

## TÜRKİYE'DE UYGULANAN GENİŞ ÖLÇEKLİ TESTLERİN ÇOK BOYUTLULUĞUNUN ANALİZİ<sup>1</sup>

### THE ANALYSIS OF LARGE SCALE TESTS APPLIED IN TURKEY IN TERMS OF THEIR MULTIDIMENSIONALITY

Yeşim ÖZER ÖZKAN<sup>2</sup>

Meltem ACAR GÜVENDİR<sup>3</sup>

#### Özet

Test yapısını ampirik olarak değerlendiren yöntemler öncelikle testin boyutluluğu olmak üzere bazı yapısal özellikler hakkında bilgi sağlamalıdır. Test boyutluluğu testin yapı geçerliliği, test puanlarının hesaplanması ve raporlanması üzerinde doğrudan bir etkiye sahiptir. Bu çalışmanın amacı Öğrenci Başarılarının Belirlenmesi Sınavının (ÖBBS) dört alt testinin boyutlarını parametrik olmayan çok boyutluluk yöntemlerinden DIMTEST T istatistiği kullanılarak belirlemektir. Araştırma betimsel türde temel bir çalışmadır. Araştırmanın verileri 2002, 2005 ve 2008 yılı ÖBBS'nin Türkçe, matematik, fen ve teknoloji, sosyal bilgiler ve İngilizce testlerine verilen yanıtlardan elde edilmiştir. Bu çalışma kapsamında verinin boyutlu olup olmadığının test edilmesi amaçlandığı için DIMTEST T istatistiği kullanılmıştır. Hipotez testi sonucu 2002, 2005 ve 2008 de uygulanan ÖBBS'nin tüm alt testlerinin çok boyutlu olduğu belirlenmiştir. Bilişsel testlerde yer alan soruları çözmek için gerekli olan beceriler düşünüldüğünde, tek boyutluluğun sağlanmadığı görülmektedir.

*Anahtar kelimeler:* çok boyutluluk, DIMTEST, geniş ölçekli testler, yapı geçerliliği, tek boyutluluk.

#### Abstract

The methods that empirically evaluate test structure need to primarily provide information about test dimensionality and other structural characteristics. Test dimensionality has a direct impact on a test's structure validity, score calculation, and result recording. The purpose of this study is to identify the dimensions of four subtests that constitute the parts of Evaluation of Student Achievement Test (ÖBBS) by using DIMTEST T statistics which is a nonparametric multidimensionality method. This study is a descriptive one. The data in this study is obtained from the responses provided for the Turkish, Mathematics, Science and Technology, and Social Sciences tests in ÖBBS 2002. In order to test the multidimensionality of the data DIMTEST T statistics was used. As a result of the hypothesis test it was found that all the subtests in ÖBBS 2002, 2005, and 2008 were multidimensional. When the skills that are necessary to solve the questions in cognitive tests are considered, it can be noticed that unidimensionality was not established.

*Keywords:* multidimensional, DIMTEST, large scale tests, construct validity, unidimensional.

<sup>1</sup>Bu çalışma 5-7 Eylül 2013 tarihinde Eskişehir Osmangazi Üniversitesi tarafından düzenlenen 22. Ulusal Eğitim Bilimleri Kurultayı'nda sözlü bildiri olarak sunulmuştur.

<sup>2</sup>Yrd. Doç. Dr., Gaziantep Üniversitesi Gaziantep Eğitim Fakültesi, Eğitim Bilimleri Bölümü, yozer80@gmail.com

<sup>3</sup>Yrd. Doç. Dr., Trakya Üniversitesi Eğitim Fakültesi, Eğitim Bilimleri Bölümü, meltemacar@gmail.com

## Giriş

Testin yapısal nitelikleri olan güvenirlik, geçerlik ve kullanışlılığı ampirik olarak değerlendiren yöntemler öncelikle testin boyutluluğu hakkında bilgi sağlamalıdır (Tate, 2003). Geleneksel olarak testin boyutluluğu belirli bir veri kümesindeki madde cevapları arasındaki ilişkilerin nedenini açıklayan gizil değişkenlerin sayısı olarak tanımlanır (Camilli, Wang ve Fesq, 1995). Başka bir ifade ile boyutluluk yeteneğin bir yerel bağımsız ve aynı zamanda örtük değişken modeli oluşturabilmesi için gerekli olan en az sayıdaki boyutu ifade eder (Stout, Froelich ve Gao, 2001). Test boyutluluğu testin yapı geçerliliği, test puanlarının hesaplanması ve raporlaştırılması üzerinde doğrudan bir etkiye sahiptir.

Eğitimde ve psikolojide kullanılan testler farklı alt bölümlerden oluşmaktadır. Örneğin, matematik bilgisini ölçen bir test aritmetik ve geometri alt testlerinden oluşabilir. Genel yeteneği ölçen bir test ise sözel, sayı ilişkileri ve problem çözme gücü, şekiller arası ilişkiler gibi alt testlerden oluşabilir. (Sinharay, Haberman ve Puhan, 2007). Benzer şekilde cebirle ilgili bir sınavda eşitliğe dayalı kısa sorulu problemler ve günlük yaşam problemleri olarak iki tip problem sorulmuş olsun: ilk tipteki sorular öğrencilerden cebirsel ifadeleri gösterme becerisini, ikinci tip ise soru metnini okumayı, anlamayı ve cebirsel ifadelerle gösterme becerisini içermektedir. Bu durumda testte, okuma becerisi, çevirme becerisinden cebirsel ifadeleri gösterme becerisine kadar farklı bileşenler yer almaktadır.

Eğitimde ve psikolojide bir yapıyı ya da özelliği ölçmede veya testlerin geliştirilmesinde ve değerlendirilmesinde iki ölçme kuramı geliştirilmiştir. Bunlar Madde Tepki Kuramı (MTK) ve Klasik Test Kuramıdır (KTK). KTK ve Tek Boyutlu MTK modelleri tek boyutluluk varsayımı üzerine kurulmuştur.

Testlerde tek boyutluluk, ölçme sonuçlarının temelinde yer alan tek bir kavramsal yapının veya özelliğin varlığı olarak belirtilebilir (McDonald, 1981). Bu durum, testteki tüm maddelerin aynı özelliği ölçüyor olması demektir. Bu açıdan düşünüldüğünde tek boyutlu MTK modellerinin, k sayıda özellik ve bu özelliklerin her birinin en az bir madde tarafından temsil edildiği n maddeden oluşan bir test düşünüldüğünde, k tane özelliğin tümünü örten ve her biri bir özelliğe karşılık gelen, özelliklerden maddelere bir bağıntı tanımlanmış olacaktır. Diğer bir anlatımla tek boyutluluk, kişinin testteki performansını belirleyen faktörün testin ölçmeye çalıştığı faktör olmasıdır.

Bireyin performansını etkileyen birden fazla faktör olduğunda tek boyutluluk sağlanamaz; ancak kişinin testteki performansını belirleyen faktörün testin ölçmeye çalıştığı faktör olması esasına dayanan tek boyutluluk varsayımı test edilirken, bireyin test performansını etkileyen tek bir özelliğin olduğunu ispatlamanın olanaksız olduğu gözden

kaçırılmaması gereken bir durumdur (Hambleton ve Swaminathan, 1985). Tek boyutluluk, kişinin testteki performansını belirleyen faktörün testin ölçmeye çalıştığı faktör olmasıdır. Başka bir ifade ile yerel bağımsız ve monoton olan bir örtük yetenek modeli üretebilmek amacı ile yetenek ( $\Theta$ ) için gerekli en az boyutluluğun  $d$  olduğu durumda test  $d$ -boyutlu olarak değerlendirilir.  $d=1$  olduğu durumda test, tek boyutlu test olarak adlandırılır. Tek boyutlu bir test, bireylerin yalnızca tek bir örtük yeteneğini ölçmek için tasarlanmıştır (Gao, 1997).

Tek boyutluluğu değerlendirmek için en yaygın kullanılan yöntem faktör analizidir. Faktör analizi ile tek boyutluluğun testinde bazı problemler ile karşılaşılacağından, tetrakorik korelasyon matrisine dayalı faktör analizi iki kategorili veriler için önerilmektedir. Tek boyutluluk için maddeler arası tetrakorik korelasyon matrisinin tek boyutlu olması önemlidir; ancak tetrakorik korelasyon matrisi, korelasyon sınırları dışında değerler üretebildiği için her zaman pozitif tanımlı olmamaktadır. Bu da faktör analizinde problem yaratır. Tetrakorik korelasyon matrisine dayalı faktör analizinin yanı sıra KR-20 tek boyutluluğun değerlendirilmesinde kullanılmaktadır; ancak KR-20 test uzunluğu ve grup homojenliğinden etkilendiği için her zaman tek boyutluluğun bir ölçüsünü vermemektedir. Bir diğer yöntem ise maddeler arası korelasyon matrisinin özdeğerlerinin grafiğini incelemek ve grafikteki kırılma noktasına göre test boyutluluğuna karar vermektir. Bu yöntemlerin yanı sıra, test içerik bakımından alt testlere ayrılarak, madde parametreleri hem alt testlerden hem de tüm test için elde edilir ve iki gruptan elde edilen madde parametreleri karşılaştırılır (Hambleton, Swaminathan ve Rogers, 1991).

Tek boyutluluk yok ise “Lumbsden Yöntemi” ile test tek boyutluluğa indirgenebilir. Bu yöntemde test geliştirici, bir madde havuzu oluşturur. Bu maddelere ait görgül verilerle yapacağı faktör analizi sonucunda başat faktörü ölçmeyen maddeleri analiz dışında bırakarak analizi tekrarlar. Bu işlem yeterli sonuç alınıncaya kadar devam eder. Yöntem ayrıca ilk faktör varyansının ikinci faktör varyansına oranlanarak “tek boyutluluk indeksi” hesaplanabileceğini ifade etmiştir (Hambleton ve Swaminathan, 1985).

Testlerin veya testlerde bulunan maddelerin sadece tek bir örtük özelliği ölçtüğünü varsaymak ve bu varsayım altında ölçmeler yaparak birey hakkında karar vermek 1980’lerden sonra daha fazla tartışılır hale gelmiştir (Ackerman, 1989; Ansley ve Forsyth, 1985; Drasgow ve Parsons, 1983; Harrison, 1986; Way, Ansley ve Forsyth, 1986). Bu çalışmalar tek boyutluluk varsayımının karşılanmadığı durumlarda yetenek ve madde parametre tahminlerinin etkilendiğini ortaya koymuştur (Özer Özkan, 2012). Traub (1983) özellikle bilişsel testlerde yer alan soruları çözmek için gerekli olan beceriler düşünüldüğünde, tek boyutluluğun kuraldan fazlasını içermediğini, belirtmiştir. Eğer maddeler çeşitli yetenek

düzeylerinde ve cevaplayıcılar da bu becerilerin bir tanesinden fazla yeterlik seviyesinde farklılaşıyorsa bu etkileşim çok boyutluluk ile modellenmelidir (Ackerman, 1994). Bu durumlarda ise çok boyutlu yöntemler kullanılmaktadır.

Testin tek boyutluluğuna karşın testin temelinde yer alan kavramsal boyut birden fazla ise bu durumda örtük değişken modeli çok boyutlu olarak adlandırılır (Stout ve diğ., 2001). Örneğin, bir testte yer alan maddeler cevaplayandan birden fazla davranış göstermesini isteyebilir. Başka bir ifade ile test maddeleri birden fazla davranış içeriyor demektir. Bu durumda testin tek boyutluluk varsayımı sağlanmaz ve çok boyutlu modellerden söz edilebilir (Smith, 2009).

Çok boyutluluğun analizinde birçok yöntem kullanılmaktadır. Bu yöntemleri parametrik ve parametrik olmayan olmak üzere iki grupta toplanabilir. Bazı durumlarda parametrik MTK modellerinin başarısızlığı ve madde sayısının az olduğu durumlarda parametrik olmayan yöntemlerin daha iyi sonuçlar vermesi parametrik olmayan yöntemlere dayalı araştırmaların yapılmasını arttırmıştır (Junker ve Sijtsma, 2001). Parametrik olmayan yöntemler incelendiğinde bir testin boyutluluğunu belirlemek amacıyla Stout, Douglas, Junker ve Roussos (1993) tarafından üç yöntem geliştirilmiştir. Bu yöntemlerden boyutluluğun testi DIMTEST (Stout, 1987; Nandakumar ve Stout, 1993), hiyerarşik kümeleme analizi HCA/CCPROX (Roussos, Stout ve Marden, 1998) ve boyutluluğun DETECT indeksi ise DETECT (Zhang ve Stout, 1999a) programları ile test edilmektedir. Bu yöntemler testlerin çok boyutluluğunun değerlendirilmesinde birtakım avantajlar sunmaktadır (Mroch ve Bolt, 2006; Stout ve diğ., 2001; Tate, 2003; Van Abswoude, Van der Ark ve Sijtsma, 2004). Bu yöntemler kullanılarak yapılan boyutluluk değerlendirmesi, iki ana alana ayrılmaktadır: Bir testin tek boyutlu mu, yoksa çok boyutlu mu olduğunun değerlendirilmesi ve gerekli olması halinde, test öğelerinin çok boyutlu yapısının belirlenmesidir.

Bu analiz yöntemlerinden DIMTEST, genellikle farklı koşullardaki etkileri göstermede ve hipotezlerin testinde kullanılır. Birçok araştırmacı farklı durumlardaki boyutluluk değerlendirmesi için doğrusal faktör analizini uygun bulmamışlardır (Hulin, Drasgow ve Parson, 1983; McDonald, 1981). Froelich ve Habing (2003) tek boyutlu madde kümelerinin seçimi için DETECT ve HCA yöntemlerini önermişlerdir çünkü bu yöntemler doğrusal faktör analizinden daha güçlüdür. Özellikle Froelich (2001), DIMTEST süreci için bile alt testleri seçmek için kullanılan doğrusal faktör analizinin bazı çok boyutlu modellerde azalmış hipotez test gücü ürettiğini belirtmiştir.

DIMTEST'in temeli, çift madde koşullu kovaryanslardır. Diğer iki yöntem olan DETECT ve HCA/CCPROX bu çift madde koşullu kovaryansları kullanarak geliştirilmiştir

(Kim, 1994; Zhang ve Stout, 1999a; Roussos ve diğ., 1998). DETECT, ayrı kümelerdeki test maddelerini sınıflandırmak ve bir testin boyutluluğunun sayısını belirlemek için bir algoritma kullanır (Kim, 1994; Zhang ve Stout, 1999b). DIMTEST ve DETECT'e bakıldığında, her iki yöntem de parametrik olmayan yöntemlerdir. Çünkü bu yöntemler verinin boyutluluğunu araştırmak amacıyla geliştirilmişlerdir ve boyutlar, madde ortak değişkenlerinin random ağırlıklarında olduğu gibi bireysel farklılıkları tanımlar. Bu yöntemler heterojenliğe tanısal bir yaklaşım sunabilirler. DIMTEST ve DETECT madde ortak değişkeni kullanmazlar. Bu yüzden, bilginin bir bölümü veri analizinde kullanılmaz (Balázs, Hidegkuti ve Boeck, 2006). HCA/CCPROX yöntemi ise çift madde koşullu kovaryansları kullanarak hesaplayan yakın bir matrise dayalı test maddelerinin hiyerarşik kümeleme analizini düzenler (Roussos ve diğ., 1998). DETECT ve HCA/CCPROX yöntemlerine bakıldığında her iki yöntem de özünde açıklayıcıdır. Bu yöntemler DIMTEST yöntemi için alt testleri seçmede doğal bir tercihtir (Froelich, 2001).

Testlerin boyutluluğunun belirlenmesi üzerine Türkiye'de birkaç çalışma (Özbek Baştuğ, 2012 ve Özer Özkan, 2012) yapılmıştır. Bu çalışmalardan Özbek Baştuğ (2012) tarafından yapılan araştırma, Orta Okullar Arası Seçme Sınavının Sosyal Bilgiler alt testinin boyutlarını parametrik ve parametrik olmayan yöntemlerini birlikte kullanarak tanımlanmasını içerir. Çalışmada Sosyal Bilgiler alt testinin çok boyutlu olduğu sonucuna ulaşılmıştır. Bu doğrultuda, testin gizil çok boyutlu yapısını tanımlamak ve sonunda bu yapıya göre tek boyutlu olarak puanlamanın önemli olduğunu vurgulamıştır. Özer Özkan (2012) ise doktora tez çalışmasının bir aşamasında sekizinci sınıflara uygulanan ÖBBS'nin 25 maddelik çoktan seçmeli Türkçe ve matematik testlerinin boyutluluğunu analiz etmiş ve testlerin çok boyutlu olduğunu belirlemiştir.

Yurtdışı alanyazında testlerin boyutluluğunun analizi, boyutluluk analizinde kullanılan parametrik ve parametrik olmayan yöntemlerin test edilmesi ve bu yöntemlerin karşılaştırılması, yöntemlerin simülatif ve gerçek veri üzerinde uygulamaları üzerine yoğunlaşan çalışmaların mevcut olduğu görülmektedir. Bu çalışmalardan Balázs ve diğ. (2006), lojistik regresyon analizinde heterojenlik testini ele aldığı çalışmalarında, testlerin boyutluluğunun belirlenmesinde kullanılan DIMTEST, DETECT ve temel bileşenler analizi yöntemlerini kullanmışlardır. Böylelikle küçük veri setlerinde heterojenliğin testini farklı yöntemler kullanarak belirlemeye çalışmışlardır. Smith ve Roussos (2010), çalışmalarında Değerlendirme Alt Kümesi (AT) ve Bölüntüleme Alt Kümesi (PT) alt testlerinin seçiminde kullanılan yöntemleri araştırmışlardır. Hattie, Krakowski, Rogers ve Swaminathan (1996) sınavların bir bölümü üzerinde doğrusal faktör analizini kullanarak AT ve PT ayrımını

yapmışlardır. Froelich ve Habing (2003) ise çalışmalarında testlerdeki boyutluluğu belirlemek amacıyla DIMTEST performansını önemli ölçüde geliştiren koşullu kovaryans tabanlı iki aşamalı bir bölümlenme (ayrıştırma) yöntemi kullanmışlardır. Roussos ve diğ. (1998) ise çalışmalarında boyutlu olarak ayrılmış madde kümelerindeki ayrıştırılmış (bölümlenmiş) test maddeleri için yeni bir yaklaşımı tanıtmışlardır. Bu yaklaşımın temelini, hiyerarşik kümeleme analizi ile kullanılabilir yeni bir çift madde koşullu kovaryans tabanlı yakınlık ölçüsü olarak açıklamışlardır. Stout, Habing, Douglas, Kim, Roussos ve Zhang (1996) boyutluluğun değerlendirilmesinde kullanılan parametrik olmayan yöntemlerden biri olan DIMTEST, DETECT ve HCA/CCPROX yöntemleri üzerinde durarak bu yöntemleri koşullu kovaryansların tahmininde kullanmışlardır. Jang ve Roussos (2007) da benzer şekilde koşullu kovaryans tabanlı parametrik olmayan yöntemleri kullanarak Yabancı Dil İngilizce Testi (Test of English as a Foreign Language-TOEFL) sınavının boyutluluğunu değerlendirmişlerdir.

Seo, Rutgers ve Roussos (2008) Uluslararası okuma becerilerinde gelişim projesi (Progress in International Reading Literacy Study-PIRLS) 2001 okuma becerilerine açılımcı ve doğrulayıcı DIMTEST kullanarak çok boyutluluğunu test etmiştir. Çalışmada farklı okuma parçalarının kullanıldığı on kitapçıktan sekizi çok boyutlu ikisinin ise tek boyutlu olduğu sonucuna ulaşılmıştır. Geniş ölçekli testlerin çok boyutluluğunun test edilmesi üzerine yapılan bir diğer çalışma da Douglas, Kim, Roussos, Stout ve Zhang (1999) tarafından yapılmıştır. Araştırmada Hukuk Fakültesi Giriş Sınavının (Law School Admission Test- LSAT) 1991 Aralık, 1992 Haziran ve 1992 Ekim uygulamalarından elde edilen veriler kullanılmıştır. Yapılan analizler sonucunda veriler çok boyutlu olduğu belirlenmiştir.

Yurt içi ve yurtdışında yapılan çalışmalar incelendiğinde bilişsel testlerde yer alan soruları çözmek için gerekli olan beceriler düşünüldüğünde, tek boyutluluğun kuraldan fazlasını içermediğinden hareketle bir testin altında yatan gizil bileşenlerin belirlenmesi test sonuçlarına göre verilen kararların doğruluğu açısından önem kazanmaktadır. Bu çalışmanın amacı 2002, 2005 ve 2008 yıllarında uygulanan Öğrenci Başarılarının Belirlenmesi Sınavının (ÖBBS) alt testlerinin boyutluluğunun belirlenmesidir.

### **Yöntem**

Araştırmanın bu bölümünde çalışma verilerine ve bu verilere ilişkin betimsel istatistikler ile verilerin analizine yer verilmiştir. Çalışmada geniş ölçekli bir test verisinin boyutluluğunun belirlenmesi amaçlandığı için öncelikle kullanılan testlere ilişkin veriler sunulmuştur.

## Veri

Çalışmanın verileri ÖBBS 2002, 2005 ve 2008 yıllarında ilköğretim sekizinci sınıf öğrencilerine uygulanan düzey belirleme testlerine verilen cevaplardan tesadüfî örneklem alınarak elde edilmiştir. Yıllara göre alt testlere ilişkin betimsel istatistikler Tablo 1, Tablo 2 ve Tablo 3'te sunulmuştur.

Tablo 1.

*ÖBBS 2002 Alt Test Puanlarına İlişkin Betimsel İstatistikler*

	N	Soru Sayısı	Ortalama	Standart Sapma	Çarpıklık Katsayısı	Basıklık Katsayısı
Türkçe	7000	25	13.62	5.56	.15	-.93
Matematik	7000	25	10.56	5.22	.85	.16
Fen ve Teknoloji	7000	20	9.00	3.90	.50	-.25
Sosyal Bilgiler	7000	20	9.50	4.29	.28	-.76

Tablo 1 incelendiğinde, ÖBBS 2002'de öğrencilerin dört alt test içerisinde Türkçe testi puanlarının en yüksek, Fen ve Teknoloji test puanlarının ise en düşük olduğu görülmektedir. Çarpıklık katsayısının -1 ile +1 sınırları içinde kalması puanların normalden aşırı bir sapma göstermediği şeklinde yorumlanabilir (Büyüköztürk, Çokluk ve Köklü, 2010). Dört alt testte de çarpıklık ve basıklık katsayısı değerlerinin belirtilen sınırlar içinde olması verilerin oluşturduğu dağılımın simetriğe çok yakın bir şekil aldığı göstergesi olarak kabul edilebilir.

Tablo 2.

*ÖBBS 2005 Alt Test Puanlarına İlişkin Betimsel İstatistikler*

	N	Soru Sayısı	Ortalama	Standart Sapma	Çarpıklık Katsayısı	Basıklık Katsayısı
Türkçe	7000	25	14.66	5.89	-.09	-1.08
Matematik	7000	25	11.20	5.48	.69	-.31
Fen ve Teknoloji	7000	20	9.03	4.10	.45	-.46
Sosyal Bilgiler	7000	20	9.91	4.53	.14	-.98

Tablo 2'deki betimsel istatistikler incelendiğinde öğrencilerin ÖBBS 2002 sonuçlarına benzer şekilde Türkçe testi puanlarının en yüksek, Fen ve Teknoloji test puanlarının ise en düşük olduğu görülmektedir. Tüm test ve alt boyutlara ilişkin çarpıklık değerleri incelendiğinde elde edilen sonuçların -1 ile +1 arasında olması verilerin oluşturduğu dağılımın normalden aşırı sapma göstermediği şeklinde yorumlanabilir.

Tablo 3.

*ÖBBS 2008 Alt Test Puanlarına İlişkin Betimsel İstatistikler*

	N	Soru Sayısı	Ortalama	Standart Sapma	Çarpıklık Katsayısı	Basıklık Katsayısı
Türkçe	3286	25	15.64	6.12	-.23	-1.14
Matematik	3286	25	11.53	6.08	.60	-.67
Fen ve Teknoloji	3286	20	9.74	4.50	.38	-.78
Sosyal Bilgiler	3286	20	10.14	4.49	.02	-1.03
İngilizce	3286	20	7.80	4.00	.48	-.35

Tablo 3 incelendiğinde ise 2002 ve 2005 yıllarında benzer şekilde Türkçe testi puanlarının en yüksek olduğu görülmektedir. Fakat iki yıldan farklı olarak test puanları en düşük dersin İngilizce olduğu söylenebilir.

## Verilerin Analizi

Bu çalışma kapsamında verinin boyutlu olup olmadığının test edilmesi amaçlandığı için ve alt testlerde madde sayısının az olması nedeniyle DIMTEST T istatistiği kullanılmıştır. Stout (1987), bir test verisi setinin tek boyutluluğunun parametrik olmayan hipotez testi için DIMTEST programını geliştirmiştir. DIMTEST, yerel madde bağımsızlığı olarak gösterilen testin boyutluluk hipotezini test etmek için yaygın olarak kullanılan bir programdır.

DIMTEST programında analizler yapılırken veri kümesi, AT ve PT olarak iki alt kümeye ayrılır. Boyutsal olarak homojen olması gereken madde setlerinden biri AT olarak seçilir. Kalan madde kümesi ise PT olarak gruplandırılır. Bu maddeler, testin uygulandığı kişilerin puan gruplarına ayrılması için kullanılır. Sonrasında program, testin uygulandığı kişileri PT'deki puanlarına göre kategorilere ayırır ve PT puanlarında bağıl AT maddelerinin eşdeğerlilik çiftlerini hesaplar. AT maddelerinin bağıl eşdeğerliliklerini esas alarak DIMTEST bir T istatistiği üretir ve hipotez testi yapar (Özer Özkan, 2012).

$H_0$ : Test tek boyutludur.

$H_1$ : Test çok boyutludur.

Stout'un T testi AT1 ve AT2 arasında standardize edilmiş varyans farklılıklarını karşılaştırmaktadır ve aşağıdaki formülle hesaplanır.

$$T = \frac{(T_L - T_B)}{\sqrt{2}}$$

Formülde yer alan  $T_L$  ve  $T_B$  AT1 ve AT2 arasında standardize edilmiş varyans farklılıklarıdır. Stout'un T istatistiği ortalaması sıfır standart sapması bir olan normal dağılım



göstermektedir (Meara, Robin ve Sireci, 2000). Eğer DIMTEST T istatistiği sifıra yakınsa  $H_0$  hipotezi korunmuştur. Böylelikle bir veri kümesinin tek boyutlu olduğu sonucuna varılır. Elde edilen p değerinin manidar çıkması ise verinin çok boyutlu olduğuna işaret etmektedir.

Testin 2.000'den fazla kişiye uygulanması durumunda DIMTEST programı, yukarıdaki işlemleri otomatik olarak her bir alt kümede 2.000'i geçmeyecek şekilde olabilecek en büyük boyuttaki eşit alt kümelere ayırır. Bu alt kümelerin her biri için DIMTEST programı PT seçimi ve T hesaplaması ayrı olarak yapılarak her bir alt küme için ayrı bir T hesaplaması elde edilir (Stout, Douglas, Junker ve Roussos,1993).

### Bulgular

Araştırmada elde edilen bulgular ÖBBS 2002, 2005 ve 2008 yıllarının alt testleri bazında verilmiştir. Yıllar bazında çok boyutluluk analizi sonucu elde edilen bulgular aşağıda sunulmuştur.

Tek boyutluluğun testi için mevcut olan birçok test yöntemi içinde teorik olarak doğrulama vermesinden dolayı (Elias, Hattie ve Douglas, 1998) DIMTEST T istatistiği kullanılmıştır. DIMTEST T istatistiğini elde etmek amacıyla Türkçe, matematik, fen ve teknoloji, sosyal bilgiler ve İngilizce testlerine ait veriler, AT ve PT olarak iki alt kümeye ayrılmıştır. Bir grup madde AT olarak seçilmiştir ve kalan maddeler ise PT olarak gruplandırılmıştır. AT maddelerinin bağıl eşdeğerliliklerini esas alarak DIMTEST programı ile bir T istatistiği üretilmiş ve hipotez testi yapılmıştır. Elde edilen p değerinin manidar çıkması verinin çok boyutlu olduğuna işaret etmektedir. Dört alt teste ilişkin hipotez testi sınanması ile elde edilen boyutluluk analizi sonuçları Tablo 4'te verilmiştir.

Tablo 4.

*ÖBBS 2002 Alt Testlere İlişkin Açıklayıcı Çok boyutluluk Analiz Sonuçları*

Yıl	Ders	DIMTEST	
		T	P
2002	Türkçe	3.77	.00
	Matematik	6.02	.00
	Fen ve Teknoloji	5.85	.00
	Sosyal Bilgiler	8.13	.00

Tablo 4 incelendiğinde dört teste ilişkin gözlemlenen istatistik ile ilişkili olasılık seviyesi 0.05'ten daha az olduğu için  $H_0$  hipotezi reddedilmiş ve veri çok boyutlu olarak kabul edilmiştir. Benzer analizler 2005 ÖBBS için yapılmış ve Tablo 5'teki sonuçlara ulaşılmıştır.

Tablo 5.

*ÖBBS 2005 Alt Testlere İlişkin Açıklayıcı DIMTEST Analizi Sonuçları*

Yıl	Ders	DIMTEST	
		T	P
2005	Türkçe	1.42	.01
	Matematik	5.43	.00
	Fen ve Teknoloji	4.80	.00
	Sosyal Bilgiler	8.70	.00

Tablo 5 incelendiğinde alt testlerin 0.05 gözlemlenen istatistik değerleri 0.05'ten daha az olduğu için test çok boyutlu olarak kabul edilmiştir. Benzer analizler 2008 ÖBBS için yapılmış ve Tablo 6'daki sonuçlara ulaşılmıştır.

Tablo 6.

*ÖBBS 2008 Alt Testlere İlişkin Açıklayıcı DIMTEST Analizi Sonuçları*

Yıl	Ders	DIMTEST	
		T	P
2008	Türkçe	3.68	.00
	Matematik	4.04	.00
	Fen ve Teknoloji	3.09	.00
	Sosyal Bilgiler	2.87	.00
	İngilizce	2.62	.00

Tablo 6 incelendiğinde dört teste ilişkin gözlemlenen istatistik ile ilişkili olasılık seviyesi 0.05'ten daha az olduğu için veri çok boyutlu olarak kabul edilmiştir.

### Sonuç ve Tartışma

Bu çalışmanın amacı 2002, 2005 ve 2008'de uygulanan ÖBBS'nin alt testlerinin boyutluluğunun belirlenmesidir. Araştırmanın verileri ÖBBS 2002, 2005 ve 2008 yıllarının Türkçe, matematik, fen ve teknoloji, sosyal bilgiler ve İngilizce testlerine verilen yanıtlardan elde edilmiştir. Bu çalışma kapsamında verinin boyutlu olup olmadığının test edilmesi amaçlandığı ve alt testlerdeki madde sayısı az olduğu için parametrik olmayan çok boyutluluk yöntemlerinden DIMTEST T istatistiği kullanılmıştır.

DIMTEST T istatistiğini elde etmek amacıyla Türkçe, matematik, fen ve teknoloji, sosyal bilgiler ve İngilizce testlerine ait veriler üzerinde her bir alt test için ayrı ayrı analiz gerçekleştirilmiştir. Her bir alt test analiz edilirken AT ve PT olarak iki alt küme ayrıştırılmıştır. Bir grup madde AT olarak seçilmiştir ve kalan maddeler ise PT olarak

gruplandırılmıştır. AT maddelerinin bağıl eşdeğerliliklerini esas alarak DIMTEST programı ile bir T istatistiği üretilmiş ve hipotez testi yapılmıştır.

Çalışmada 2002, 2005 ve 2008’de uygulanan ÖBBS’nin tüm alt testlerinin çok boyutlu olduğu sonucuna ulaşılmıştır. Özellikle öğrenciler hakkında geçti-kaldı kararı verilmek istendiğinde seçme ve yerleştirmelerde ya da bir işe eleman alımında tüm/alt test puanlarının doğruluğu önem kazanmaktadır. ÖBBS’nin madde ve test istatistikleri KTK’ye göre elde edilmekte ve bu doğrultuda geliştirilmektedir. KTK ise tek boyutluluk varsayımı üzerine kurulmuştur. Sonuçlar ÖBBS’nin tüm alt testlerinin çok boyutlu olduğunu göstermektedir. Bu yüzden testlerin veya testlerde bulunan maddelerin sadece tek bir örtük özelliği ölçtüğü varsayımı altında ölçmeler yaparak birey hakkında verilen kararların doğruluğu çok tartışılan bir konudur. Dolayısıyla bu durum testin örtük çok boyutlu yapısının tanımlanmasının ve bu yapıya göre yetenek kestirimi yapmanın önemli olduğunu göstermektedir. Araştırmanın bu sonucu geniş ölçekli testlerde çok boyutluluğun belirlenmesi amacıyla yapılan çalışmaların (Douglas ve diğ., 1999; Seo ve diğ., 2008; Özbek Baştuğ, 2012 ve Özer Özkan, 2012) sonuçları ile örtüşmektedir.

Yapılan çalışmalar (Adams, Wilson ve Wang, 1997; Kelderman,1996; Rost ve Carstensen, 2002; Yao ve Schwarz, 2006; Socha ve DeMars, 2013) birden fazla alt test içeren testlerde tek boyutluluk ihlal edildiğinde güvenilirlik ve geçerlilik sorunları ortaya çıktığını, madde parametrelerinin tahmininde yanlışlık oluşturduğunu ve bu testlerin çok boyutluluk ile modellenmesinin hataları en aza indirmeye yardımcı olacağını ortaya koymuştur.

Yetenek kestirimlerinin daha etkili yapılabilmesi için sınavların Türkçe, matematik, fen ve teknoloji, sosyal bilgiler ve İngilizce dersi öğretim programlarında yer alan öğrenme alanlarına göre hazırlanması önemlidir. Sınavların bu yöntemle hazırlanması alt boyutlar bazında değerlendirmeler yapılarak öğrencilerin eksik olduğu öğrenme alanını ortaya çıkaracağı ve birey hakkında daha tanılayıcı bilgiler vermesine katkı sağlayacağı düşünülmektedir.

İlköğretim (6-7-8. sınıflar) Türkçe Dersi Öğretim Programı incelendiğinde öğrenme alanlarının okuma, dinleme/izleme, konuşma, yazma öğrenme alanları ile dil bilgisinden oluştuğu görülmüştür (Milli Eğitim Bakanlığı-MEB, 2006a). MEB’e göre bu öğrenme alanları hem kendi içlerinde hem de birbirleriyle bir bütünlük içerisinde ele alınmış ve ilişkilendirilmiştir. Ancak ÖBBS çoktan seçmeli testlere dayalı bir sınav olduğu için yapılan sınavda okuduğunu anlama becerisi ve dil bilgisine yönelik sorular hazırlanmıştır (MEB, 2002; 2007; 2009). Benzer şekilde İlköğretim Matematik Öğretim Programında sekizinci

sınıflar için öğrenme alanları, sayılar, geometri, ölçme ve cebir olarak belirtilmiştir (MEB, 2006b).

Alt öğrenme alanları düzeyinde değerlendirme yapabilmek için öncelikle soru sayısının artırılması gerekmektedir. Araştırmada kullanılan testlerde soru sayılarının az olması ve birbirinden farklı alt alanları kapsamaması puanlama açısından sorunlara neden olmaktadır. Örneğin Türkçe testinde soru sayısının 25 olması, dilbilgisi ve okuduğunu anlama becerisi gibi alanları içermesi puanlama açısından sorunlar oluşturmaktadır. Bu durum diğer tüm dersler için de benzer sorunların temelini oluşturmaktadır. Burada Türkçe ve matematik için verilen örnekler diğer dersler için de genişletilebilir. Diğer bir anlatımla tüm dersler düzeyinde alt boyutların sınav öncesi belirlenmesi ve bu boyutları test edebilecek miktarda soru ile değerlendirilmesi gerekmektedir.

### Kaynaklar

- Ackerman, T.A. (1989). Unidimensional IRT calibration of compensatory and non-compensatory multidimensional items. *Applied Psychological Measurement*, 13, 113-127.
- Ackerman T.A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement In Education*, 7(4), 255–278.
- Adams, R. J., Wilson, M., & Wang, W.C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21, 1–23.
- Ansley, T.N. & Forsyth, R.A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement*, 9, 37-48.
- Balázs, K., Hidegkuti, I., & De Boeck, P.(2006). Heterogeneity in logistic regression models. *Applied Psychological Measurement*. Volume 30 Number 4 July.
- Büyüköztürk, Ş., Çokluk, Ö. ve Köklü, N. (2010). *Sosyal bilimler için istatistik* (6. baskı). Ankara: Pegem A yayınevi Tic. Ltd. Şti.
- Camilli, G., Wang, M., & Fesq, J. (1995). The effects of dimensionality on equating the law school admission test. *Journal of Educational Measurement*, 32(1), 79–96.
- Douglas, J., Kim, H. R., Roussos, L., Stout, W., & Zhang, J. (1999). *LSAT Dimensionality analysis for the december 1991, june 1992, and october 1992 administrations* [Law School Admission Council Statistical Report 95-05].
- Drasgow, F. & Parsons, C.K. (1983). Application of Unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement*, 7, 189-199.
- Elias, S., Hattie, J., & Douglas, G. (1998). An assessment of various item response model and structural equation model fit indices to detect unidimensionality. *Paper presented at the annual meeting of the National Council on Measurement in Education*, San Diego, CA.

- Froelich, A.G. (2001). *A new bias correction method for DIMTEST*. Unpublished manuscript.
- Froelich, A. G. & Habing B. (2003). *Conditional covariance based subtest selection for DIMTEST. in william stout institute for measurement, dimtest version 2.0[software manual]*. Urbana-Champaign, IL: William Stout Institute for Measurement.
- Gao, F. (1997). *DIMTEST Enhancements and some parametric IRT asymptotics*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Hambleton, R. K. & Swaminathan H. (1985). *Item response theory: principles and application*. Kluwer, Nijhoff Publishing a Member of Kluwer Academic Publisher Group.
- Hambleton, R. K., Swaminathan H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. California: Sage Publications Inc.
- Harrison, D.A. (1986). Robustness of IRT parameter estimation to violations of the unidimensionality assumption, *Journal of Educational Statistics*, 11, 91-115.
- Hattie, J., Krakowski, K., Rogers, J., & Swaminathan, H. (1996). An assessment of stout's index of essential dimensionality. *Applied Psychological Measurement*, 20, 1-14.
- Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory*. Dow Jones-Irwin, Homewood, IL.
- Junker, B. W. & Sijtsma, K. (2001). Nonparametric item response theory in action: An overview of the special issue. *Applied Psychological Measurement*, 25, 211-220.
- Kelderman, H. (1996). Multidimensional rasch models for partial-credit scoring. *Applied Psychological Measurement*, 20, 155–168.
- Kim, H. R. (1994). *New techniques for dimensionality assessment of standardized test data*. Unpublished doctoral dissertation. University of Illinois at Urbana-Champaign, Department of Statistics.
- McDonald, R. P. (1981). The dimensionality of test and items. *British Journal of Mathematical and Statistical Psychology*, 34, 100-117.
- Meara, K., Robin, F., & Sireci, S.G. (2000). Using Multidimensional Scaling to Assess the Dimensionality of Dichotomous Item Data, *Multivariate Behavioral Research*, 35 (2), 229–259.
- Milli Eğitim Bakanlığı, Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı (2002). *Öğrenci Başarılarının Belirlenmesi Sınavı Durum Belirleme Raporu: Türkçe, Matematik, Fen Bilgisi, Sosyal Bilgiler*. Milli Eğitim Bakanlığı.
- Milli Eğitim Bakanlığı, Talim ve Terbiye Kurulu Başkanlığı (2006a). *İlköğretim Türkçe dersi (6, 7, 8. Sınıflar) öğretim programı*. Milli Eğitim Bakanlığı.
- Milli Eğitim Bakanlığı, Talim ve Terbiye Kurulu Başkanlığı (2006b). *İlköğretim matematik dersi (6, 7, 8. Sınıflar) öğretim programı*. Milli Eğitim Bakanlığı.
- Milli Eğitim Bakanlığı, Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı (2007). *Öğrenci başarılarının belirlenmesi sınavı durum belirleme raporu: Türkçe, matematik, fen*

*bilgisi, sosyal bilgiler, İngilizce*. Eğitimi Araştırma ve Geliştirme Dairesi Yayınları.  
Milli Eğitim Bakanlığı

- Milli Eğitim Bakanlığı, Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı (2009). *Öğrenci başarılarının belirlenmesi sınavı durum belirleme raporu: Türkçe, matematik, fen bilgisi, sosyal bilgiler, İngilizce*. Eğitimi Araştırma ve Geliştirme Dairesi Yayınları. Milli Eğitim Bakanlığı.
- Mroch, A. A. & Bolt, D. M. (2006). A simulation comparison of parametric and nonparametric dimensionality detection procedures. *Applied Measurement in Education*, 19 (1), 67-91.
- Nandakumar, R. & Stout, W.F. (1993). Refinements of Stout's procedure for assessing unidimensionality. *Journal of Educational Statistics*, 18, 41-68.
- Özer Özkan, Y. (2012). *Klasik test kuramı, tek boyutlu ve çok boyutlu madde tepki kuramı modellerinden kestirilen öğrenci başarısı belirleme sınavı (ÖBBS) başarı ölçülerinin karşılaştırılması*. Yayımlanmamış Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Özbek Baştuğ, Ö.Y. (2012). Assessment of Dimensionality in Social Science Subtest. *Educational Sciences: Theory & Practice*. 12(1), Winter: 382-385.
- Rost, J. & Carstensen, C. H. (2002). Multidimensional rasch measurement via item component models and faceted designs. *Applied Psychological Measurement*, 26, 42–56.
- Roussos, L.A., Stout, W., & Marden, J. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of educational measurement*, 35, 1-30.
- Seo, M., Rutgers, C., & Roussos, L. (2008). Evaluating the dimensionality of the 2001 PIRLS reading assessment: An application of DIMTEST with DESM and CFA. [http://www.iea.nl/fileadmin/user\\_upload/IRC/IRC\\_2008/Papers/IRC2008\\_Seo\\_ChIU\\_et\\_al.pdf](http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2008/Papers/IRC2008_Seo_ChIU_et_al.pdf)
- Sinharay, S., Haberman, S. J., & Puhan, G. (2007). Subscores Based on Classical Test Theory: to Report or Not to Report. *Educational Measurement: Issues and Practice*, 26 (4), 21–28.
- Smith, J. (2009). *Some issues in item response theory: Dimensionality assessment and models for Guessing*. Unpublished Doctoral Dissertation. University of South California.
- Smith, J. & Roussos, L. (2010). *Effect of refined subtest selection on DIMTEST performance with an application to nonsimple structure multidimensionality*. Presented at the 2010 annual meeting of the American Education Research Association May 2.
- Socha, A. & DeMars, C.E. (2013). An investigation of sample size splitting on ATFIND and DIMTEST. *Educational and Psychological Measurement*, 73, 631-647.
- Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, 52, 589-617

- Stout, W., Froelich, A. G., & Gao, F. (2001). Using resampling methods to produce an improved DIMTEST procedure. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essay on item response theory* (pp. 357-375). New York: Springer.
- Stout, W. F., Habing, B., Douglas, J., Kim, H. R., Roussos, L., & Zhang, J. (1996). Conditional covariance based nonparametric multidimensionality assessment. *Applied Psychological Measurement*, 19, 331-354.
- Stout, W.F., Douglas, J., Junker, B., & Roussos, L.A. (1993). *DIMTEST Manual*. Unpublished manuscript available from W.F. Stout, University of Urbana-Champaign, Illinois.
- Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement*, 27, 159–203.
- Traub, R.E. (1983). *A priori consideration in choosing an item response model*. In R.K.
- Van Abswoude, A. A. H., Van der Ark, L.A., & Sijtsma, K. (2004) A comparative study of test data dimensionality assessment procedures under nonparametric IRT models. *Applied Psychological Measurement*. January, vol. 28, no. 1 3-24.
- Yao, L. & Schwarz R. (2006). A Multidimensional partial credit model with associated item and test statistics: an application to mixed format tests. *Applied Psychological Measurement*, 30, 469–492.
- Zhang, J. & Stout, W. (1999a). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika*, 64, 213-249.
- Zhang, J. & Stout, W.F. (1999b). Conditional covariance structure of generalized compensatory multidimensional items. *Psychometrika*, 64, 129–152.
- Way, W. D., Ansley, T.N., & Forsyth, R. A. (1988). The comparative effects of compensatory and non-compensatory two dimensional data on unidimensional IRT estimates. *Applied Psychological Measurement*, 12, 239-252.

### **Extended Abstract**

#### **Purpose**

The methods that empirically evaluate test structure need to primarily provide information about test dimensionality and other structural characteristics. Test dimensionality has a direct impact on a test's structure validity, score calculation, and result recording. There is a continuous debate on whether to use the score that belongs to a whole test or not in order to reveal the student performance, if the test is a multidimensional one. The purpose of this study is to identify the dimensions of four subtests that constitute the parts of Evaluation of Student Achievement Test (ÖBBS) by using DIMTEST T statistics which is a nonparametric multidimensionality method.

Initially, the descriptive statistics values of ÖBBS 2002, 2005, and 2008 were considered. Of the four subtests in ÖBBS 2002 while Turkish test scores were the highest,

science and technology test scores were the lowest. In all four subtests the skewness and kurtosis coefficient values were within the mentioned limits. Therefore, the distribution formed by the data can be accepted as very close to symmetrical. Similar to 2002 values the scores in ÖBBS 2005 show that Turkish test scores were the highest and the science and technology test scores were the lowest. When the skewness values of all tests and the sub-dimensions are analyzed it can be argued that the distribution formed by the data does not display an extreme deviation from the normal since the results obtained are between the values of -1 and +1. The analysis of ÖBBS 2008 data show that Turkish test score is the highest of all. However, contrary to 2002 and 2005 test scores the lowest one in ÖBBS 2008 is the English test score.

This study is a descriptive one. The data in this study is obtained from the responses provided for the Turkish, Mathematics, Science and Technology, and Social Sciences tests in ÖBBS 2002. In order to test the multidimensionality of the data DIMTEST T statistics was used.

#### Result

In order to find the DIMTEST T statistics of the four subtests in ÖBBS 2002, 2005, and 2008, the data of Turkish, Mathematics, Science and Technology, and Social sciences test data was divided in to two subsets as AT and PT. A group of items was selected as AT and the rest was grouped as PT.

T statistics was created by the DIMTEST program and the conditional equivalence of AT items was considered. As a result of the hypothesis test it was found that all the subtests in ÖBBS 2002, 2005, and 2008 were multidimensional.

#### Discussion

The study results show that all the subsets of ÖBBS are multidimensional. When the skills that are necessary to solve the questions in cognitive tests are considered, it can be noticed that unidimensionality was not established. If the items vary in various skill levels and if the respondents vary in more than one of these skills, this interaction has to be modeled as multidimensional. Moreover, the dimensions of a test can be composed of various sub-dimensions. Thus, the evaluation of dimensionality is an important term for educational and psychological measurement. In situations when unidimensional assumption is violated, serious deviations that belong to the evaluation of skills of the test taker can be found. In order to find reliable results, the unidimensionality assumption should be confirmed before the application of psychometric methods that require unidimensionality of the data.



### Conclusion

In order to implement skill predictions more efficiently, tests should be designed according to the learning fields mentioned in Turkish, mathematics, science and technology, and English course programs. Thus, the learning fields that are problematic for students can be identified and more diagnostic individual information can be collected.