



Volume: 11 Issue: 1 Year: 2014

A comparison of estimated achievement scores obtained from student achievement assessment test utilizing classical test theory, unidimensional and multidimensional IRT¹

Öğrenci başarılarının belirlenmesi sınavından klasik test kuramı, tek ve çok boyutlu madde tepki kuramı modelleri ile kestirilen başarı puanlarının karşılaştırılması

Yeşim Özer Özkan²

Abstract

The focus of this research is to test the estimation of achievement measurements in the test battery and to empirically compare the results after applying classical test theory, unidimensional and multidimensional item response theory models to Student Achievement Assessment Test (ÖBBS-2008) subtests of Turkish and Mathematics. It also tries to put forward the best model that estimates students' achievement with less error as the comparison is being made. From the analysis of Turkish test's data results, it is identified that the ability parameters estimated obtained from the whole test under multidimensional IRT, have partially less error scores and reached more precise measurement than ability parameters estimated obtained from unidimensional IRT on the basis of sub dimensions and test scores obtained from CTT. Similar results were obtained in mathematics test results. Finally, it is found that parameters, obtained within the scope of multidimensional IRT, have partially less error scores.

Özet

Bu araştırmada, bir test bataryasındaki başarı ölçüleri kestiriminin doğruluğunun belirlenmesi ve ampirik olarak Klasik Test Kuramı (KTK), tek ve çok boyutlu Madde Tepki Kuramı (MTK) modellerinin Öğrenci Başarılarının Belirlenmesi Sınavı'nın (ÖBBS-2008) Türkçe ve matematik alt testi verilerine uygulanarak elde edilen başarı ölçülerinin karşılaştırılması amaçlanmıştır. Bu karşılaştırmalar yapılırken başarı ölçülerini daha az hata ile kestiren en iyi model ortaya konulmaya çalışılmıştır. Türkçe testi verilerinin analizi sonucunda tüm testten çok boyutlu MTK ile kestirilen yetenek parametrelerinin alt boyutlar bazında tek boyutlu MTK'ye göre kestirilen yetenek parametreleri ve KTK'ye göre elde edilen test puanlarına kıyasla kısmen daha düşük standart hataya sahip olduğu belirlenmiştir. Matematik testi verilerinin analizi sonucunda, yetenek parametrelerinin kestiriminde en düşük hatanın çok boyutlu MTK'ye göre; en yüksek hatanın ise matematik testinin alt boyutlarından tek boyutlu MTK ve tüm testten KTK'ye göre belirlenen puanlardan elde edildiği belirlenmiştir.

¹ Bu çalışma, 2012 yılında Ankara Üniversitesi Eğitim Bilimleri Enstitüsü'nde Prof. Dr. Ezel Tavşancıl danışmanlığında yürütülmüş olan doktora tez çalışmasından hazırlanmıştır.

²Yrd. Doç. Dr., Gaziantep Üniversitesi, Gaziantep Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı. Gaziantep. yozer80@gmail.com

Keywords: Unidimensional and multidimensional item response theory, classical test theory, ability estimation, dimensionality, ÖBBS.

Anahtar Kelimeler: Tek ve çok boyutlu madde tepki kuramı, klasik test kuramı, yetenek kestirimi, boyutluluk, ÖBBS.

[\(Extended English abstract is at the end of this document\)](#)

Giriş

Bilgi çağı olarak adlandırdığımız 21. yüzyılda eğitime olan talep ve beklentiler giderek artmaktadır. Toplumun ve bireylerin daha yüksek eğitim talebi, eğitim isteklerinin sunumunda sıkıntılara neden olmaktadır. Bu nedenle talebin yüksek olduğu alanlarda öğrencilerin başarılarına göre sıralanarak seçilmesi zorunlu hale gelmiştir. Bununla birlikte insana yapılan yatırımın en değerli yatırım olduğu anlayışı eğitim hedeflerinin gerçekleşip gerçekleşmediğini kontrol etmeyi gerektirmektedir. Bahsedilen nedenlerle öğrencilerin eğitim ihtiyaçlarının belirlenmesinde, hedeflere ulaşma düzeyinin saptanmasında ve eğitim politikalarına yön vermede testlerin önemi giderek artmaktadır. Eğitim ve psikolojide yaygın olarak kullanılan testler psikolojik yapıların anlaşılması, teşhis edilmesi ve diğer yapılarla ilişkilerinin ortaya çıkarılmasını sağlarken öğrenciler hakkında eğitim kararlarının verilmesi, öğrenme zorluklarının teşhisi, öğrenci başarısının saptanması, seçme ve yerleştirme gibi çeşitli amaçlar için kullanılmaktadır (Baykul, 2000). Bu testlerden elde edilen puanlar bireyler hakkında verilen kararların doğruluğunda temel bir dayanak oluşturmaktadır.

Türkiye’de, eğitimin öğrenci başarısı üzerindeki etkisini görmek amacı ile ulusal ve uluslararası alanda ve değişik sınıf düzeylerinde standart testler ve anketler kullanılarak ölçme değerlendirme çalışmaları yapılmaktadır. Çalışmalarını uluslararası düzeyde yürüten kuruluşlardan biri Uluslararası Eğitim Başarısını Değerlendirme Kuruluşu (International Association for The Evaluation of Educational Achievement-IEA-)'dur. Diğeri ise kurucu üyesi olduğumuz İktisadi İşbirliği ve Kalkınma Teşkilatı (Organisation for Economic Co-operation and Development-OECD-)'dir. IEA'nın yaptığı çalışmalar; Uluslararası Okuma Becerilerinde Gelişim Projesi (PIRLS) ve Üçüncü Uluslararası Matematik ve Fen Bilgisi Araştırma Projesi (TIMSS)'dir. OECD'nin yürüttüğü çalışma ise Uluslararası Öğrenci Değerlendirme Projesi (Program for International Student Assessment-PISA-)'dir.

Uluslararası izleme ve değerlendirme çalışmaları yanında ulusal düzeyde yapılan Öğrenci Başarılarının Belirlenmesi Sınavı (ÖBBS), Seviye Belirleme Sınavı (SBS), Yükseköğretime Geçiş Sınavı (YGS) ve Lisans Yerleştirme Sınavı (LYS) gibi durum belirleme ve seçme amaçlı yapılan sınav sonuçları da okul öğrenmelerinin izlenmesine ışık tutabilecek sonuçlar vermektedir. Ulusal düzeyde üç yıllık periyotlarla ilköğretim ve ortaöğretim öğrencilerine uygulanan ÖBBS bir durum belirleme çalışmasıdır. Bu sınav ilköğretim düzeyinde dördüncü, beşinci, altıncı, yedinci ve sekizinci

sınıflara, Türkçe, Matematik, Fen ve Teknoloji, Sosyal Bilgiler ve İngilizce temel derslerinde yapılmaktadır. Genel amacı temel eğitimde öğrencilerin ulusal boyutta durumunu belirlemek ve eğitimin kalitesini ortaya koymaktır (MEB, 2009).

ÖBBS’de öğrenci düzey belirleme testleri ile birlikte öğrenci ve öğretmen anketleri kullanılmaktadır. Öğretmen anketlerinde öğretmenlerden yaş, cinsiyet, öğretmenlikte kıdem, esas branş, okutulan ders, son olarak bitirilen okul, evinde bulunan ders kitapları dışındaki kitap sayısı ile ilgili kişisel bilgiler ve öğretmenlik mesleğine bağlılık, mesleğinin toplum ve öğrenciler tarafından takdir edilip edilmediği ile ilgili görüşler alınmıştır. Öğrenci anketinde ise kişisel bilgiler, öğrencilerin eğitime ayırdıkları zaman, derslerde kendilerini başarılı bulma dereceleri gibi konularda sorular yöneltilmiştir (MEB, 2009). Bu çalışmada veri toplama aracı olarak kullanılan öğrenci düzey belirleme testleriyle ise öğrencilerin başarı durumları değerlendirilmeye çalışılmaktadır. Öğrenci düzey belirleme sınavı öğrencinin başarı düzeyini gösteren bir test sonucu vermektedir.

Türkiye’de testlerin ortaöğretim ve yükseköğretim kurumlarına girişte öğrencileri seçme ve yerleştirme amacıyla kullanılması ve buna dayalı olarak bireylerin gelecekte iş ve meslek yaşamının belirlenmesi, test sonuçlarına göre verilecek kararların önemini daha da arttırmıştır.

Öğrencilerin başarı durumlarının bir göstergesi olan test sonuçları tüm test puanı veya alt test puanları olmak üzere iki şekilde hesaplanmaktadır. Türkiye’de öğrenciler hakkında önemli kararların alındığı ulusal düzeyde uygulanan geniş ölçekli test sonuçları tüm test üzerinden hesaplanmaktadır. Tüm test puanı, bireylerin ölçmeye konu olan özelliklerini, kendilerinde bulunma derecesine göre verdikleri yanıtlar doğrultusunda doğru (1 puan) ya da yanlış (0 puan) kabul edilme ölçütüne göre puanlanarak hesaplanır. Tüm test puanı öğrenci hakkında genel bir değerlendirme sonucu elde edilmesini sağlasa da ölçülen özellik tek boyutlu değil ise öğrenci hakkında kapsamlı bir değerlendirme sonucu elde etmemize engel teşkil eder. Bununla birlikte, öğrencilerin gelecekte yapacakları çalışmaları planlamak için farklı içerik bölümlerindeki (farklı alt testlerdeki) güçlü ve zayıf yönlerinin belirlenmesi gerekir.

Haladyana ve Kramer’in de (2004) belirttiği üzere “Eyaletler, kolej veya üniversiteler gibi akademik kurumlar öğrencilerin performansı hakkında bilgi sahibi olmak ve ihtiyaç olan alanlara odaklanmak ister”. Bu nedenlerle tüm testten elde edilen toplam puanı rapor etmek testin çok boyutlu yapıya sahip olması veya boyutlar arasındaki ilişkinin düşük olması gibi farklı koşullar altında uygun olmayabilir. Belirtilen bu gerekçelerden dolayı toplam test puanı yerine alt test puanlarının rapor edilmesi önem kazanmaya başlamıştır.

Eğitimde ve psikolojide kullanılan testler farklı alt bölümlerden oluşmaktadır. Örneğin, matematik bilgisini ölçen bir test aritmetik ve geometri alt testlerinden oluşabilir. Benzer şekilde

genel yeteneği ölçen bir testte sözel, sayı ilişkileri ve problem çözme gücü, şekiller arası ilişkiler gibi alt testlerden oluşabilir. Alt testlerin her biri farklı bir öğrenme alanıdır ve her bir alt öğrenme alanına ilişkin test puanları ayrıca hesaplanabilir. Bu alt testlerden hesaplanan puanlar ise alt test puanı olarak tanımlanmaktadır (Sinharay, Haberman ve Puhan, 2007).

Öğrenci hakkında her bir alt öğrenme alanındaki eksiklikleri ortaya koyarak kapsamlı tanılayıcı bilgiler vermesi, alt test puanlarına olan ilgiyi arttırmıştır (Bock, Thissen ve Zimowski, 1997; Haberman, 2008; Haberman ve Sinharay, 2010a; Tate, 2004; Yao, 2009). Özellikle öğrenciler hakkında geçti-kaldı kararı verilmek istendiğinde, seçme ve yerleştirmelerde ya da bir işe eleman alımında daha tanılayıcı bilgi veren tüm/alt test puanlarının doğruluğu önem kazanmaktadır.

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 Kuramı'dır (KTK). KTK test geliştirmeye uzun yıllardır hizmet veriyor olmakla birlikte MTK, ölçme için kuramsal bir temel olarak çok hızlı bir şekilde ana kaynak olmaya başlamıştır. Kuramsal olarak daha savunulabilir ölçme ilkelerine sahip olduğu ve pratik ölçme problemlerini daha yüksek potansiyelle çözümlendiği için standardize testler giderek artan bir şekilde MTK ile geliştirilmektedir (Emberson ve Reise, 2000). MTK modelleri ölçülen özelliğin yapısının tek veya çok boyutlu olmasına göre tek boyutlu ve çok boyutlu MTK olarak isimlendirilir. KTK ve tek boyutlu MTK modelleri tek boyutluluk varsayımı üzerine kurulmuştur. Bu varsayımın göre bütün maddeler aynı beceriyi veya beceriler bütünü ölçmektedir ve testlerde tek boyutluluk sağlanmalıdır (Emberson ve Reise, 2000). Fakat bu varsayım çoğu zaman karşılanamamaktadır.

Eğitimde yapılan birçok ölçme durumunda testi oluşturan maddeler kendi başına farklı beceri veya beceriler bütünü ölçebilmektedir. Örneğin, cebirle ilgili bir sınavda eşitliğe dayalı kısa sorulu problemler ve günlük yaşam problemleri olarak iki tip problem sorulmuş olsun: İlk 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.

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 tartışılır hale gelmiştir (Ackerman, 1989; Ansley ve Forsyth, 1985; Drasgow ve Parsons, 1983; Harrison, 1986; Way, Ansley ve Forsyth, 1986). Bu araştırmalar tek boyutluluk varsayımının karşılanmadığı durumlarda yetenek ve madde parametre tahminlerinin etkilendiğini ortaya koymuştur. Ayrıca yapılan araştırmalarda (Adams ve diğerleri, 1997; Kelderman, 1996; Rost ve Carstensen, 2002; Yao ve Schwarz, 2006) çok boyutlu MTK'nin, birden fazla alt test içeren testlerde

tek boyutlu MTK uygulandığında ortaya çıkan geçerlik ve güvenirlik sorunlarını en aza indirmeye yardımcı olduğu ortaya konmuştur.

Alan yazın incelendiğinde birçok araştırmada (Anderson, 1999; Courville, 2005; Çelen, 2008; Demirtaşlı, 2002; Fan, 1998; Lawson, 1991; MacDonald ve Paunonen, 2002; Ndalichako ve Rogers, 1997; Progar ve Sočan, 2008; Rogers ve Ndalichako, 2000; Tomkowicz ve Rogers, 2005) ikili puanlanan maddelerden oluşan testlerde test geliştirmede ve değerlendirmede KTK ile tek boyutlu MTK karşılaştırılmıştır ve her iki kurama göre benzer sonuçların elde edildiği ortaya konmuştur.

Tek boyutlu MTK ile KTK karşılaştırması yapılan araştırmaların yanı sıra yetenek kestiriminde tek boyutlu ve çok boyutlu MTK'nın karşılaştırıldığı çalışmalar da (Chang, 1992; de la Tore ve Patz, 2005; Köse, 2010; Luecht, 2003; Spencer, 2004; Sümbül, 2011) bulunmaktadır. 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 yeterli seviyesinde farklılaşırsa bu etkileşim çok boyutluluk ile modellenmelidir (Ackerman, 1994).

Araştırma sonuçları incelendiğinde, çok boyutlu MTK'nın tek boyutluluk varsayımının getirdiği sınırlılıkları giderip gidermediğine karar verebilecek düzeyde sonuçlara ulaşılamamış olması, yapılan çalışmaların çoğunlukla simülasyon veri üzerinde veya araştırmacı tarafından geliştirilen testlerin bir grup öğrenci üzerine uygulanması ile yürütülmesi, geniş ölçekte uygulanan gerçek bir veriden elde edilen test sonuçları üzerine yapılan karşılaştırma çalışmalarının oldukça az olması gibi nedenlerle daha az hatalı yetenek kestirimini elde edebilecek bir modeli karşılaştırmalı olarak ortaya koymak bir gereklilik olarak görülmüştür.

Çalışmanın Amacı

Bu araştırmanın amacı bir test bataryasındaki başarı ölçüleri kestiriminin doğruluğunu belirlemek, ampirik olarak KTK, tek ve çok boyutlu MTK modellerini ÖBBS verilerine uygulayarak karşılaştırmaktır. Bu amaçla aşağıdaki sorulara cevap aranmıştır:

1. Türkçe testinden ve testin alt boyutlarından (“anlam bilgisi” ve “dil bilgisi”) KTK'ya göre elde edilen puanların hata kestirimleri ile tek ve çok boyutlu MTK'ya göre belirlenen yetenek parametrelerinin hata kestirimleri arasında manidar fark var mıdır?
2. Matematik testinden ve testin alt boyutlarından (“aritmetik” ve “sayısal mantık”) KTK'ya göre elde edilen puanların hata kestirimleri ile tek ve çok boyutlu MTK'ya göre belirlenen yetenek parametrelerinin hata kestirimleri arasında manidar fark var mıdır?

3. Türkçe testinden KTK'ya göre elde edilen puanlar ile tek ve çok boyutlu MTK'ya göre belirlenen yetenek parametrelerinin kestirimleri arasında manidar bir ilişki var mıdır?
4. Matematik testinden KTK'ya göre elde edilen puanlar ile tek ve çok boyutlu MTK'ya göre belirlenen yetenek parametrelerinin kestirimleri arasında manidar bir ilişki var mıdır?
5. Türkçe ve matematik testlerinden KTK'ya göre elde edilen puanlar ile tek ve çok boyutlu MTK'ya göre belirlenen yetenek kestirimlerinin güvenilirlikleri ne düzeydedir?

Çalışmanın Önemi

Eğitim alanında uygulanan testlerin amacı, bireyin verdiği tepkilerden yararlanarak geçerli ve güvenilir sonuçlara ulaşmaktır. Türkiye'de geniş ölçekli test sonuçlarına göre bireyler hakkında eğitimsel ve iş yaşamına ilişkin önemli kararlar verilmektedir. Geniş ölçekli testleri oluşturan maddeler kendi başına farklı beceri ya da beceri bütünü ölçebilmektedir. Türkiye'de bu testlerin geliştirilmesi ve değerlendirilmesinde KTK modellerinden yararlanılmaktadır. Test geliştirme çalışmalarında ve psikolojik ölçeklerin puanlanmasında KTK yaygın olarak kullanılan bir kuram olmakla birlikte KTK'da yaşanan sınırlılıklara alternatif olarak geliştirilen MTK da giderek daha fazla uygulama alanı bulmuştur. KTK ve tek boyutlu MTK tek boyutluluk varsayımı üzerine kurulduğundan dolayı bu kuramlara dayalı analizler tartışma konusu olmuştur. Eğitimde ve psikolojide kullanılan testlerin çok boyutlu özelliğe sahip olması testlerin geliştirilmesinde ve değerlendirilmesinde çok boyutlu MTK'nın kullanımını bir gereklilik haline getirmiştir.

Alanyazındaki araştırmalar incelendiğinde çok boyutlu MTK'nın tek boyutluluk varsayımının getirdiği sınırlılıkları giderip gidermediğine karar verebilecek düzeyde sonuçlara ulaşamamıştır. Bu yönüyle araştırmanın çok boyutlu testlerin geliştirilmesine ve yorumlanmasına katkı sağlayacağı düşünülmektedir.

Öğrencilerin başarı durumlarını belirlemede kullanılan testlerin tek boyutlu olarak kabul edilmesi ve test sonuçlarının toplam test üzerinden hesaplanması öğrenciler hakkında genel bir değerlendirme sonucu elde etmemizi sağlarken öğrenci hakkında kapsamlı bir bilgi elde etmemize engel teşkil etmektedir. Çok boyutlu MTK, öğrenciler hakkında tanılayıcı bilgiler sunmakta ve bu durumun bir sonucu olarak öğrencilerin farklı alt testlerdeki güçlü ve zayıf yönlerinin belirlenerek gelecekte yapacakları çalışmalarını planlamalarına ve daha isabetli kararlar vermelerine katkı sağlamaktadır. Bu yönüyle araştırmanın özellikle geniş ölçekli testlerin değerlendirilmesinde çok boyutlu MTK'nın kullanımının bir avantajı olan daha kapsamlı bilgiler elde etmemizi sağlaması yönünde uygulayıcılara farkındalık oluşturması açısından önemli görülmektedir.

Yetenek kestiriminin en az hata ile yapılması uygun modelin kullanılması ile sağlanabilir. Özellikle öğrenciler hakkında önemli kararların verildiği geniş ölçekli testlerden elde edilen sonuçların doğruluğu bireysel ve toplumsal açıdan önem arz etmektedir. Bu yönüyle çalışmanın, elde edilen ölçme sonuçlarının doğruluğuna ve dolayısıyla eğitim sistemine katkı sağlayacağı düşünülmektedir.

YÖNTEM

Araştırma Modeli

Araştırma ölçme modellerinin, bir testin tümünden ya da alt testlerinden elde edilen puanların psikometrik özellikleri üzerindeki etkisine odaklanmıştır. Bu araştırma ile ÖBBS'nin Türkçe ve matematik testinden elde edilen verilerin KTK, tek ve çok boyutlu MTK modellerine göre analizinde yetenek parametrelerinin kestiriminde ortaya çıkan hata miktarının ve güvenilirliğinin kuramlara göre farklılaşıp farklılaşmadığı ele alınmaktadır. Araştırma var olan kuramların gerçek veri üzerine sınanması yapıldığından bir verinin psikometrik özelliklerinin belirlenmesi durum saptamaya yönelik olduğundan araştırma betimsel türde temel bir çalışmadır.

Evren ve Örneklem

Araştırmada MEB EARGED tarafından ilköğretim sekizinci sınıf öğrencilerine uygulanan ÖBSS verilerinden yararlanıldığı için bu bölümde, ÖBBS'nin evren ve örneklemeden bahsedilmiştir. ÖBBS'nin evrenini 2008 yılında Türkiye'deki resmî ve özel ilköğretim okullarının sekizinci sınıfına devam eden öğrenciler oluşturmaktadır. Evreni temsil eden örnekleme ise MEB Eğitim Araştırma ve Geliştirme Dairesi (EARGED) tarafından MEB e-okul veritabanı 2008 verileriyle Türkiye İstatistik Kurumu tarafından belirlenen Türkiye'nin Ekonomik ve Sosyal Kalkınmışlık Düzeyi Düzey-2 verileri kullanılarak 36 ilden, 270 ilköğretim okulundan tabakalı örnekleme yoluyla belirlenmiş 9876 sekizinci sınıf öğrencisi oluşturmaktadır. Düzey 2'de belirtilen 26 ekonomik ve sosyal kalkınmışlık bölgesinin her birinden bölgedeki il sayısına göre 36 il belirlenmiştir.

Örnekleme seçilirken yedi coğrafi bölgeyi temsil edecek iller, illerin gelişmişlik düzeyi, o bölgeyi temsil edebilme özellikleri ve ilgili sınıf seviyelerindeki öğrenci sayıları göz önüne alınmıştır. Okullar çalışmaya dahil edilen her ildeki ilköğretim okullarından eş olasılıkla seçilmiştir. İllerde hangi okulların örnekleme dahil edileceğine ek olarak bu okulların dördüncü, beşinci, altıncı, yedinci ve sekizinci sınıflarındaki hangi şubelerin örnekleme dahil edileceği eş olasılık yöntemiyle belirlenmiştir (MEB, 2009).

Veri Toplama Aracı ve Verilerin Elde Edilmesi

Bu çalışmada, MEB tarafından 2008 yılında ilköğretim sekizinci sınıf öğrencilerine uygulanan ÖBBS verilerinden yararlanılmıştır. Araştırma kapsamında kullanılan veriler, bu sınavın 25'er çoktan seçmeli maddeden oluşan Türkçe ve matematik dersleri için hazırlanan düzey belirleme testlerinden elde edilmiştir. Bu alt testlere öğrencilerin verdikleri yanıtlar, MEB Eğitim Teknolojileri Genel Müdürlüğü Ölçme ve Değerlendirme Dairesi Başkanlığından alınmıştır.

Verilerin Analizi

Madde ve test parametrelerinin kestirilmesinde kullanılan kuramlardan biri MTK'dır. MTK'nın önemli varsayımlarından biri, bütün maddelerin aynı beceriyi veya beceriler bütününe ölçtüğüdür. Fakat pek çok ölçme durumunda testi oluşturan maddeler kendi başına farklı beceri veya beceriler bütününe ölçebilmektedir. Bu nedenle ilk olarak testin tek boyutlu mu, yoksa çok boyutlu mu olduğunun değerlendirilmesi gerekmektedir. Stout (1987), bir test verisi setinin boyutluluğunun belirlenmesinde parametrik olmayan hipotez testi için lineer faktör analizi yöntemini geliştirmiştir. Tek boyutluluğun testi için mevcut olan birçok test yöntemi içinde kuramsal olarak doğrulama vermesinden lineer faktör analizine dayalı DIMTEST T istatistiği, tek boyutluluğun değerlendirilmesi için başarılı göstergelerden biri olarak bilinmektedir (Elias, Hattie ve Douglas, 1998).

DIMTEST T istatistiğini elde etmek amacıyla Türkçe testine ait veriler, Değerlendirme Alt Testi (Assessment Subtest-AT) ve Bölümlenme Alt Testi (Partitioning Subtest-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. Hipotez testinin sınanması ile elde edilen boyutluluk analizi sonuçları incelendiğinde ($p=0.0001$, $p<0,01$) Türkçe testi çok boyutlu olarak kabul edilmiştir.

Çok boyutlu olduğu ortaya konan testin ölçtüğü yapıyı sınamak ve alt boyutları belirlemek amacıyla veriler faktör analizine tabi tutulmuştur. Faktör analizine geçmeden öncelikle verilerin faktör analizi için uygun olup olmadığı KMO (Kaiser-Meyer-Olkin) katsayısı ve Barlett Küresellik testi ile incelenmiştir. KMO katsayısı değeri 0.95 bulunmuş ve verilerin faktör analizi için uygun olduğu belirlenmiştir. Benzer şekilde Barlett Küresellik testi manidardır. ($p<0,01$).

Faktörleri ilk elde edilen şekli ile yorumlamak zordur. Bunun sebebi faktör elde ederken aralarında yüksek ilişki olan değişkenlerin farklı faktörler altında yer alabilmeleri ve bunun hesaplamada göz ardı edilmesidir. Tatlıdil (2002) elde edilen ilk faktörlerin döndürülmesi ile daha iyi

yorumlanabilecek basit yapılara ulaşılabileceğini belirtmiştir. Bu noktadan hareketle verilere Varimaks eksen döndürme yöntemi kullanılarak testin faktör yapısı ortaya çıkarılmış ve Çizelge 1'deki sonuçlara ulaşılmıştır.

Çizelge 1. Türkçe Testinde Yer Alan Maddelerin Döndürülmüş Faktör Analizi Sonuçları

Madde	Faktör Yükleri	
	Boyut 1	Boyut 2
T3	0,604	
T20	0,566	
T10	0,561	
T1	0,283	
T4	0,537	
T11	0,338	
T16	0,492	
T7	0,492	
T21	0,448	
T24	0,424	
T5	0,417	
T13	0,375	
T25	0,354	
T6		0,505
T8		0,553
T9		0,584
T22		0,526
T2		0,519
T14		0,495
T18		0,477
T17		0,449
T12		0,443
T15		0,398
T19		0,389
T23		0,365

Çizelge 1'deki maddelerin faktörlerde aldıkları yük değerleri incelendiğinde, 25 maddeden 13'ünün en yüksek yük değerini birinci faktörde, 12'sinin ise ikinci faktörde yer aldığı görülmektedir.

Türkçe testine ilişkin boyutları isimlendirmek amacıyla İlköğretim Türkçe Öğretim Programı incelendiğinde öğrenme alanlarının okuma, dinleme/izleme, konuşma, yazma öğrenme alanları ile dil bilgisinden oluştuğu görülmüştür. MEB'e (2005) 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. Bu nedenle boyutların isimlendirilmesinde de bu öğrenme alanları esas alınmıştır. Birinci boyutta Türkçenin yapısını oluşturan ve işleyiş kurallarını tanımlayan bilgilere ilişkin kazanımlar yer aldığı için “*dil bilgisi*”, ikinci boyut ise öğrenme, araştırma, yorumlama, tartışma, eleştirel düşünmeyi sağlayan okuma kuralları, okunan metinlerin anlaşılması ve çözümlenmesi, anlamlandırılması, okunan metinlerin değerlendirilmesi ve okumanın bir alışkanlık hâline dönüştürülmesine yönelik olarak çeşitli kazanımlar yer aldığı için “*anlam bilgisi*” olarak isimlendirilmiştir.

Araştırmada kullanılan bir diğer alt test olan matematik testi verisinin boyutluluğunun belirlenmesi için DIMTEST programı ile tetrakorik korelasyon matrisi kullanılarak lineer faktör analizi yöntemi ile bir T istatistiği üretilmiş ve hipotez testi yapılmıştır. Hipotez testinin sınanması sonucu elde edilen boyutluluk analizi sonuçları incelendiğinde ($p=0,000$, $p<0,01$) matematik testinin çok boyutlu olduğu belirlenmiştir.

Matematik testi için testin alt boyutlarını belirlemek amacıyla veriler faktör analizine tabi tutulmuştur. Verilerin faktör analizi sonucu hesaplanan; KMO katsayısı değeri 0.93 bulunmuştur ve verilerin faktör analizi için uygun olduğu belirlenmiştir. Barlett Küresellik testi ise manidardır ($p<0,01$). Elde edilen faktörleri yorumlanmasına açıklık kazandırmak amacıyla döndürme işlemi yapılmıştır. Varimaks eksen döndürme yöntemi kullanılarak testin faktör yapısı ortaya çıkarılmış ve Çizelge 2 ‘deki yapı elde edilmiştir.

Çizelge 2. Matematik Testinde Yer Alan Maddelerin Döndürülmüş Faktör Analizi Sonuçları

Maddeler	Faktör Yükleri	
	Boyut 1	Boyut 2
M8	0,578	
M9	0,527	
M1	0,496	
M20	0,488	
M21	0,474	
M4	0,445	
M7	0,438	
M15	0,424	
M5	0,421	
M11	0,394	
M3	0,386	

M19	0,385
M2	0,331
M10	0,302
M22	0,365
M23	0,508
M17	0,294
M13	0,542
M12	0,537
M25	0,535
M14	0,489
M16	0,477
M18	0,475
M24	0,474
M6	0,422

Matematik testine ilişkin boyutları isimlendirmek amacıyla İlköğretim Matematik Öğretim Programındaki öğrenme ve alt öğrenme alanları (sekizinci sınıflar için) ile bu alanlardaki kazanımlar incelenmiştir.

Öğrenme alanları ve alt öğrenme alanları incelendiğinde her bir alt öğrenme alanının bilgi ve beceri boyutunu ölçen uygun bir kazanım cümlesi hazırlanan sınavda yer almadığından boyutlar genel olarak isimlendirilmiştir. Birinci faktör temel matematiksel işlemlerden yararlanarak sayısal muhakeme yapma gücünü ölçen sorulardan oluştuğu ve matematiğin sayılarla ilgili bölümünü temel alan, sayıların özellik ve işlemlerini anlatan kolu aritmetik olarak adlandırıldığı (Türk Dil Kurumu Sözlüğü) için bu alt boyut “*aritmetik*” olarak isimlendirilmiştir. İkinci boyut ise üç boyutlu geometrik ilişkiler ve ölçme gibi üst düzey muhakeme gücünü ölçen sorulardan oluştuğu için bu boyut “*sayısal mantık*” olarak isimlendirilmiştir.

Türkçe ve matematik testlerine uygulanan faktör analizleri sonucunda her iki testin de iki alt boyutunun olduğu belirlenmiştir. Her bir alt boyut tek bir yapıyı ölçmektedir. Bu bulgular doğrultusunda verinin alt boyutlar düzeyinde tek boyutlu MTK’ye göre yetenek kestirimleri yapabilmek için karşılaması gereken tek boyutluluk varsayımını sağladığı sonucuna ulaşılmıştır. Tek boyutluluğun sağlandığında yerel bağımsızlık varsayımının da sağlanacağı görüşünden hareketle uygulama sonucunda elde edilen verilerin yerel bağımsızlık varsayımını da karşıladığı sonucu elde edilmiştir.

Türkçe, matematik testlerinin ve bu testlerin alt boyutlarından KTK’ye, tek ve çok boyutlu MTK’ye göre elde edilen puanlar üzerinde aşağıdaki analizler yapılmıştır.

Türkçe ve matematik testinden KTK'ya dayalı test puanları elde etmek amacıyla ikili puanlama yöntemi kullanılmıştır. Türkçe testinden KTK'ya dayalı puanları elde etmek için her bir cevaplayıcının doğru yanıtları toplanmıştır. Benzer işlemler matematik testi içinde yapılmıştır. Matematik testinden KTK'ya dayalı puanları elde etmek için her bir cevaplayıcının doğru yanıtları toplanmıştır. Yapılan hesaplamalar sonucunda matematik ve Türkçe testine ilişkin test puanı elde edilmiştir.

Tek boyutlu MTK'ye dayalı yetenek parametrelerinin kestiriminde her bir testin alt boyutları için yetenek kestirimi yapılmıştır. Türkçe testi için “anlam bilgisi” ve “dil bilgisi” alt boyutlarından elde edilen verilerden yetenek kestirimleri yapılmıştır. Matematik testi için ise “aritmetik” ve “sayısal mantık” alt boyutlarından elde edilen verilerden yetenek kestirimleri yapılmıştır.

Türkçe ve matematik testlerinin çok boyutlu MTK'ya dayalı yetenek parametrelerinin kestiriminde ise Beklenen Posteriori Değeri (Expected A Posteriori) yöntemi esas alınarak iki-faktör (bi-factor) modeli kullanılarak TESTFACT programıyla kestirim yapılmıştır.

Verilerin çözümlenmesine yönelik analizlere geçilmeden önce Türkçe ve matematik testlerinin alt boyutlarının MTK modellerine uyumu BILOG programıyla incelenmiştir. Tek boyutlu MTK modellerinden hangisine daha iyi uyum sağladığını belirlemek için $-2\log$ (olabilirlik) değerlerinin farkı alınarak kay-kare ile test edilmiştir.

Araştırmanın genel amacı çerçevesinde cevap aranan sorulardan birinci ve ikinci soruya cevap bulabilmek için KTK analizlerinden elde edilen puanların içerdiği hata miktarı standart hata ile tek ve çok boyutlu MTK analizlerinden elde edilen yetenek kestirimlerinin içerdiği hata miktarı ise hataların ortalama karekökü (the root mean square standart deviations-RMS) değerlerinden faydalanılarak karşılaştırılmıştır. MTK bireysel yetenek tahmini için standart hata sağlar, KTK'da olduğu gibi tüm bireyler için tek bir tahmin hatası vermez (Emberson ve Reise, 2000). Bu nedenle KTK'da elde edilen ölçmenin standart hatası ile MTK'da elde edilen standart hata aynı ölçek üzerinde değildir. MTK'da elde edilen puanların dağılımı standart sapması 1 ortalaması 0 olan bir dağılımdır. Hataların karşılaştırılabilmesi için KTK'da elde edilen test puanları standartlaştırılmıştır. Standart sapması 1 ortalaması 0 olan z puanına dönüştürülmüştür. Elde edilen standart puanlar üzerinden hata kestirimi yapılmıştır (Luecht, 2003).

Araştırmada cevap aranan üçüncü ve dördüncü soru için, öncelikle MTK'ya göre kestirilen yetenek parametreleri ile KTK'ya göre belirlenen test puanları z puanlarına dönüştürülerek aynı ölçek üzerinde incelenmesi sağlanmıştır. Test puanları ile yetenek parametreleri arasındaki ilişkiler ikişerli olarak Pearson Momentler Çarpım Korelasyon Katsayısı ile incelenmiştir

Araştırmada cevap aranan beşinci soru çerçevesinde Türkçe ve matematik testlerinden KTK'ya göre elde edilen puanlar ile tek ve çok boyutlu MTK'ya göre belirlenen yetenek kestirimlerinin güvenilirlikleri karşılaştırılmıştır. KTK'ya dayalı güvenilirlik kestirim yöntemi KR-20 güvenilirliği olarak belirlenmiştir. Tek boyutlu MTK'ya dayalı güvenilirlik BILOG programı ile hesaplanan Lord'un güvenilirlik katsayısı ile belirlenmiştir. Çok boyutlu MTK'ya dayalı güvenilirliğin kestiriminde ise TESTFACT programı ile hesaplanan ampirik güvenilirlik katsayısından yararlanılmıştır.

BULGULAR VE YORUMLAR

Araştırmanın bu bölümünde, alt amaçların veriliş sırası dikkate alınarak verilerin analizinden elde edilen bulgulara ve bunlara dayalı yorumlara yer verilmiştir.

Türkçe Testinden Elde Edilen Puanların/Yetenek Parametrelerinin Hata Kestirimleri

Türkçe testi için KTK'ya göre elde edilen puanların standartlaştırılması ile standart puanlar hesaplanmıştır. Standart puanlardan elde edilen ölçmenin standart hatası ile tek ve çok boyutlu MTK'da yetenek kestirimlerinin içerdiği standart hatayı ortaya koymada kullanılan hataların ortalama karakökü (RMS) değerleri Çizelge 3'de, verilmiştir.

Çizelge 3. Türkçe Testi Puanlarının/Yetenek Kestiriminin Standart Hataları

Standart Hatanın Elde Edildiği Kuram/ Kestirim Yöntemi	Test/Alt Boyut	Standart Hata
KTK (Ölçmenin Standart Hatası)	Türkçe	0,44
Tek Boyutlu MTK (RMS)	Anlam Bilgisi	0,42
	Dil Bilgisi	0,38
Çok Boyutlu MTK (RMS)	Türkçe	0,36

Çizelge 3 incelendiğinde Türkçe testinden KTK'ya göre belirlenen puanların hatası ile tek ve çok boyutlu MTK'ya göre kestirilen yetenek kestirimlerinin standart hataları karşılaştırıldığında en yüksek hatanın (0,44) KTK'ya göre yapılan analizlerde ortaya çıktığı belirlenmiştir.

Çok boyutlu MTK ile elde edilen kestirimlerin daha az hatalı (0,37) ve ölçme duyarlılığının fazla olduğu ortaya konmuştur. Ayrıca çok boyutlu MTK ile elde edilen yetenek kestiriminin alt boyutlar bazında tek boyutlu MTK'ya göre elde edilen yetenek kestirimlerine kıyasla kısmen daha

düşük standart hataya sahip olduğu belirlenmiştir. Araştırmanın bu bulguları çok boyutlu MTK altında elde edilen yetenek kestirimlerinin, tek boyutlu MTK ile elde edilen yetenek kestirimlerine kıyasla, daha düşük standart hatalarının olduğu belirlenen (Ackerman ve Davey, 1991; Chang,1992; Köse, 2010; Spencer, 2004) çalışmalarından elde edilen bulgularla tutarlılık göstermektedir.

Araştırma bulgularını Luecht'in (2003) KTK, tek ve çok boyutlu MTK'yi karşılaştırdığı çalışmanın bulguları da destekler niteliktedir. Belirtilen çalışmada çok boyutlu yapıya sahip bir testi tek boyutlu kabul ederek MTK'ya göre analiz edildiğinde hata miktarının KTK'ya göre elde edilen sonuçlara göre daha yüksek olduğunu; en az hatayı ise alt testler düzeyinde tek boyutlu MTK ve tüm test düzeyinde çok boyutlu MTK ile yapılan analizlerde elde edildiği vurgulanmıştır. Bu bulgulara göre araştırmacı tek boyutluluk varsayımının ihlal edilmesinin parametre kestirimlerinde farklılıklara yol açtığını belirtmiştir. Ayrıca Walker ve Bretvas (2003) yaptıkları araştırmada çok boyutlu verinin tek boyutlu kuramla incelenmesinin, bireyin yeteneğinin yorumlanmasında yanlış çıkarıma yol açabileceğini vurgulamışlardır.

Matematik Testinden Elde Edilen Puanların/Yetenek Parametrelerinin Hata Kestirimleri

Matematik testi için KTK'ya göre elde edilen puanların standartlaştırılması ile elde edilen standart puanlardan elde edilen ölçmenin standart hatası ile tek ve çok boyutlu MTK'da yetenek kestirimlerinin içerdiği standart hatayı ortaya koymada kullanılan hataların ortalama karekök (RMS) değerleri Çizelge 4'te, verilmiştir.

Çizelge 4. Matematik Testi Puanlarının/Yetenek Kestiriminin Standart Hataları

Standart Hatanın Elde Edildiği Kuram/ Kestirim Yöntemi	Test/Alt Boyutu	Standart Hata
KTK (Ölçmenin Standart Hatası)	Matematik	0,47
Tek Boyutlu MTK (RMS)	Aritmetik	0,43
	Sayısal Mantık	0,44
Çok Boyutlu MTK (RMS)	Matematik	0,37

Çizelge 4 incelendiğinde en düşük hataya (0,37) çok boyutlu MTK'ye göre belirlenen yetenek parametrelerinin kestiriminde en yüksek hataya (0,47) ise matematik testinden KTK'ye göre elde edilen puanların sahip olduğu görülmektedir.

Tek boyutlu MTK'ye göre yapılan analizlerde ortaya çıkan hata miktarının ("aritmetik" alt boyutu için 0,43, "sayısal mantık" alt boyutu için 0,44) çok boyutlu MTK'ye göre daha yüksek ancak KTK'ye göre daha düşük olduğu ortaya konulmuştur. Bu bulguları Chang'ın (1992) yaptığı okuma becerisi ve matematik alt testlerinden elde edilen gerçek ve simülasyon veri üzerinden yetenek parametrelerinin doğruluğunu tek ve çok boyutlu MTK modellerine göre karşılaştırdığı çalışmanın bulguları desteklemektedir. Bu çalışmada çok boyutlu MTK ile elde edilen sonuçların daha büyük ölçme keskinliğine sahip olduğu ve çok boyutlu kuram altında elde edilen yetenek kestirimlerinin, tek boyutlu kuram altında elde edilen yetenek kestirimlerine kıyasla, daha düşük standart hatalarının olduğu şeklinde vurgulanmıştır.

Türkçe Testinden Elde Edilen Puanlar ve Yetenek Parametreleri Arasındaki İlişki

Türkçe testi için KTK'ya göre elde edilen puanlar ile tek ve çok boyutlu MTK'ya göre kestirilen yetenek parametreleri ve tek ve çok boyutlu MTK'ya göre kestirilen yetenek parametreleri arasındaki ilişkileri ortaya koyan Pearson Momentler Çarpım Korelasyon Katsayıları Çizelge 5'te, verilmiştir.

Çizelge 5. Türkçe Testine İlişkin Test Puanları/Yetenek Kestirimleri Arasındaki İlişki

		KTK	Tek Boyutlu MTK		Çok Boyutlu MTK
		Türkçe	Anlam Bilgisi	Dil Bilgisi	Türkçe
KTK	Türkçe	-	0,91	0,93	0,98
	Anlam Bilgisi	-	-	0,71	0,92
Tek Boyutlu MTK	Dil Bilgisi	-	-	-	0,93

Çizelge 5 incelendiğinde KTK'ya göre elde edilen test puanları ile tek boyutlu MTK'ya göre anlam alt boyutundan elde edilen yetenek kestirimleri arasında yüksek ve manidar bir ilişki olduğu görülmektedir ($r=0,91$, $p<0,01$). Benzer şekilde KTK'ye göre elde edilen test puanları ile tek boyutlu MTK'ye göre "dil bilgisi" alt boyutundan elde edilen yetenek kestirimleri arasında yüksek ve manidar bir ilişki elde edilmiştir ($r=0,93$, $p<0,01$). Bu bulgular (Anderson, 1999; Courville, 2005; Çelen, 2008; Demirtaşlı, 2002; Fan, 1998; Lawson, 1991; MacDonald ve Paunonen, 2002; Ndalichako ve Rogers, 1997; Progar ve Sočan, 2008; Rogers ve Ndalichako, 2000; Tomkowicz ve Rogers, 2005; Jimelo ve Silvestre-Tipay, 2009) çalışmalarından elde edilen bulgularla tutarlılık

göstermektedir. Ayrıca KTK'ya göre elde edilen test puanları ve çok boyutlu MTK'ya göre kestirilen yetenek parametreleri arasında da yüksek ve manidar düzeyde bir ilişki bulunmuştur ($r=0,98$, $p<0,01$).

Çizelge 5 incelendiğinde “anlam bilgisi” alt boyutundan tek boyutlu MTK'ye göre elde edilen yetenek kestirimleri ile çok boyutlu MTK'ye göre elde edilen yetenek kestirimleri arasında yüksek ve manidar bir ilişki olduğu görülmektedir ($r=0,92$, $p<0,01$). Benzer şekilde “dil bilgisi” alt boyutundan tek boyutlu MTK'ye göre elde edilen yetenek kestirimleri ile çok boyutlu MTK'ye göre elde edilen yetenek kestirimleri arasında da yüksek ve manidar bir ilişki elde edilmiştir ($r=0,93$, $p<0,01$). Chang'ın (1992) yaptığı araştırmada okuma becerileri testinin alt testler bazında tek boyutlu MTK'ye göre elde edilen yetenek kestirimleri ile tüm testin çok boyutlu MTK'ye göre elde edilen yetenek kestirimleri arasında yüksek ilişki (0,91) bulunmuştur. Bu bulgular KTK, tek ve çok boyutlu MTK farklı varsayımlara sahip olmasına rağmen kuramlardan elde edilen test puanları ve yetenek kestirimleri arasında büyük bir farklılığın bulunmadığını göstermektedir.

Matematik Testinden Elde Edilen Puanlar ve Yetenek Parametreleri Arasındaki İlişki

Matematik testi için KTK'ya göre elde edilen puanlar ile tek ve çok boyutlu MTK'ya göre kestirilen yetenek parametreleri ve tek ve çok boyutlu MTK'ya göre kestirilen yetenek parametreleri arasındaki ilişkileri ortaya koyan Pearson Momentler Çarpım Korelasyon Katsayıları Çizelge 6'da, verilmiştir.

Çizelge 6. Matematik Testine İlişkin Test Puanları/Yetenek Kestirimleri Arasındaki İlişki

		KTK	Tek Boyutlu MTK	Çok Boyutlu MTK	
		Matematik	Aritmetik	Sayısal Mantık	Matematik
KTK	Matematik	-	0,93	0,85	0,99
Tek Boyutlu	Aritmetik	-	-	0,66	0,90
MTK	Sayısal	-	-	-	0,87
	Mantık	-	-	-	-

Çizelge 6 incelendiğinde KTK'ya göre elde edilen test puanları ile “aritmetik” alt boyutundan tek boyutlu MTK'ya göre elde edilen yetenek kestirimleri arasında yüksek ve manidar bir ilişki olduğu görülmektedir ($r=0,93$, $p<0,01$). Benzer şekilde “sayısal mantık” alt boyutundan tek boyutlu MTK'ya göre elde edilen yetenek kestirimleri arasında da yüksek ve manidar bir ilişkinin olduğu

sonucuna ulaşılmıştır ($r=0.85$, $p<0,01$). Araştırmanın bu sonucu Hwang'ın (2002) bulgularıyla benzerlik göstermektedir. Hwang sekizinci sınıflara uyguladığı 15 çoktan seçmeli maddeden oluşan matematik testinden KTK'ya göre elde ettiği test puanları ile Rasch modeline göre MTK'dan kestirilen yetenek parametreleri arasındaki ilişkileri incelemiş ve korelasyonu 0,98 olarak belirlemiştir.

MacDonald ve Paunonen'in (2002) simülasyon veri üzerinde yaptıkları çalışmada da KTK'ya göre elde ettiği test puanları ile iki parametrelili lojistik modele göre MTK'dan kestirilen yetenek parametreleri arasındaki korelasyon katsayılarının 0,97 ile 0,99 arasında değiştiği belirlenmiştir.

Tüm testten KTK'ya göre elde edilen test puanı ile çok boyutlu MTK'ya göre kestirilen yetenek kestirimleri arasında da ilişki yüksek olarak belirlenmiştir. Çizelge 6 incelendiğinde "aritmetik" alt boyutunda tek boyutlu MTK'ya göre elde edilen yetenek kestirimleri ile çok boyutlu MTK'ya göre elde edilen yetenek kestirimleri arasında yüksek ve manidar bir ilişki elde edilmiştir ($r=0.90$, $p<0,01$). Benzer şekilde "sayısal mantık" alt boyutunda tek boyutlu MTK'ya göre elde edilen yetenek kestirimleri ile çok boyutlu MTK'ya göre elde edilen yetenek kestirimleri arasında yüksek ve manidar bir ilişki olduğu belirlenmiştir ($r=0.87$, $p<0,01$).

Türkçe ve Matematik Testinden Elde Edilen Puanların ve Yetenek Kestirimlerinin Güvenirlikleri

Türkçe testi için KTK'da elde edilen KR-20 güvenirlilik katsayısı, tek boyutlu MTK'ya elde edilen Lord'un Güvenirlilik Katsayısı ve çok boyutlu MTK'ye elde edilen ampirik güvenirlilik katsayıları Çizelge 7'de verilmiştir.

Çizelge 7. Türkçe Testi Puanlarının ve Yetenek Kestiriminin Güvenirlilik Katsayıları

Güvenirlilik Katsayısının Elde Edildiği Kuram/ Kestirim Yöntemi	Test/Alt Boyut	Güvenirlilik Katsayısı
KTK (KR-20)	Türkçe	0,80
Tek Boyutlu MTK (Lord'un Güvenirlilik Katsayısı)	Anlam Bilgisi	0,76
	Dil Bilgisi	0,80
Çok Boyutlu MTK (Ampirik Güvenirlilik Katsayısı)	Türkçe	0,84

Çizelge 7’de verilen yetenek kestirimlerinin güvenilirlik değerleri incelendiğinde, çok boyutlu MTK’ya göre elde edilen güvenilirliğin kısmen daha yüksek (0,84) olduğu görülmektedir. En düşük güvenilirlik katsayısı (0,76) ise anlam alt boyutunda tek boyutlu MTK’ya göre elde edilen katsayıdır. KTK ve dil alt boyutunda tek boyutlu MTK’ya göre elde edilen güvenilirlik katsayıları (0,80) ise aynı çıkmıştır. Yetenek parametrelerinin kestiriminde tek ve çok boyutlu MTK modelleri kapsamında yapılan çalışmalar sonucunda çok boyutlu MTK modellerinden elde edilen güvenilirlik katsayısının tek boyutlu MTK modellerine göre daha yüksek olduğu belirlenmiştir (Köse, 2009; Chang, 1992).

Çizelge 8’de, matematik testi için KTK’da elde edilen KR-20 güvenilirlik katsayısı, tek boyutlu MTK’da elde edilen Lord’un Güvenirlik Katsayısı ve çok boyutlu MTK’da elde edilen ampirik güvenilirlik katsayıları verilmiştir.

Çizelge 8. Matematik Testi Puanlarının ve Yetenek Kestirimlerinin Güvenirlik Katsayıları

Güvenirlik Katsayının Elde Edildiği Kuram/ Kestirim Yöntemi	Test/Alt Boyutu	Güvenirlik Katsayısı
KTK (KR-20)	Matematik	0,78
Tek Boyutlu MTK (Lord’un Güvenirlik Katsayısı)	Aritmetik	0,74
	Sayısal Mantık	0,63
Çok Boyutlu MTK (Ampirik Güvenirlik)	Matematik	0,81

Çizelge 8’de verilen yetenek kestirimlerinin güvenilirlikleri incelendiğinde, çok boyutlu MTK’ye göre elde edilen güvenilirliğin (0,81) kısmen daha yüksek olduğu görülmektedir. En düşük güvenilirlik katsayısının ise alt testler bazında tek boyutlu MTK’ye göre elde edilen değer olduğu tespit edilmiştir.

KTK’den elde edilen puanların güvenilirlik değerleri ile çok boyutlu MTK’den kestirilen yetenek parametrelerinin güvenilirlik katsayıları karşılaştırıldığında çok boyutlu MTK’nin daha güvenilir sonuçlar verdiği görülmektedir. Yetenek parametrelerinin kestiriminde çok boyutlu MTK modellerinden elde edilen güvenilirlik katsayısının tek boyutlu MTK modellerine göre daha yüksek olduğu araştırmalarla (Walker ve Bretvas, 2003; Spencer, 2003, Haberman ve Sinharay, 2010 ve Wiberg, 2012) desteklenmiştir.

SONUÇ VE TARTIŞMA

Araştırmada Türkçe ve matematik testlerinden ve testlerin alt boyutlarından KTK'ya göre elde edilen puanların hata kestirimleri ile tek ve çok boyutlu MTK'ya göre belirlenen yetenek parametrelerinin hata kestirimleri karşılaştırılmıştır. Ayrıca Türkçe ve matematik testlerinden KTK'ya göre elde edilen puanlar ile tek ve çok boyutlu MTK'ya göre belirlenen yetenek parametrelerinin kestirimleri arasındaki ilişki incelenmiştir. Araştırmanın son aşamasında ise Türkçe ve matematik testlerinden KTK'ya göre elde edilen puanlar ile tek ve çok boyutlu MTK'ya göre belirlenen yetenek kestirimlerinin güvenilirlikleri belirlenmiştir. Araştırmada aşağıdaki sonuçlara ulaşılmıştır.

- Türkçe testinden ve testin alt boyutlarından çok boyutlu MTK ile elde edilen yetenek kestirimlerinin alt boyutlar bazında tek boyutlu MTK'ya göre elde edilen yetenek kestirimlerine ve KTK'ya göre elde edilen test puanlarına kıyasla kısmen daha düşük standart hataya sahip olduğu sonucuna ulaşılmıştır. KTK'nın temelinde tek boyutluluk varsayımı olduğu için tek boyutluluk varsayımının sağlanamadığı durumlarda KTK'ya dayalı analizlerde hata miktarının arttığı belirlenmiştir.

- Matematik testinden ve testin alt boyutlarından yetenek parametrelerinin kestiriminde ortaya çıkan hata miktarları karşılaştırıldığında en düşük hataya çok boyutlu MTK'ya göre kestirilen yetenek parametrelerinin; en yüksek hataya ise matematik testinden ve testin alt boyutlarından KTK'ya göre elde edilen puanların sahip olduğu gözlemlenmiştir.

- Türkçe testinden KTK'ya göre elde edilen puanlar ile tek ve çok boyutlu MTK'ya göre belirlenen yetenek parametrelerinin kestirimleri arasındaki ilişkiler incelendiğinde KTK'ya göre elde edilen test puanları ile tek boyutlu MTK'ya göre elde edilen yetenek kestirimleri arasında yüksek ve manidar ilişkiler olduğu belirlenmiştir ($r=0,91$ ve $r=0,92$, $p<0,01$). Bu durum, farklı varsayımlara sahip iki farklı kurama göre kestirilen test puanları ve yetenek kestirimlerinin benzerlik içinde olduğunu ve kuramlardan elde edilen kestirimler arasında büyük bir farklılığın bulunmadığını göstermektedir. KTK'ya göre elde edilen test puanları ve çok boyutlu MTK'ya göre kestirilen yetenek parametreleri arasında ise yüksek ve manidar düzeyde bir ilişki elde edilmiştir ($r=0,98$, $p<0,01$).

- Türkçe testinden tek boyutlu MTK'ya göre elde edilen yetenek kestirimleri ile çok boyutlu MTK'ya göre elde edilen yetenek kestirimleri arasında yüksek ve manidar bir ilişki olduğu sonucuna ulaşılmıştır ($r=0,92$, $r=0,93$, $p<0,01$). İlişkilerin yüksek olması kuramlar arasında yapılan kestirimlerde elde edilen sonuçların farklılaşmadığını göstermektedir. Bu durum iki kuramın öğrencileri yeteneklerine göre sıralamada bir fark oluşturmadığı sonucuna ulaştırmaktadır. Bu sonuç

testen her iki kurama göre elde edilen yetenek kestirimlerinin birbirlerinin yerine kullanılabileceğini göstermektedir.

- Matematik testinden KTK'ya göre elde edilen puanlar ile tek ve çok boyutlu MTK'ya göre belirlenen yetenek parametrelerinin kestirimleri arasındaki ilişki incelendiğinde KTK'ya göre elde edilen test puanları ile tek boyutlu MTK'ya göre elde edilen yetenek kestirimleri arasında yüksek ve manidar bir ilişki elde edilmiştir ($r=0,93$, $r=0,85$, $p<0,01$). Bu durum, farklı varsayımlara sahip iki farklı kurama göre kestirilen test puanları ve yetenek kestirimlerinin benzerlik içinde olduğunu göstermektedir.

- Türkçe testlerinden KTK'ya göre elde edilen puanlar ile tek ve çok boyutlu MTK'ya göre belirlenen yetenek kestirimlerinin güvenilirlikleri incelendiğinde çok boyutlu MTK'ya göre elde edilen güvenirlığın (0,84) diğer kuramlarla kestirilen güvenirlik değerlerinden kısmen daha yüksek olduğu belirlenmiştir. Bu bulgu Türkçe testi için çok boyutlu MTK'ya göre yapılan kestirimlerde elde edilen gözlenen puanların gerçek puanlara daha yakın olduğu sonucuna ulaştırmaktadır.

- Matematik testlerinden KTK'ya göre elde edilen puanlar ile tek ve çok boyutlu MTK'ya göre belirlenen yetenek kestirimlerinin güvenilirlikleri incelendiğinde, çok boyutlu MTK'ya göre elde edilen güvenirlığın (0,81) diğer kuramlarla kestirilen güvenirlik değerlerinden kısmen daha yüksek olduğu belirlenmiştir. Bu bulgu sonucunda matematik testi için çok boyutlu MTK'ya göre yapılan kestirimlerde elde edilen gözlenen puanların gerçek puanlara daha yakın olduğu söylenebilir. Diğer bir anlatımla en az hatalı kestirim çok boyutlu MTK'ya göre elde edildiği belirlenmiştir. En düşük güvenirlige ise alt testler bazında tek boyutlu MTK'ya göre kestirilen puanların sahip olduğu görülmüştür.

Araştırmanın genel bir sonucu olarak, iki farklı alt boyutu olan Türkçe ve matematik testinin tek boyutlu olarak ölçülüp tek bir yetenek kestirimi yapıldığında tanılayıcı bilgiler tam olarak verilemediği için, bu testin çok boyutlu olarak değerlendirilmesinin daha doğru olacağı düşünülmektedir. Çünkü çok boyutlu yapıya sahip testlerin alt boyutlar düzeyindeki yetenek kestiriminde daha etkili olduğu ve tanılayıcı bilgiler verdiği yapılan araştırmalarla da desteklenmektedir.

Bu sonuçlarından hareketle ve ilgili alanyazın taramalarından yararlanarak elde edilen bilgiler doğrultusunda araştırma bulgularına dayalı ve uygulayıcılara gelecekte yapılacak araştırmalara ışık tutabilecek olmak üzere aşağıdaki önerilerde bulunulmuştur.

1. Yetenek kestirimlerinin daha etkili yapılabilmesi için sınavların Türkçe ve Matematik dersi öğretim programlarında yer alan öğrenme alanlarına göre hazırlanması önerilmektedir. Sınavların bu ilke doğrultusunda hazırlanmasının ve bunun birey hakkında daha tanılayıcı karar

verme gücüne sahip olan çok boyutlu MTK modelleri kapsamında değerlendirilmesinin öğrencilerin eksik olduğu öğrenme alanını ortaya çıkaracağı ve bu şekilde öğrenci hakkında daha kapsamlı bilgi edinilebileceği düşünülmektedir.

2. Araştırmada madde ve test istatistikleri KTK'ya göre elde edilen ve bu doğrultuda geliştirilen geniş ölçekli bir test verisinin KTK ile tek ve çok boyutlu MTK'ya göre karşılaştırılması yapılmıştır. Aynı şekilde çok boyutlu MTK'ya göre hazırlanan bir test üzerinde benzer kuram araştırması yapılarak sonuçlar karşılaştırılabilir.

3. Araştırma gerçek veri üzerinden yürütüldüğünden boyutlar arasındaki ilişki manipüle edilememiştir. Aynı çalışma, boyutlar arasındaki ilişki miktarları belirlenerek ve simülasyon verileri kullanılarak yapılabilir.

4. Yapılan çalışma test sonuçlarından kestirilen yetenek parametrelerinin tahmini ile sınırlıdır. Benzer çalışma aynı veri kullanılarak farklı kuramlardan kestirilen madde parametreleri karşılaştırılarak yapılabilir.

Kaynakça

- Ackerman, T.A. (1989). Unidimensional IRT Calibration of Compensatory and Non-Compensatory Multidimensional Items. *Applied Psychological Measurement*, 13, 113–127.
- Ackerman, T. A. and Davey, T. C. (1991). Concurrent adaptive measurement of multiple abilities. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Adams, R. J., Wilson, M., and Wang, W.C. (1997). The Multidimensional Random Coefficients Multinomial Logit Model. *Applied Psychological Measurement*, 21, 1–23.
- Anderson, J.O. (1999). Does Complex Analysis (IRT) Pay Any Dividends in Achievement Testing?. *The Alberta Journal of Educational Research*, XLV, 344-352.
- Ansley, T.N. and 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.
- Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulanması*. Ankara: ÖSYM Yayınları.
- Bock, R. D., Thissen, D. and Zimowski, M. F. (1997). IRT Estimation of Domain Scores. *Journal of Educational Measurement*, 37(3), 197–211.
- Chang, Y.W. (1992). *A Comparison of Unidimensional and Multidimensional IRT Approaches to Test Information in a Test Battery*. Unpublished doctoral dissertation, University of Minnesota.
- Courville, T. G. (2005). *An Empirical Comparison of Item Response Theory and Classical Test Theory Item/Person Statistics*. Unpublished doctoral dissertation, Texas A&M University.
- Çelen, Ü. (2008). *Klasik Test Kuramı ve Madde Tepki Kuramına Dayalı Olarak Geliştirilen İki Testin Psikometrik Özelliklerinin Karşılaştırılması*. Yayımlanmamış Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- de la Torre J. and Patz R. J.(2005). Making The Most of What We Have: A Practical Application of Multidimensional IRT in Test Scoring. *Journal of Educational and Behavioral Statistics*, 30, 295–311.

- Özer Özkan, Y. (2014). A comparison of estimated achievement scores obtained from student achievement assessment test utilizing classical test theory, unidimensional and multidimensional IRT. *International Journal of Human Sciences*, 11(1), 20-44. doi: 10.14687/ijhs.v11i1.2739
-
- Demirtaşlı, Ç.N. (2002). A Study Of Raven Stndart Progressive Matrices Tests' Item Measures Under Clasic and Item Response Models: An Empirical Comparison. *Ankara University, Journal of Faculty of Educational Sciences*, 35, 1-2.
- Drasgow, F. and Parsons, C.K. (1983). Application of Unidimensional Item Response Theory Models to Multidimensional Data. *Applied Psychological Measurement*, 7, 189–199.
- Elias, S., Hattie, J., and 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.
- Embretson, S.E. and Reise, S.P. (2000). *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Fan, X. (1998). Item Response Theory and Classical Test Theory: An Empirical Comparison of Their Item/Person Statistics. *Educational and Psychological Measurement*, 58, 357–381.
- Haberman, S. J. (2008). When Can Subscores Have Value?. *Journal of Educational and Behavioral Statistics*, 33 (2), 204–229.
- Haberman, S. J. and Sinharay, S. (2010a). Reporting of Subscore Using Multidimensional Item Response Theory, *Psychometrika* 75 (2), 209–227.
- Haladyna, T. M. and Kramer, G. A. (2004). The Validity of Subscores for a Credentialing Test. *Evaluation and the Health Professions*, 27 (4), 349–368.
- Harrison, D.A. (1986). Robustness of IRT Parameter Estimation to Violations of The Unidimensionality Assumption, *Journal of Educational Statistics*, 11, 91–115.
- Hwang, D.Y. (2002). Classical Test Theory and Item Response Theory: Analitical and Empirical Comparison. Speeches/meeting paper, presented at the Annual Meeting of the Southwest Educational Research Association.
- Jimelo L. and Silvestre-Tipay. (2009). Item Response Theory and Classical Test Theory: An Empirical Comparison of Item/Person Statistics in a Biological Science Test. *The International Journal of Educational and Psychological Assessment*, 1(1), 19-31.
- Kelderman, H. (1996). Multidimensional Rasch Models for Partial-Credit Scoring. *Applied Psychological Measurement*, 20, 155–168.
- Köse, A. (2010). *Madde Tepki Kuramına Dayalı Tek Boyutlu ve Çok Boyutlu Modellerin Test Uzunluğu ve Örneklem Büyüklüğü Açısından Karşılaştırılması*. Yayımlanmamış Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Lawson, S. (1991). One Parameter Latent Trait Measurement: Do The Results Justify The Effort?. In B. Thompson (Ed.), *Advances in Educational Research: Substantive Findings, Methodological Developments*, Greenwich, CT: JAI Press, 1, 159–168.
- Luecht R. M. (2003). Applications of Multidimensional Diagnostic Scoring for Certification and Licensure Tests. *Paper presented at the meeting of the National Council on Measurement in Education*, Chicago, IL.
- MacDonald, P. and Paunonen, S. (2002). A Monte Carlo Comparison of Item and Person Statistics Based on Item Response Theory Versus Classical Test Theory. *Educational and Psychological Measurement*, 62, 921–943.
- MEB (2009). *İlköğretim Öğrencilerinin Başarılarının Belirlenmesi Raporu-Türkçe, Matematik, Fen Bilgisi, Sosyal Bilgiler*. Eğitim Araştırma ve Geliştirme Dairesi Başkanlığı.
- Ndalichako, J. L. and Rogers, W. T. (1997). Comparison of Finite State Score Theory, Classical Test Theory, and Item Response Theory in Scoring Multiple-Choice Items. *Educational and Psychological Measurement*, 57, 580–589.
- Progar, S. and Sočan, G. (2008). An Empirical Comparison of Item Response Theory and Classical Test Theory, *Horizons of Psychology*, 17 (3), 5–24.
- Rogers, W.T. and Ndalichako, J. (2000). Number-Right, Item-Response, and Finite-State Scoring: Robustness With Respect to Lack of Equally Classifiable Options and Item Option Dependence, *Educational and Psychological Measurement*, 60(1), 5–19.

Özer Özkan, Y. (2014). A comparison of estimated achievement scores obtained from student achievement assessment test utilizing classical test theory, unidimensional and multidimensional IRT. *International Journal of Human Sciences*, 11(1), 20-44. doi: 10.14687/ijhs.v11i1.2739

- Rost, J. and Carstensen, C. H. (2002). Multidimensional Rasch Measurement Via Item Component Models and Faceted Designs. *Applied Psychological Measurement*, 26, 42–56.
- Sinharay, S., Haberman, S. J., and Puhon, G. (2007). Subscores Based on Classical Test Theory: to Report or Not to Report. *Educational Measurement: Issues and Practice*, 26 (4), 21–28.
- Spencer, G.S. (2004). *The Strength of Multidimensional Item Response Theory in Exploring Construct Space That is Multidimensional and Correlated*. Unpublished doctoral dissertation, Brigham Young University.
- Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, 52, 589–617.
- Stout, W. F., Douglas, J., Junker, B. and Roussos, L.A. (1993). *DIMTEST manual*. Unpublished manuscript available from W. F. Stout, University of Illinois at Urbana-Champaign, Champaign.
- Sünbül, Ö. (2011). [Çeşitli Boyutluluk Özelliklerine Sahip Yapılarda, Madde Parametrelerinin Değişmezliğinin Klasik Test Teorisi, Tek Boyutlu Madde Tepki Kuramı ve Çok Boyutlu Madde Tepki Kuramı Çerçevesinde İncelenmesi](#). Yayınlanmamış doktora tezi, Mersin Üniversitesi Sosyal Bilimler Enstitüsü, Mersin.
- Tatlıdil, H. (2002). *Uygulamalı Çok Değişkenli İstatistiksel Analiz*. Ankara: Akademi Matbaası.
- Tate, R. L. (2004). Implications of Multidimensionality for Total Score and Subscore Performance. *Applied Measurement in Education*, 17(2), 89–112.
- Tomkowicz, J. ve Rogers, W.T. (2005). [The Use of One-, Two-, and Three-Parameter and Nominal Item Response Scoring in Place of Number-Right Scoring in the Presence of Test-Wisness](#), *The Alberta Journal of Educational Research*, 51(3), 200–215.
- Traub, R.E (1983). *A Priori Consideration In Choosing An Item Response Model*. In R.K.
- Van der Linden, W. J. and Hambleton, R. K. (Eds.) (1997). *Handbook of Modern Item Response Theory*. New York: Springer.
- Walker, C.M. ve Beretvas, S.N. (2003). Comparing Multidimensional and Unidimensional Proficiency Classifications: Multidimensional IRT As a Diagnostic Aid. *Journal of Educational Measurement*, 40 (3), 255-275.
- Way, W. D., Ansley, T.N. and 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.
- Wiberg, M. (2012). Can a multidimensional test be evaluated with unidimensional item response theory? *Educational Research and Evaluation*, 18(4): 307-320
- Yao, L. and 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.
- Yao, L. (2009). *Reporting Valid and Reliable Overall Score and Domain Score*. Paper presented at the meeting of the National Council on Measurement in Education, San Diego, CA.

[Extended English Abstract](#)

The information age ,21 century, the demand for education and expectations is gradually increasing. Society and the individual demand for education, causes difficulties the provision of training requests. For this reason, selection of students are ranked according to their achievements has become forced. Tests are widely used in education and psychology for various purposes such as while understanding of the psychological structures, diagnose and revealing relationships other structures, providing educational decisions about students, diagnosed with learning difficulties, determination of student achievement, selection and placement" .The scores of these tests constitutes a fundamental basis the accuracy of the decisions made about individuals.

In Turkey, measurement and evaluation are carried out in order to see the effect of education on student achievement national and international level and different grade levels using the standardized tests and surveys. The test results are an indication of students' achievement status to be calculated in two ways total test score or sub test scores. The large-scale test results applied at the national level that taken important decisions about students in Turkey calculated over the total test.

Many educational and psychological tests are inherently multidimensional, meaning these tests measure two or more dimensions or constructs. Provide comprehensive diagnostic information and putting out the shortcomings of each sub learning area about student increased interest in sub-test scores. The tests or items of the test measure only one latent by measurements under this assumption deciding on an individual, it has become more questionable after the 1980s. These studies revealed that the ability and item parameter estimates are affected when the assumption of unidimensionality not met.

The focus of this research is to test the estimation of achievement measurements in the test battery and to empirically compare the results after applying classical test theory, unidimensional and multidimensional item response theory models to Student Achievement Assessment Test (ÖBBS) subtests of Turkish and Mathematics. It also tries to put forward the best model that estimates students' achievement with less error as the comparison is being made. Research data was obtained by implementing 25-item Turkish and Mathematics Student Achievement Test to eighth graders.

This research focused on measurement models that the impact of the psychometric properties of scores obtained from total test or subtest. Research is descriptive study of the basic type. Descriptive research “is designed to provide a picture of a situation as it naturally happens”. It may be used to justify current practice and make judgment and also to develop theories.

The researcher first conducted preliminary study in order to determine whether the data obtained from the tests and sub dimensions of the tests met the model assumptions. Firstly, dimensions of the tests were analyzed and found to be multidimensional. Then, factor analysis was applied and sub dimensions were determined for each subtest. The results of this analysis revealed the Turkish test contains two dimensions: “reading comprehensions” and “grammar”; the mathematics test also contained two dimensions: “arithmetic” and “mathematical logic”.

From the analysis of Turkish test's data results, it is identified that the ability parameters estimated obtained from the whole test under multidimensional IRT, have partially less error scores and reached more precise measurement than ability parameters estimated obtained from unidimensional IRT on the basis of sub dimensions and test scores obtained from CTT. When the reliability of the scores obtained according to CTT, and the ability parameters estimated identified according to the unidimensional and multidimensional IRT are examined, it is noticed that the reliability, obtained from the multidimensional IRT is partially higher. When the relation between the scores and the ability parameters are examined, it is found that there is a meaningful and great relation in between the ability parameter, identified from CTT's scores and a unidimensional and multidimensional IRT and the between ability unidimensional IRT and the multidimensional IRT. Similar results were obtained in mathematics test results.

Finally, it is found that parameters, obtained within the scope of multidimensional IRT, has partially less error scores and reached more precise measurement than ability parameters estimated obtained from unidimensional IRT on the basis of sub dimensions and test scores obtained from CTT.

Özer Özkan, Y. (2014). A comparison of estimated achievement scores obtained from student achievement assessment test utilizing classical test theory, unidimensional and multidimensional IRT. *International Journal of Human Sciences*, 11(1), 20-44. doi: 10.14687/ijhs.v11i1.2739

Recommendations were made taking advantage of a literature and in light of diagnoses obtained at the end of the research. The tests is prepared according to the learning area in Turkish and mathematics curriculum is recommended for ability estimates can be made more effective. In this research, multidimensional analysis were made based on the 'TESTFACT' program. Item and ability estimation is repeated using programs MPLUS, BMIRT, MIRTE, MAXLOG, NOHARM.