorisi (Latent Trait Theory) adı altında bir teori daha geliştirilmiştir. Bu teori madde analizi yanında, bireyleri belli özelliklerine göre bir ölçekte sıralamaya da fırsat vermektedir (Baker, 1977, 151). Örtük Özellikler Teorisi bireyin ölçülen özelliği ile maddeye ilişkin gösterdiği performansı arasındaki ilişkinin matematiksel olarak ortaya konulabileceğini varsaymaktadır. Bu matematiksel tanım, madde karakteristik fonksiyonu ve eğrisi olarak adlandırılır (Stocking, 1997). Örtük Özellikler Teorisi'nde madde puanının θ vektörü üzerindeki regresyonu “madde karakteristik fonksiyonu” olarak adlandırılır ve $P_g(\theta)$ ile gösterilir. $P_g(\theta)$ fonksiyonu, tanımı gereği bir gruptan diğerine değişmez ve $P_g(\theta)$ değerini belirleyen parametreler sabit parametreler haline gelirler (Lord ve Novick, 1968, 360).

Örtük Özellikler Teorisi'nde madde karakteristik fonksiyonlarının matematiksel yapılarına bağlı olarak farklı modeller ortaya çıkmıştır. Bu modellerin kullanılmasında Örtük Özellikler Teorisi'nin bazı varsayımlarının karışılması gerekmektedir. Bu varsayımlar, dağılımın normal olması, tek boyutlu olması ve yerel bağımsızlıktır.

Örtük Özellikler Teorisi'ndeki modeller çok boyutlu geliştirilmesine rağmen, geleneksel olarak ölçülen madde havuzunun tek boyutlu olduğu varsayımı üzerine kurulmuştur. Bütün tek boyutlu Örtük Özellikler Teorisi'nde ki modeller şu varsayıma sahiptir: Tek bir belirleyici örtük yapı (θ) her bir test maddesine verilmiş olan gözlenen tepkilerin bir ürünüdür. θ 'nın maddeye verilen tepkileri oluşturuluş biçimine bağlı olarak farklılıklar göstermektedir (Harvey ve Hammer, 1999).

Bu farklılıklar farklı Örtük Özellikler Teorisi modellerini oluşturmuştur. Bu modeller bütün durumlarda madde parametreleri θ ile gözlenen madde tepkileri arasında var olan ilişki biçimini tanımlamaktadır. Yapılan çalışmada, iki kategorili ölçekler için geliştirilmiş olan 3 önemli Örtük Özellikler Teorisi modelinden söz edilmiş ve araştırmada kullanılan bu modellerden biri olan iki parametrelili lojistik model hakkında aşağıda bilgi verilmiştir.

Örtük Özellikler Teorisi modellerinden biri olan, bir parametrelili lojistik modelin en önemli sınırlılığı, bütün maddelerin eşit ayırıcılık gücüne sahip olduğu varsayımdır. Bir parametrelili lojistik model madde cevap sürecini temsil etmek için sadece tek bir madde parametresini varsaymaktadır. Bu parametre, madde güçlüğü olarak adlandırılır ve b_i parametresi olarak gösterilir. Bütün maddelerin eşit ayırıcılık gücüne sahip olduğu varsayımı, teoride mümkün olurken, uygulamada, pek rastlanan bir durum değildir. Bu nedenle, madde karakteristik fonksiyonları hesaplanırken, maddelerin ayırıcılık gücünün de (a_i parametresi) dikkate alınması gerçeği ile birlikte, iki parametrelili lojistik model adı verilen bir model oluşturulmuştur.

Bu modelde kullanılan $P(\theta)$, b_i , a_i ve θ değerleri normal ogive modelinde kullanılanla aynı anlamdadır. Bu model D ölçekleme faktörünü içermektedir. İki parametrelili lojistik modelin iki parametrelili normal ogive modelinden farkı a_i parametresinin 1,7 olan D değeri ile çarpıldığında normal ogive modeline yaklaşmasıdır (Hambelton ve Swaminathan, 1985, 37). Bu modelin madde karakteristik fonksiyonu aşağıda belirtilmiştir.

$$P_i(\theta) = \frac{e^{D a_i(\theta - b_i)}}{1 + e^{D a_i(\theta - b_i)}} \quad (i=1,2,3,\dots,n) \quad (1)$$

Bir parametrelili lojistik modelde kestirilmeye çalışılan tek madde parametresi, örtük özelliğin düzeyi ya da madde güçlük düzeyi anlamındaki b_i parametresidir. Bir maddenin b_i parametresi, maddenin 0,50 doğru cevaplandırılma olasılığına yetenek ölçeği üzerinde karşılık gelen değerdir. b_i parametresi θ 'ya dayalı olarak tanımlandığından, hem b_i hem de θ aynı ölçek üzerindedir. b_i parametresi, örtük özelliğin düzeyini ya da maddenin güçlük düzeyini gösterir ve teoride $(-\infty, +\infty)$ aralığında değerler aldığı kabul edilir. Hambelton ve Swaminathan (1985, 36) bir grubun puanlarını $\bar{X}=0$ ve $S_x=1$ olarak transfer ettiğimizde uygulama da bu parametrenin $(-2; +2)$ aralığında değerler aldığı belirtilmiştir. Madde kolaylaştıkça bu değer -2'ye, madde zorlaştıkça bu değer +2'ye yaklaşır. Madde güçlük indeksi 0 civarında olan maddeler orta güçlükte maddelerdir. Bu parametre maddenin en iyi ölçtüğü yetenek düzeyi olarak da adlandırılır.

İki parametrelili lojistik model ise ayırıcılık gücü ya da a_i adını verdiğimiz yeni bir parametreyi modele katmaktadır. Ve bu parametre madde karakteristik eğrilerinin farklı maddeler için farklı eğimlere sahip olmasını sağlamaktadır. a_i parametresi, maddenin kalitesi yani elde edilen θ yetenek ölçüsünün gerçek θ hakkında ne kadar bilgi taşımakta olduğunu gösterir (Lord ve Novick, 1968, 368). Buna göre a_i değeri arttıkça maddenin ayırıcılık gücü ve maddenin testin bütünü ile ölçülen özelliğe sağladığı katkı ya da bilgi de artmaktadır. Bu bakımdan a_i parametresi madde geçerlik ölçüsü olarak da tanımlanabilir. Madde ayırt edicilik indeksi olan a_i parametresi teorik olarak $(-\infty, +\infty)$ aralığında değer alır. Ama yetenek testlerinde negatif ayırt edici maddeler çıkartılır. Bu nedenle genellikle uygulamalarda a_i parametresi +2'den büyük değer almadığından bu parametrenin $(0, +2)$ aralığında değerler aldığı belirtilmektedir (Hambelton ve Swaminathan, 1985, 36). Maddenin a_i parametre değeri arttıkça bu değer 1'in üstünde +2'ye yakın değerler almaktadır. a_i parametre değeri 1 olan maddeler en uygun ayırıcılığa sahip olan maddelerdir. Maddenin ayırıcılığı arttıkça bu değer yükselir. a_i parametre değeri +2'ye yaklaşan maddeler, sadece üst yetenek grubundaki öğrenciler için ayırıcı olmaktadır. Farklı yetenek grubunda olan öğrencileri birbirlerinden ayıramayan maddelerin ayırıcılık güçleri düşüktür ve bu tür maddelerin ayırıcılık güçleri 0'a yaklaşır. Bu durumda maddenin testin bütünü ile ölçülen özelliğe sağladığı katkı ya da bilgi miktarı azalmaktadır.

İki parametrelili lojistik modelin önemli bir sınırlılığı ise maddenin şansla doğru cevaplandırılma olasılığını (c_i

parametresi) dikkate alınmasıdır. Bu sınırlılığı gidermek amacı ile iki parametrelili modele c_1 parametresi eklenerek üç parametrelili lojistik model geliştirilmiştir.

Örtük Özellikler Teorisi'nde modellerin kullanılabilirliği için bazı varsayımların karşılanması gerekmektedir. Bu varsayımlar, dağılımın normal olması, tek boyutlu olması ve yerel bağımsızlıktır. Örtük Özellikler Teorisi'nin gücü bu sayıtların yerine getirilmesine bağlıdır.

Klasik Test Teorisi'ndeki güvenilirlik kavramının Örtük Özellikler Teorisi'ndeki karşılığı, madde ve test bilgi fonksiyonları olarak görülmektedir ve Klasik Test Teorisi'nde güvenilirlik ve ölçme hataları ölçmenin yapıldığı gruba bağımlıyken, Örtük Özellikler Teorisi'ndeki madde ve test bilgi fonksiyonları ile ölçme hataları, ölçmenin yapıldığı gruptan bağımsızdır. Ayrıca, Klasik Test Teorisi'ne dayalı yöntemlerle kestirilen güvenilirlik katsayısı, farklı yetenek gruplarındaki bütün bireyler için aynı anlamı taşıırken, Örtük Özellikler Teorisi'nde, farklı yetenek düzeyleri için elde edilen ölçme sonuçlarının doğruluğu (precision) ve ölçme hataları ayrı ayrı kestirilebilmektedir. Örtük Özellikler Teorisi'nde test bilgi fonksiyonunu bütün yetenek düzeylerinde tek biçimli (uniform) bir dağılım vermesi durumunda, her bir yetenek düzeyi için elde edilen test bilgi fonksiyonlarının ortalamasının alınması suretiyle tek bir güvenilirlik katsayısının (marginal reliability) elde edilmesi de mümkün görünmektedir (Da Ayala, 1993). Harvey ve Hammer (1999)'ın da belirttiği gibi, Örtük Özellikler Teorisi, bütün ranjlardaki puanlar bakımından her bir madde ve ölçek düzeyinde farklı yetenek düzeyleri için madde ve test düzeyinde güvenilirliği verir.

Problem Durumu

Eğitim ve psikolojideki ölçmelerde, insan özellikleri konu edildiğinden, bu alanlardaki ölçmelerde kullanılan ölçme araçlarının geliştirilmesi, psikometrinin önemli bir konu alanı içinde yer almaktadır. Ölçme araçlarının geliştirilmesinde kullanılan yöntemlerin hepsinde amaç, uygun maddeleri seçerek istenilen nitelikte testler meydana getirmektir. Test geliştirme çalışmalarında, testin ve onu oluşturan maddelerin özellikleri (test ve madde istatistikleri) ile bu iki grup özelliği birbirine bağlayan bağıntılardan yararlanır. Bu nedenle de test geliştirme çalışmaları büyük ölçüde madde istatistiklerine dayanmaktadır. Madde istatistikleri, test geliştirme adımlarından madde analiz çalışmalarıyla belirlenmektedir. Bu durum madde istatistiklerinin testin oluşturulması sırasında hazır olmasını gerektirir. Bu gereklilik test geliş-

firmada, deneme çalışmasının yerine getirilmesiyle mümkün olmaktadır (Baykul, 1991). Test geliştirilirken madde yazım çalışmaları tamamlandıktan sonra maddelerin özelliklerini kestirmek için deneme uygulaması yaparak, testi oluşturan maddelerin yapısı hakkında bilgi edinmeye çalışılır. Yani istenen nitelikte bir test geliştirebilmek için madde analiz çalışmalarında deneme uygulaması yapılmasına ihtiyaç duyulmaktadır. Fakat deneme uygulamasının yapılmasının sakıncalı olduğu bazı durumlar vardır. Baykul (1991), kullanılan testin gizli olduğu ve gizliliğin korunması gerektiği bazı hallerde testin bütününe veya bazı maddelerin gizli kalmasının sık rastlanan bir durum olduğunu belirtmiştir. Gizliliğin korunmadığı veya böyle bir endişenin var olduğu durumlarda deneme yapmaktan kaçınılır. Örneğin, yükseköğretime giriş sınavlarında veya Milli Eğitim Bakanlığı'nca yapılan parasız yatılı, Anadolu ve Fen liselerine giriş sınavlarında gizliliğin korunması nedeniyle deneme uygulaması yapılmamaktadır. Deneme uygulamasının yapılamadığı durumlarda, madde istatistikleri grup özelliklerine göre belirli bir yaklaşıklıkla tahmin edilebilirse, istenen özelliklere bir ölçüde sahip testler yapma olanağı elde edilmiş olur. Baykul (1991)'un da belirttiği gibi, denemenin yapılamadığı durumlarda, grubu tanıyan kişiler tarafından maddelerin güçlük ve ayıricılık gücü indeksleri tahmin edilebilir. Bu tahminlerin hataları hesaplanarak madde güçlük indeksleri ve ayıricılık gücü indeksleri daha az hata ile kestirebilir ve kestirilen bu istatistikler deneme sonuçları yerine kullanılabilir.

Bu araştırmada, deneme uygulaması yapılamayan durumlarda uzman tahminlerine göre madde özellikleri tahmin edilerek Klasik Test Teorisi ve Örtük Özellikler Teorisi'nin iki parametrelili logistik modeline dayalı hesaplanan test güvenilirliğini tahmin etme gücünün nasıl olduğu belirlenmeye çalışılacaktır. Deneme uygulaması yapılamayan durumlarda madde istatistiklerini tahmin ederek, test istatistiklerinin nasıl ve hangi yollarla kestirilebileceği konusunda Klasik Test Teorisi'ne dayalı test geliştirme çalışmalarına rastlanmıştır. Fakat günümüzde test geliştirme çalışmalarında gittikçe artan bir önem kazanan Örtük Özellikler Teorisi'ne dayalı ve bu iki teorisinin karşılaştırılmasına yönelik çalışmalara literatürde rastlanmamıştır. Bu nedenle bu çalışmanın deneme uygulaması yapılmayan durumlardaki test geliştirme çalışmalarına önemli bir katkı getireceği düşünülmektedir.

Problem Cümlesi

Deneme uygulaması yapılamayan durumlarda madde özelliklerinin uzman tahminlerine dayalı kestirimlerinin

Klasik Test Teorisi ve Örtük Özellikler Teorisi'nin iki parametrelili logistik modeline dayalı olarak hesaplanan test güvenilirliğini tahmin etme gücü nasıldır?

Alt Problemler

1. Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı matematik alt testinin uzman tahminlerine dayalı kestirimlere göre hesaplanan testin güvenilirliği ile Klasik Test Teorisi'nde madde istatistiklerine göre hesaplanan testin güvenilirliği arasında fark var mıdır?

2. Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı matematik alt testinin uzman tahminlerine dayalı kestirimlere göre hesaplanan test bilgi fonksiyonu ile Örtük Özellikler Teorisi'ne göre hesaplanan test bilgi fonksiyonları nasıl bir değişim göstermektedir?

Yöntem

Araştırmanın Türü

Bu çalışmada, deneme uygulaması yapılmayan durumlarda madde özelliklerinin uzman tahminlerine dayalı kestirimlerinin Klasik Test Teorisi ve Örtük Özellikler Teorisi'nin iki parametrelili logistik modeline dayalı olarak hesaplanan test güvenilirliğini tahmin etme gücünün nasıl olduğu incelenmeye çalışılmıştır. Yapılan çalışmada, örneklem bilgilerinden bir evrene genelleme yapma amacı güdülmüştür. Bu araştırma, temel bir araştırma niteliğindedir.

Araştırmanın Yapılacağı Grup

Araştırma, Milli Eğitim Bakanlığı'nın 1999-2000 öğretim yılı Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı'na giren rasgele olarak seçilmiş 9914 öğrenci üzerinde yapılmıştır.

Verilerin Toplanması

Araştırmanın verilerini, Milli Eğitim Bakanlığı'nın 1999-2000 öğretim yılı Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavına giren ilköğretim öğrencilerinin Matematik alt testlerine vermiş oldukları cevapları oluşturmaktadır.

Temel Varsayımların Test Edilmesi

Örtük Özellikler Teorisi'ne göre matematik alt testinin madde parametre değerleri (b_g ve a_g) hesaplanmadan önce, bu teorisinin temel varsayımlarının karşılanıp karşılanmadığı SSPS paket programından yararlanılarak aşağıda belirtildiği gibi test edilmiştir.

Normal Dağılım

Örtük Özellikler Teorisi'nde bireyin ölçülen özelliği-

nin evrende normal dağılım gösterdiği varsayılmaktadır. Araştırmada matematik alt test puan dağılımını betimleyen istatistikler Tablo 1'de verilmiştir.

Tablo 1

Matematik Testi Puan Dağılımına Ait Betimsel İstatistikler

İstatistikler	
Aritmetik Ortalama	8,52
Varyans	19,32
Standart Kayma	4,40
Basıklık	0,80
Çarpıklık	0,94
Mod	6
Ortanca	8
Madde Sayısı	25
N	9914

Tablo 1 incelendiğinde matematik alt testinin basıklık katsayısının 0 (sıfır)'a yakın pozitif bir değer olması, dağılımın normale yaklaşmakla birlikte sivrilme gösterdiğini ortaya koymaktadır. Çarpıklık katsayısına bakıldığında bu katsayının 0'a yakın pozitif bir değer aldığı görülmektedir. Çarpıklık katsayısının pozitif değer olması, dağılımın sağa çarpık bir ivme gösterdiğini ortaya koymaktadır. Ayrıca matematik alt test puanlarındaki aritmetik ortalama, mod ve ortanca değerlerinin birbirine yakın değerler almış olması da puan dağılımının simetrik olduğu hakkında bilgi vermektedir. Yapılan çalışma bir simülasyon çalışması olmadığından, seçme sınavlarının yapısı gereği dağılım üzerinde bir değişiklik yapılmadan çalışmaya devam edilmiştir.

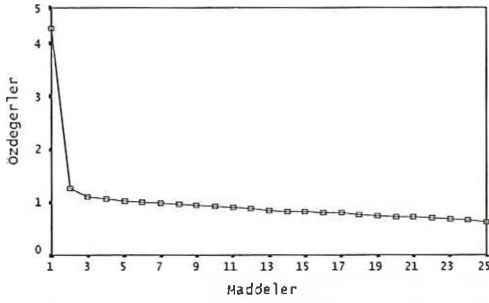
Tek Boyutluluk

Tek boyutluluğun test edilmesinde Hamblton ve Swaminathan (1985, 21) faktör analizi sonucuna bakarak, maddeler eğer bir boyutta toplanıyorsa, testin tek boyutlu olduğunu söylenebileceğini belirtmişlerdir. Araştırmada matematik alt testinin tek boyutlu olup olmadığını test etmek için, faktör analizi tekniğinden yararlanılmıştır.

Matematik testi faktör analizi sonuçları incelendiğinde, 25 maddeden 20'sinin yani 4., 17., 20., 23 ve 24. madde dışındaki bütün maddelerin birinci boyutta faktör yükünün en büyük olduğu belirlenmiştir. Ayrıca tek boyutluluk için özdeğerlere bakıldığında; toplam açıklanan varyansın %17.084'ünün birinci faktör, %5.060'ının ikinci faktör yükü ve % 4.370'inin üçüncü faktör yükü tarafından açıklandığı görülmüştür. Belirlenen faktörlerin öz-

değerleri birinci faktör için 17.084 ikinci faktör için 5.060 olduğundan; birinci faktör yüküne ait özdeğerin (eigen value) ikinci faktör yüküne ait özdeğerden çok farklı olmasının ve ikinci faktör yüküne ait özdeğerin diğer faktör yüklerine ait özdeğerlerden çok farklı olmaması ölçülen özelliğin tek boyutlu olduğunun bir göstergesidir. Yapılan faktör analizi sonuçlarına göre özdeğerlerin grafiği Şekil 1'de verilmiştir.

Faktör analizi sonucunda birinci faktör yükünden ikinci faktör yüküne ait özdeğerlerdeki ani düşüş, matematik testinde yer alan maddelerin birinci boyutta yeterli büyüklükte faktör yüküne sahip olduğunu göstermektedir.



Şekil 1. Matematik Testindeki Maddelerin Özdeğerlerinin Grafiği

Yerel Bağımsızlık

Örtük Özellikler Teorisi'nin karşılanması gereken üçüncü temel varsayımı yerel bağımsızlıktır. Lord (1980) tek boyutluluk varsayımının karşılanması durumunda, belli bir yetenek düzeyindeki bireylerin maddelere vermiş oldukları tepkiler arasındaki korelasyonun sıfır olduğunu ve dolayısıyla tek boyutlu olan ölçme araçlarının aynı zamanda yerel bağımsızlık varsayımını da karşıladığını ileri sürmektedir. Bu anlamda tek boyutluluk ve yerel bağımsızlık birbirinin eşdeğeri olarak kullanılmaktadır. Buna göre, yapılan faktör analizi sonucunda matematik alt testinin tek boyutluluk göstermesi, alt testi oluşturan maddelerin birbirinden yerel bağımsız olduğu şeklinde yorumlanmıştır.

Verilerin Analizi

Araştırmanın verilerinin çözümlenmesinde öncelikle gözlenen değerleri belirlemek için Klasik Test Teorisi ve Örtük Özellikler Teorisi'ne göre ayrı ayrı madde parametre değerleri hesaplanmıştır. Öğrencilerin, matematik alt testine ilişkin vermiş oldukları cevaplardan yararlanarak Klasik Test Teorisi'ne göre madde parametre değerleri yani madde güçlüğü ve madde ayırıcılık gücü (p_j ve

r_{jx}) indeksleri ITEMAN for Windows 3.1 paket programı yardımı ile hesaplanmıştır.

Örtük Özellikler Teorisi'nin temel varsayımlarına yönelik yukarıdaki açıklamalardan sonra matematik alt testine ait madde parametre değerleri yani madde güçlüğü ve madde ayırıcılık gücü (a_j ve b_j) indeksleri Örtük Özellikler Teorisi'ne göre Bilog 3.0 paket programı yardımı ile hesaplanmıştır. Matematik alt testine ait madde parametre değerleri Klasik Test Teorisi ve Örtük Özellikler Teorisi'ne göre ayrı ayrı hesaplandıktan sonra uzman tahminlerinin kendilerine verilen metrik ölçek üzerinde her bir maddenin parametre değerlerini her iki teoriye göre ayrı ayrı kestirmeleri istenmiştir. Uzman tahmin değerlerini almak için hazırlanan farklı iki form öncelikle Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı matematik alt testini hazırlayan 6 uzmana verilmiş, fakat araştırma kapsamında uzman sayısı yetersiz bulunarak bu sınava hazırlık kursları veren farklı dersanelerdeki 10 matematik öğretmeni de çalışmaya dahil edilmiştir. Uzman tahminlerine dayalı madde parametrelerini belirlemek için, uzmanlardan araştırmacı tarafından hazırlanan, 10 cm'lik bir cetvel üzerinde her bir test maddesinin test uygulanmadan önceki madde parametre değerlerini, yani madde güçlük ve ayırt edicilik gücü indeks değerlerini tahmini olarak işaretlemeleri istenmiştir. Klasik Test Teorisi'ne göre hazırlanan madde parametrelerinin yer aldığı formda, madde güçlük indeksi olan p_j parametresi ve madde ayırıcılık gücü indeksi olan r_{jx} parametrelerinin ne anlama geldiği uzmanların anlayabileceği açıklıkta ifade edilmiştir. Örtük Özellikler Teorisi'ne göre hazırlanan formda ise madde güçlük indeksi olan b_j parametresi ve madde ayırıcılık gücü indeksi olan a_j parametresi açıklanmıştır. Bu açıklamaların yetersiz olabileceği düşüncesiyle uzmanların hem klasik hem de örtük özelliklerdeki madde parametrelerini tutarlı bir şekilde tahmin etmelerini sağlamak için araştırmacı tarafından ayrı bir eğitim programı hazırlanmıştır. Bu eğitim programı içerisinde testi hazırlayan 6 uzmana bir ve 10 uzmana da üç haftalık farklı gün ve saatlerde olmak üzere toplam dört haftalık eğitim verilerek araştırmacı tarafından her iki teoride yer alan bu parametreler anlatılmış ve uzmanlarla madde parametreleri önceden belirlenmiş farklı yıllara ait örnek test maddeleri üzerinde bireysel çalışmalar yapılmıştır. Farklı yıllara ait örnek maddeler üzerinde çalışılarak araştırmacı tarafından konunun anlaşıldığı belirlendikten sonra uzmanlardan 1999-2000 yılına ait matematik alt testindeki maddelerin madde güçlük ve ayırt edicilik gücü indeks değerlerini

metrik ölçek üzerinde işaretlemeleri istenmiştir. Uzmanlardan elde edilen metrik ölçek üzerindeki bu tahmin değerleri 10 cm' lik bir cetvel yardımıyla ölçülmüş ve her bir uzmana ait bu iki teori için ayrı ayrı madde parametrelerinin uzman tahmin değerleri belirlenmiştir. Bir grup uzmanın her birisi için kestirme hatasının, o uzman grubuna ait tahmin değerlerinin ortalaması alınarak hesaplanan kestirme hatasından büyük olduğu, bir uzmana ait güvenilirliğin toplam tahmin güvenilirliğinden daha düşük ve buna bağlı olarak da hata varyansının büyük olduğu yapılan çalışmalarda ortaya konmuştur (Sezer, 1992). Bu nedenle tek tek her bir uzmana ait kestirimde bulunmanın çalışmanın amacına da uygun olmadığı düşünüldüğünden, araştırmada 16 uzmana ait madde tahmin değerlerinin ortalaması alınmıştır.

Matematik alt testine ait farklı iki teoriye göre hesaplanan gözlenen değerler ve uzman tahmin değerleri belirlendikten sonra uzman tahminlerine dayalı kestirimlere göre testin güvenilirliğini belirlemek için, madde istatistiklerinden yararlanılarak test istatistikleri hesaplanmıştır. Uzman tahminlerine dayalı kestirimlere göre hesaplanan testin güvenilirliği ve Klasik Test Teorisi' nde madde istatistiklerine göre hesaplanan testin güvenilirliğini belirlemek için KR-20 güvenilirlik katsayısı eşitliğinden yararlanılmıştır.

$$KR - 20 = \frac{K}{K - 1} \left[1 - \frac{\sum_{j=1}^k p_j q_j}{S_x^2} \right] \quad (2)$$

K = Testteki madde sayısı

p_j = j . Maddenin güçlük indeksi

q_j = $1 - p_j$

S_x^2 = Test puanlarının varyansı

Araştırmanın 1. alt probleminde yer alan uzman tahminlerine dayalı kestirimlere göre hesaplanan testin güvenilirliği ile Klasik Test Teorisi' nde madde istatistiklerine göre hesaplanan testin güvenilirliği arasındaki farkın test edilmesinde ise Fisher'ın z dönüşüm formülünden yararlanılmıştır (Akhun 1988, 18).

$$Z = \frac{z_K - z_G}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \quad (3)$$

Z_K : Kestirim değerlerinin birim normal cinsinden değeri

Z_G : Gözlenen değerlerin birim normal cinsinden değeri

n_1, n_2 : Testin madde sayısı

Araştırmanın 2. alt probleminde yer alan matematik alt testinin uzman tahminlerine dayalı kestirimlere göre hesaplanan test bilgi fonksiyonu ile Örtük Özellikler Teorisi'ne göre hesaplanan test bilgi fonksiyonlarının nasıl bir değişim gösterdiğini belirlemek için öncelikle iki parametrelili logistik dağılım fonksiyonu giriş bölümünde açıklanan 1. eşitlik yardımıyla hesaplanmıştır.

İki parametrelili logistik dağılım fonksiyonu hesaplandıktan sonra uzman tahmin değerleri ve Örtük Özellikler Teorisi'ne göre ayrı ayrı elde edilen parametre değerlerinden yararlanılmış ve her bir maddeye ait madde bilgi fonksiyonları ve test bilgi fonksiyon değeri farklı yetenek düzeyleri için hesaplanmıştır. Madde bilgi fonksiyonları ve test bilgi fonksiyonu θ 'nın -2 ve +2 aralığındaki farklı yetenek düzeyleri için 4. eşitlikte yer alan formülden yararlanılarak hesaplanmıştır. Ayrıca θ 'nın -2 ve +2 aralığındaki farklı yetenek düzeyleri için standart hata değerleri 5. eşitlikten yararlanılarak hesaplanmıştır (Hambelton 1985, 104).

$$I(\theta) = \sum_{i=1}^n \frac{P_i'(\theta)^2}{P_i(\theta)Q_i(\theta)} \quad (4)$$

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (5)$$

Bulgular ve Yorum

Bu araştırmada, deneme uygulaması yapılamayan durumlarda uzman tahminlerine göre madde özellikleri tahmin edilerek Klasik Test Teorisi ve Örtük Özellikler Teorisi'nin iki parametrelili logistik modeline dayalı hesaplanan test güvenilirliğini tahmin etme gücünün nasıl olduğu üzerinde durulmaktadır.

Araştırmada alt problemlere ait bulguları açıklamadan önce, bu bulguların elde edilmesinde kullanılan uzman tahmin değerlerinin güvenilirliğini belirlemenin uygun olacağı düşünülmüştür.

Uzman tahmin değerlerinin güvenilirliğini belirlemek için 16 uzmanın bu iki teori için ayrı ayrı belirledikleri madde parametre değerleri için varyans analizi uygulanmıştır. Böylelikle her bir madde parametresine ilişkin uzman tahminleri arasındaki tutarlılık belirlenmiştir. Varyans analizi sonucunda, 16 uzmanın Klasik Test Teorisi'ndeki madde güçlük indeks değerleri ve Örtük Özellikler Teorisi'ndeki b_g parametrelerine ilişkin tahmin değerleri arasında manidar fark bulunmazken, Klasik Test

Teorisi' ndeki madde ayırıcılık gücü indeks değerleri ve Örtük Özellikler Teorisi'ndeki a_g parametrelerine ilişkin tahmin değerleri arasında manidar fark bulunmuştur ($p < .01$). Yani her iki teorinin madde güçlük indeks parametrelerine ilişkin tahmin değerlerinde uzmanlar arası tutarlılık görülmüş; fakat madde ayırıcılık gücü indekslerine ilişkin tahmin değerlerinde uzmanlar arasında farklılık gözlenmiştir.

1. Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı matematik alt testinin uzman tahminlerine dayalı kestirimlere göre hesaplanan testin güvenilirliği ile Klasik Test Teorisi' nde madde istatistiklerine göre hesaplanan testin güvenilirliği arasında fark var mıdır?

Bu alt probleme cevap bulmak için öncelikle 16 matematik uzmanının madde parametrelerine ilişkin olarak vermiş oldukları kestirim değerleri dikkate alınarak testin güvenilirliğini hesaplamak için gerekli olan test parametreleri yani testin ortalaması ve varyansı hesaplanmıştır. Test parametreleri hesaplandıktan sonra uzman tahminlerine dayalı testin KR-20 güvenilirlik katsayısı hesaplanmış ve 0,49 bulunmuştur. Daha sonra Klasik Test Teorisi madde istatistiklerine göre testin KR-20 güvenilirlik katsayısı hesaplanmış ve 0,77 bulunmuştur.

Uzman tahminlerine ve Klasik Test Teorisi' ne göre hesaplanan KR-20 güvenilirlik katsayıları arasında anlamlı bir farkın olup olmadığını belirlemek için Fisher'in z dönüşüm formülünden yararlanılmış ve 0,05 düzeyinde anlamlı bir fark bulunmamıştır. KR-20 güvenilirlik katsayıları arasında anlamlı bir farkın bulunmamış olması, uzman tahmin değerlerinden elde edilen güvenilirlik katsayısının gözlenen değerlerden elde edilen güvenilirlik katsayısına eşit olduğu anlamına gelmemelidir.

2. Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı matematik alt testinin uzman tahminlerine dayalı kestirimlere göre hesaplanan test bilgi fonksiyonu ile Örtük Özellikler Teorisi'ne göre hesaplanan test bilgi fonksiyonları nasıl bir değişim göstermektedir?

Bu alt problemde matematik alt testinin uzman tahminlerine dayalı kestirimlere göre hesaplanan test bilgi fonksiyonu ile Örtük Özellikler Teorisi'ne göre hesaplanan test bilgi fonksiyonlarının nasıl bir değişim gösterdiğini belirlemeye çalışılmıştır. Bunun için öncelikle uzmanların matematik alt testindeki 25 maddeye ait vermiş oldukları tahmini b_g ve a_g parametre ortalamalarından yararlanılarak θ 'nın -2 ve +2 aralığındaki farklı yetenek düzeyleri için iki parametrelili logistik dağılım fonksiyonu hesaplanmıştır. Aynı işlem Örtük Özellikler Teorisi'nin iki parametrelili logistik modeline dayalı olarak he-

saplanan gözlenen b_g ve a_g parametre değerleri kullanılarak θ 'nın -2 ve +2 aralığındaki farklı yetenek düzeyleri tekrar hesaplanmıştır. Uzman tahmin değerlerine ve Örtük Özellikler Teorisi'ndeki gözlenen değerlere göre farklı yetenek düzeyleri için iki parametrelili logistik dağılım fonksiyonu her bir madde için hesaplandıktan sonra, her bir maddeye ait madde bilgi fonksiyonları ve test bilgi fonksiyon değerleri farklı yetenek düzeyleri için hesaplanmıştır.

Madde bilgi fonksiyonları, ölçeğin bütünü ile ölçülmek istenen özelliğin doğru (güvenilir) bir şekilde ölçülmesine her bir maddenin sağladığı katkıyı belirlemesi bakımından önemlidir. Madde bilgi fonksiyonlarının toplamı, testin bilgi fonksiyonunu verir. Madde bilgi fonksiyonlarının değerleri maddenin "a" ve "b" değerleri ile yakından ilişkilidir. "a" değeri yüksek olan maddelerin sahip oldukları madde bilgi fonksiyonu değerlerinin de yüksek olduğu gözlenmektedir. Maddelerin "b" değerleri ise maddenin konumu (location) hakkında bilgi vermektedir. "b" değeri -1 ve altında olan maddelerin düşük yetenek düzeyinde, +1 ve üzerinde olan maddelerin ise yüksek yetenek düzeyinde daha fazla bilgi verdiği gözlenmektedir. Buna ek olarak test bilgi fonksiyonu değerleri azaldıkça standart hata değerleri artmakta, test bilgi fonksiyonu değerleri arttıkça standart hata değerleri de azalmaktadır.

Uzman tahmin değerleri ve Örtük Özellikler Teorisi'ndeki gözlenen değerlere göre θ 'nın -2 ve +2 aralığındaki farklı yetenek düzeyleri için madde bilgi fonksiyonlarının toplamından oluşan test bilgi fonksiyonları ve standart hata değerleri Tablo 2 ve 3'te verilmiştir.

Tablo 2

Uzman Tahmin Değerlerine Göre Farklı Yetenek Düzeyleri İçin Kestirilen Test Bilgi Fonksiyonu Değerleri ve Standart Hataları

Yetenek Düzeyleri	Test Bilgi Fonksiyonları I (θ)	Standart Hata SE (θ)
-2,0	2,22	0,67
-1,5	4,33	0,48
-1,0	7,98	0,35
-0,5	13,19	0,28
0	17,84	0,24
0,5	18,10	0,24
1,0	12,96	0,28
1,5	6,70	0,39
2,0	2,88	0,59

Tablo 3
Örtük Özellikler Teorisi'ne Göre Farklı Yetenek Düzeyleri İçin Kestirilen Test Bilgi Fonksiyonu Değerleri ve Standart Hataları

Yetenek Düzeyleri	Test Bilgi Fonksiyonları I (θ)	Standart Hata SE (θ)
-2,0	1,26	0,89
-1,5	1,89	0,73
-1,0	2,70	0,61
-0,5	3,61	0,53
0	4,39	0,48
0,5	4,72	0,46
1,0	4,50	0,47
1,5	3,86	0,51
2,0	3,09	0,57

Uzman tahmin değerleri ve Örtük Özellikler Teorisi'ndeki gözlenen değerlere göre madde bilgi fonksiyonlarının toplamından oluşan test bilgi fonksiyonlarının θ 'nın -2 ve +2 aralığındaki farklı yetenek düzeyleri üzerindeki dağılımı Şekil 2'de verilmiştir.



Şekil 2. Uzman Tahmin Değerleri ve Örtük Özellikler Teorisine Göre Hesaplanan Test Bilgi Fonksiyon Grafikleri

Uzman tahmin değerlerine göre Tablo 2'de yer alan farklı yetenek düzeyleri için kestirilen test bilgi fonksiyon değerleri ve Şekil 2'deki test bilgi fonksiyonları dağılımı incelendiğinde, test bilgi fonksiyonu değerlerinin -0,5 ve altında yetenek düzeyine sahip olanlar ile +1 ve üstünde yetenek düzeyine sahip olanlar için düşük, -0,5 ile +1 aralığında yetenek düzeyine sahip olanlar için yüksek olduğu gözlenmektedir. Yani uzman tahmin değerlerine göre, test bilgi fonksiyonu değerlerinin -2 yetenek düzeyine sahip olan bireylerden 0,5 yetenek düzeyine sahip olan bireylere doğru gittikçe arttığı, 0,5 yetenek düzeyine sahip olan bireylerden +2 yetenek düzeyine

sahip olan bireylere doğru gittikçe düştüğü söylenebilir.

Örtük özellikler teorisine göre Tablo 3'te yer alan farklı yetenek düzeyleri için kestirilen test bilgi fonksiyon değerleri ve Şekil 2'deki test bilgi fonksiyonları dağılımı incelendiğinde, test bilgi fonksiyonu değerlerinin bütün yetenek düzeyleri için özellikle de -0,5'in altında yetenek düzeyine sahip olanlar için çok düşük olduğu gözlenmiştir. Fakat aynı şekilde, Örtük Özellikler Teorisi'ne göre, -2 yetenek düzeyine sahip olan bireylerden 0,5 yetenek düzeyine sahip olan bireylere doğru gittikçe test bilgi fonksiyon değerlerinin arttığı, 0,5 yetenek düzeyine sahip olan bireylerden +2 yetenek düzeyine sahip olan bireylere doğru gittikçe düştüğü gözlenmiştir.

Buna göre uzman tahmin değerlerine ve Örtük Özellikler Teorisi'ne göre farklı yetenek düzeyleri için kestirilen test bilgi fonksiyon değerlerinin -2 ve +0,5 yetenek düzeyine sahip olan bireyler için artma gösterdiği, +0,5 ve +2 yetenek düzeyine sahip olan bireyler içinde test bilgi fonksiyon değerlerinde düşme olduğu gözlenmiştir. Fakat aynı yetenek düzeyi aralıklarında artma ve düşme göstermiş olmasına rağmen, uzman tahmin değerlerine göre kestirilen test bilgi fonksiyon değerlerinin Örtük Özellikler Teorisi'ne göre elde edilen test bilgi fonksiyon değerlerinden büyük farklılık gösterdiği gözlenmiştir. Buna da uzmanların kestirimde bulunurken a_g parametre değerlerini olduğundan yüksek bir değerde kestirimde bulunmaları neden olmuş olabilir. Buradan elde edilen değerlerle hesaplanan test bilgi fonksiyon değerlerinin Örtük Özellikler Teorisi'nden elde edilen test bilgi fonksiyon değerlerinden daha yüksek değerler verdiği söylenebilir.

Bu nedenle bu sonuçlar dikkate alınarak farklı yetenek düzeylerine sahip olan bireyler için uzman tahmin değerlerine dayalı olarak elde edilen parametre değerlerinin Örtük Özellikler Teorisi'ne dayalı olarak elde edilen parametre değerlerinden daha fazla bilgi verdiği ve daha az hatayla kestirilebileceğini söylemek, bulgular bölümünün başında da ifade edildiği gibi uzman tahmin değerleri arasında tutarlılığın sağlanamaması da dikkate alındığında doğru olmayabilir.

Sonuçlar

Araştırmada, deneme uygulaması yapılamayan durumlarda uzman tahminlerine göre madde özellikleri tahmin edilerek Klasik Test Teorisi ve Örtük Özellikler Teorisi'nin iki parametreliliğine dayalı hesaplanan test güvenirliliğini tahmin etme gücünün nasıl olduğu belirlenmeye çalışılmıştır.

Araştırmada Örtük Özellikler Teorisi'nin iki parametrelili logistik modeline dayalı madde parametrelerinin analizine geçmeden önce bu teorinin temel varsayımlarının karşılanıp karşılanmadığı test edilmiştir. Örtük Özellikler Teorisi'nin temel varsayımları ile ilgili açıklamalardan sonra, aşağıda araştırma bulgularından elde edilen sonuçlara yer verilmiştir.

Araştırmada alt problemlere ait bulguları açıklamadan önce, bu bulguların elde edilmesinde kullanılan uzman tahmin değerlerinin güvenilirliğini belirlemek için 16 uzmanın her iki teori için ayrı ayrı belirledikleri madde parametre değerleri için varyans analizi uygulanmıştır. 16 uzmanın Klasik Test Teorisi'nden elde edilen madde güçlük indeksi (p_j) ve Örtük Özellikler Teorisi'nin iki parametrelili logistik modelinden elde edilen b_p parametrelerine ilişkin tahmin değerleri arasında manidar fark bulunmazken, Klasik Test Teorisi'ndeki madde ayırıcılık gücü indeksi ve Örtük Özellikler Teorisi'ndeki a_p parametrelerine ilişkin tahmin değerleri arasında manidar fark bulunmuştur ($p < .01$). Yani her iki teorinin madde güçlük indeksi parametrelerine ilişkin tahmin değerlerinde uzmanlar arası tutarlılık görülmüş, fakat madde ayırıcılık gücü indekslerine ilişkin tahmin değerlerinde uzmanlar arası tutarlılığa rastlanmamıştır.

Araştırmanın birinci alt problemünde Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı matematik alt testinin uzman tahminlerine dayalı kestirimlere göre hesaplanan testin güvenilirliği ile Klasik Test Teorisi'nde madde istatistiklerine göre hesaplanan testin güvenilirliği arasında anlamlı bir farkın olup olmadığı belirlemek için Fisher'in z dönüşüm formülünden yararlanılmış ve 0,05 düzeyinde anlamlı bir fark bulunmamıştır. KR-20 güvenilirlik katsayıları arasında anlamlı bir farkın bulunmuş olması, uzman tahmin değerlerinden elde edilen güvenilirlik katsayısının gözlenen değerlerden elde edilen güvenilirlik katsayısına eşit olduğu anlamına gelmemelidir. Elde edilen betimsel değerlere bakıldığında da gözlenen değerlerden hesaplanan güvenilirlik katsayı değerinin, uzman tahmin değerlerinden hesaplanan güvenilirlik katsayı değerine göre çok daha yüksek olduğu görülmektedir.

Araştırmanın ikinci alt problemünde ise, Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı matematik alt testinin uzman tahminlerine dayalı kestirimlere göre hesaplanan test bilgi fonksiyonu ile Örtük Özellikler Teorisi'ne göre hesaplanan test bilgi fonksiyonlarının nasıl bir değişim gösterdiği belirlenmeye çalışılmıştır.

Elde edilen bulgulara göre, uzman tahmin değerlerine ve Örtük Özellikler Teorisi'ne göre farklı yetenek düzey-

leri için kestirilen test bilgi fonksiyon değerlerinin -2 ve +0,5 yetenek düzeyine sahip olan bireyler için artma gösterdiği, +0,5 ve +2 yetenek düzeyine sahip olan bireyler içinde test bilgi fonksiyon değerlerinde düşme olduğu gözlenmiştir. Fakat aynı yetenek düzeyi aralıklarında artma ve düşme göstermiş olmasına rağmen, uzman tahmin değerlerine göre kestirilen test bilgi fonksiyon değerlerinin Örtük Özellikler Teorisi'ne göre elde edilen test bilgi fonksiyon değerlerinden büyük farklılık gösterdiği gözlenmiştir.

Buna da uzmanların kestirimde bulunurken a_p parametre değerlerini olduğundan yüksek bir değerde kestirimde bulunmalarının neden olduğu söylenebilir. Buradan elde edilen değerlerle hesaplanan test bilgi fonksiyon değerlerinin Örtük Özellikler Teorisi'nden elde edilen test bilgi fonksiyon değerlerinden daha yüksek değerler verdiği söylenebilir. Bu nedenle bu sonuçlar dikkate alınarak farklı yetenek düzeylerine sahip olan bireyler için uzman tahmin değerlerine dayalı olarak elde edilen parametre değerlerinin Örtük Özellikler Teorisi'ne dayalı olarak elde edilen parametre değerlerinden daha fazla bilgi verdiği ve daha az hatayla kestirilebileceğini söylemek, bulgular bölümünün başında da ifade edildiği gibi uzman tahmin değerleri arasında tutarlılığın sağlanamaması da dikkate alındığında doğru olmayabilir.

Kaynakça

- Akhun, İ. (1988). *İstatistiksel formüller ve tablolar*. Yayınevi belirsiz.
- Baker, B. F. (1977). Advances in item analysis. *Review of Educational Research*, 47, 151-178.
- Baykul, Y. (2000). *Eğitinde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.
- Baykul, Y. (1991). *Ön denemenin yapılamadığı hallerde madde güçlük indeksinin ve buna bağlı olan madde ve test istatistiklerinin kestirilmesi*. Malatya: Üniversite ve Çevre İlişkileri Sempozyumu.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern theory*. Cbs College Publishing.
- De Ayala, R. J. (1993). An introduction to polytomous item response theory models. *Measurement and Evaluation in Counseling and Development*, 25 (4), 172-189.
- Hambelton, J. & Swaminathan, M. (1985). *Item response theory: Principles and application*. Kluwer, Nijhoff Publishing.
- Harvey, R. J. & Allen, L. H. (1999). Item response theory. *Counseling Psychologist*, 27 (3), 353-374.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison Wesley Publishing Co.
- Lorge, I. & Kruglov, L. (1953). The improvement of estimates

- of test difficulty. *Educational and Psychological Measurement*, 13, 34-46.
- 1.E.B. Orta Öğretim Kurumları. (2000). *Öğrenci seçme ve yerleştirme sınavı test kitapçığı*. Milli Eğitim Bakanlığı Yayınları.
- ezer, S. (1992). *Ön deneme yapılamayan durumlarda madde güçlük ve ayıricılık gücü indekslerinin ve bunlara bağlı test istatistiklerinin kestirilmesi*, Yayımlanmamış Doktora Tezi, H.Ü. Sosyal Bilimler Enstitüsü, Ankara.
- tocking, M. L. (1997). Item response theory. In J. P. Keeves (ed.) *Educational research, methodology and measurement: An international handbook* (Second Edition). Elsevier Science Ltd.
- Turgut, M. F. (1997). *Eğitimde ölçme ve değerlendirme metodları*. Ankara: Yargıcı Matbaası.
- Turgut, M. F. & Baykul, Y. (1992). *Ölçekleme teknikleri*. Ankara: ÖSYM Yayınları.
- Geliş : 22 Ocak 2004
İnceleme : 26 Nisan 2004
Düzeltilme : 15 Haziran 2004
Kabul : 18 Ekim 2004