

TEOG Ortak ve Mazeret Sınavındaki Türkçe ve Matematik Alt Testlerinin Psikometrik Özelliklerinin Karşılaştırılması*

The Comparison of Psychometric Properties of Standardised and Make Up Maths and Turkish Subtest Questions in TEOG

Abdullah Faruk KILIÇ **

Hülya KELECİOĞLU ***

Öz

Bu çalışmanın amacı 2013-2014 eğitim öğretim yılında uygulanan TEOG ortak ve mazeret sınavındaki Türkçe ve matematik alt testlerinin psikometrik özelliklerini karşılaştırmaktır. Türkçe ve matematik testlerinin ortak ve mazeret sınavları madde güçlüğü, madde ayırıcılık indeksleri, testlerin güvenilirliği ve geçerliği gibi psikometrik özellikler çevresinde karşılaştırmalı olarak incelenmiştir. Araştırma Türkçe ortak sınavı için 9773, matematik ortak sınavı için 9485, Türkçe mazeret sınavı için 2747 ve matematik mazeret sınavı için 2323 öğrencinin madde puanları üzerinden yürütülmüştür. Kapsam geçerliği çalışması için uzman görüşlerine başvurulmuştur. Bunun için araştırmacılar tarafından TEOG ortak ve mazeret sınavlarındaki soruların ölçtüğü kazanımları belirlemek ve karşılaştırmak amacıyla soru ve kazanım eşleştirme formu ile soru değerlendirme formu oluşturulmuştur. Araştırma sonucunda hem matematik hem de Türkçe ortak ve mazeret sınavlarının test ve madde istatistiklerinin manidar düzeyde farklı olduğu gözlenmiştir. Uzman görüşlerine dayalı olarak değerlendirilen Türkçe ve matematik testlerinin kapsamının ortak ve mazeret sınavları için benzer olduğu görülmüştür. Matematik testi ortak ve mazeret sınavlarındaki maddelerin aynı yapıyı ölçmediği fakat Türkçe testi ortak ve mazeret sınavlarındaki maddelerin benzer yapıyı ölçtüğü sonucuna ulaşılmıştır.

Anahtar Kelimeler: TEOG, geçerlik, güvenilirlik, faktör analizi, test ve madde istatistikleri

Abstract

The aim of this study is to compare the psychometric properties of Turkish and math subtests in common and make-up examination of TEOG applied in 2013-2014 school year. Turkish and math tests of the common and make-up examinations were investigated comparatively within the frame of psychometric properties such as item difficulty, item discrimination index, test reliability and validity. Research was carried out via item scores matrix of 9773 students for Turkish common exam, 9485 students for math common exam, 2747 students for Turkish make-up exam and 2323 students for math make-up exam. It was consulted to expert opinion for the content validity study. According to this, in order to determine and compare the acquisitions that is measured in common and make-up examinations of TEOG, question-acquisition matching form and question assessment form was prepared by researcher. As a result of research, it has been observed that Turkish and math subtests test and item statistics are different in significantly. Content validity which based on expert opinions is in similar level for Turkish and math subtests. The results of the study revealed that math tests of the common and make-up exams items didn't measure the same things in terms of construct validity. Unlike math, Turkish tests of the common and make-up exams items measure the same things in terms of construct validity.

Key Words: TEOG, validity, reliability, factor analysis, test and item statistics.

* Bu makale birinci yazar tarafından ikinci yazar danışmanlığında hazırlanan yüksek lisans tezinden türetilmiştir.

** Araş. Gör., Adıyaman Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Adıyaman-Türkiye, e-posta: afarukkilic@windowslive.com

*** Prof. Dr, Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Ankara-Türkiye, e-posta: hulyakelecioglu@gmail.com

GİRİŞ

Türkiye’de eğitimin hemen hemen her düzeyinde test uygulanmakta ve bunların sonucunda bireyler hakkında kararlar verilmektedir. Örneğin ortaokul düzeyinde Temel Eğitimden Ortaöğretime Geçiş (TEOG) sınavı, lise düzeyinde Yüksek Öğretime Geçiş Sınavı (YGS), Lisans Yerleştirme Sınavı (LYS), lisans düzeyinde Akademik Personel ve Lisansüstü Eğitim Sınavı (ALES), Yabancı Dil Sınavı (YDS), Kamu Personeli Seçme Sınavı (KPSS) gibi sınavlar uygulanmaktadır. Bu sınavların ortak özelliği ise bireylerin hayatını, yüksek derecede etkileme gücüne sahip olmalarıdır. Genç nüfusun fazla olduğu ülkemizde, bireyleri belli bir amaç için sıralamak ve bu sıralamaya göre bireyler hakkında karar vermek kaçınılmazdır.

Hazırlanan bir test, ölçülmesi planlanan davranış evreninden ölçülecek davranışların bir örneklemini sağlamaktadır. Bu nedenle testin ölçülmek istenilen davranış evrenini ne derecede iyi temsil ettiği ve testin ölçülmek istenilen davranış evrenini ne derece doğru ölçtüğü sorulması gereken temel sorulardır (Alıcı ve diğerleri., 2011). Bu soruların cevaplarına ulaşabilmek için testin geçerlik ve güvenilirlik çalışmalarının yapılması gerekmektedir.

Güvenirlik; test puanlarının aynı ölçme işlemlerinden geçen tekrarlı ölçmeler sonucunda tutarlı sonuç verme derecesi olduğu gibi aynı zamanda hatalardan arınık olma derecesidir (AERA, APA, ve NCME, 2014; Özgüven, 2011; Turgut ve Baykul, 2012). Geçerlik ise testin amaca hizmet etme derecesi olarak tanımlanmakta (Turgut ve Baykul, 2012) ve test puanları hakkında testin hazırlanma amacı doğrultusunda yapılan yorumların, hem kuramsal olarak hem de toplanan kanıtlarla desteklenme derecesi olarak düşünülmektedir (AERA ve diğerleri., 2014). Tanımlardan da anlaşılacağı üzere hem geçerlik hem de güvenilirlik bir derece olarak belirtilmektedir. Bir test için var ya da yok şeklinde ifade etmek yerine *yeterli* veya *yetersiz* olarak değerlendirilmektedir.

Bir ölçme aracından elde edilen sonuçlar üzerinden geçerlik ve güvenilirlik çalışmalarının yapılması, ölçümlerin ne kadar hatasız olduğu, testte ölçülmek istenilen yapıyla ne kadar örtüştüğü gibi ölçme sonuçları hakkında önemli kanıtlar sunmaktadır. Ulusal düzeyde uygulanan sınavlar için de geçerlik ve güvenilirlik çalışmaları yürütülerek ölçümlerin ne kadar hatasız yapıldığı, ölçülmek istenilen kapsamın ne kadar yoklandığı ve ölçülmek istenilen yapıyla ilişkisi araştırılmalıdır. Ulusal düzeyde gerçekleştirilen sınavlardan biri de 2013-2014 eğitim öğretim yılından beri Millî Eğitim Bakanlığı tarafından uygulanan Temel Eğitimden Ortaöğretime Geçiş Sınavı (TEOG)’dır. Görmez ve Coşkun'da (2015) TEOG için geçerlik ve güvenilirlik çalışmaları yapılmasının önemini vurgulamıştır.

TEOG ilk kez 2013-2014 eğitim öğretim yılında 8. sınıf düzeyinde uygulamaya konmuştur. Bireylerin başarılarını sürece yayarak değerlendirmeyi hedefleyen bir sınav sistemi olduğu belirtilmektedir. TEOG’da Türkçe, matematik, fen ve teknoloji, din kültürü ve ahlak bilgisi, T.C. inkılâp tarihi ve Atatürkçülük ile yabancı dil dersleri için Millî Eğitim Bakanlığı Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğüne her dönem ortak sınavlar yapılarak öğrenciler değerlendirilmektedir (Millî Eğitim Bakanlığı [MEB], 2015).

Her dönem yapılan ortak sınavlarda ağırlıklandırılmış puanların hesaplanmasında; Türkçe, matematik, fen ve teknoloji dersleri için dört; T.C. inkılâp tarihi ve Atatürkçülük, yabancı dil ile din kültürü ve ahlâk bilgisi için iki katsayısı o dersin puanı ile çarpılarak her bir dersin ağırlıklandırılmış ortak sınav puanı hesaplanmaktadır. Puanlama 700 tam puan üzerinden yapılmaktadır. Öğrencilerin; 6, 7 ve 8’inci sınıf yılsonu başarı puanları ile 8. sınıf ağırlıklandırılmış ortak sınav puanı toplanarak, elde edilen toplam, ikiye bölünüp merkezî sistemle öğrenci alan ortaöğretim kurumlarına yerleştirmeye esas puanın hesaplanmasında kullanılmaktadır. Bu hesaplamada ise puanlama, 500 tam puan üzerinden yapılmaktadır (MEB, 2015).

TEOG’da ortak sınavlar ile mazeret sınavları adı altında iki sınav yapılmaktadır. Ortak sınavlarda her ders için çoktan seçmeli 20 soru sorulmakta, değerlendirmede yanlış cevap sayısı doğru cevap sayısını etkilememektedir. Sınav süresi her ders için 40 dakika olarak verilmektedir. Mazeret sınavları ise; ortak sınavlara katılamayan öğrencilerin girdiği sınavlardır. Mazeret sınavına girecek öğrenciler “Millî Eğitim Bakanlığı Okul Öncesi Eğitim ve İlköğretim Kurumları yönetmeliğinin 23. maddesinde belirtilen hükümler çerçevesinde değerlendirilerek okul müdürlüğüne belirlenmektedir.

Bu öğrencilerin bilgileri, aynı gün il/ilçe millî eğitim müdürlüklerine bildirilerek okul müdürlüklerince e-okul sistemine sınavlar tamamlandıktan sonra 5 (beş) gün içerisinde giriş yapılmaktadır (MEB, 2015).

Mazereti okul müdürlüğünce uygun görülen öğrenciler için Bakanlıkça belirlenen tarih ve merkezlerde mazeret sınavı yapılarak geçerli bir mazeretten dolayı sınava giremeyen öğrencilere yeni bir sınav hakkı verilmektedir. Mazeret sınavları da ortak sınavlarda olduğu gibi her ders için 20 sorudan oluşmakta, sınav süresi yine 40 dakika olarak verilmekte ve yanlış cevap sayısı doğru cevap sayısını etkilememektedir. Şad ve Şahiner (2016) tarafından öğrenci, öğretmen ve veli görüşler, üzerinden yürütülen araştırmada mazeret sınavının olması ve puanlamada düzeltme formülünün kullanılmamasının olumlu karşılandığı gözlenmiştir.

TEOG sınavlarında belirtilen altı dersten; bir dönemde iki sınav yapılan derslerin ilk sınavı, bir dönemde üç sınav yapılan derslerin ise ikinci sınavı ortak sınav şeklinde uygulanmaktadır.

TEOG sınavlarında ortak ve mazeret sınavı olmak üzere iki sınav yaklaşık 1 ay ara ile yapılmaktadır. Bu durumda her iki sınavda uygulanan testlerin ve testlerde yer alan maddelerin madde güçlüğü, madde ayırıcılığı, maddelerin ölçtüğü kazanımlar, test sonuçlarının güvenilirlik düzeyi ve testlerin ölçtüğü yapılar gibi test ve madde özelliklerinin birbiriyle eşdeğer olup olmadığı sorunu ortaya çıkmaktadır. Bireylerin aynı şekilde değerlendirildiği, aynı ölçüt üzerinden karar verildiği iki test ve testte bulunan maddeler birbirine eşdeğer olmadığında, bir grubun dezavantajlı olması durumu ortaya çıkmaktadır. Bunun için de bu testlerin psikometrik nitelikleri incelenmeli, eşdeğerlikleri sağlanmalı ve gelecek testlerde bunlara yönelik önlemler alınmalıdır.

TEOG ile ilgili yapılan araştırmalarda kapsam geçerliğini ve sınavlardaki maddelerin hangi kazanımla ilişkili olduğunu inceleyen (Dalak, 2015; Karadeniz, Eker, ve Ulusoy, 2015; Kardeş-Birinci, 2014; Yılmaz, 2014; Yılmaz-Koçar ve Aygün, 2015; Yorgancı, 2015) çalışmalarla birlikte TEOG sistemine yönelik öğrenci ve öğretmen görüşlerine göre değerlendirilmesi ön plana çıkmaktadır (Atıla ve Özeken, 2015; Dinç, Dere, ve Koluman, 2014; Karaca, Bektaş, ve Armağan, 2015; Karaca, 2015; Zayimoğlu-Öztürk ve Aksoy, 2014). Bu araştırmaların yanında Özkan, (2015) tarafından 2013-2014 eğitim öğretim yılının I. döneminde gerçekleştirilen TEOG sınavlarından matematik ortak ve mazeret sınavlarının test eşitleme çalışması yapılmıştır. TEOG sınavlarının psikometrik özelliklerine yönelik yapılacak araştırmaların da bu sınavların geleceğini belirleme konusunda öneme sahip olduğu düşünülmektedir.

Araştırmanın Amacı

Bu çalışmanın amacı 2013-2014 eğitim öğretim yılında uygulanan TEOG ortak ve mazeret sınavındaki Türkçe ve matematik alt testlerinin psikometrik özelliklerini karşılaştırmaktır. Bu amaçla çalışmada Türkçe ve matematik alt testlerine ait ortak ve mazeret sınavlarının;

1. Test ve madde istatistikleri
2. Kapsam geçerlikleri ve
3. Yapı geçerlikleri karşılaştırmalı olarak incelenmiştir.

YÖNTEM

Araştırmanın Modeli

Bu çalışma tarama modellerinden, genel tarama modelinde yürütülen bir araştırmadır. Genel tarama modelinde çok sayıda elemandan oluşan bir evrende, evren hakkında genel bir kanıya varmak amacıyla evren üzerinde ya da örnekleminde yapılan tarama düzenlemeleridir (Karasar, 2014). Ayrıca düzeyi bakımından bu araştırma betimsel bir araştırmadır. Betimsel araştırmalarda verilen bir durum dikkatli bir şekilde tanımlamaya çalışılır (Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz, ve Demirel, 2013).

Evren ve Örneklem

Araştırmanın iki farklı evreni bulunmaktadır. Bunlardan birincisi 2013-2014 eğitim öğretim yılının ikinci döneminde uygulanan TEOG Türkçe ve matematik ortak sınavına katılan 1.271.284 öğrencidir. Aynı yıl gerçekleştirilen TEOG Türkçe ve matematik mazeret sınavına katılan 2890 öğrenci araştırmanın ikinci evrenini oluşturmaktadır.

Araştırmanın örneklemini ortak sınavlara giren öğrencilerden rastgele olarak seçilen 10.000 öğrenci oluşturmaktadır. Örneklem incelenip tüm sorulara aynı cevabı veren ya da sistematik olarak cevapları tekrarlayan bireylerin cevapları temizlenerek matematik ortak sınavı için 9485, Türkçe ortak sınavı için 9773 öğrencinin verileri üzerinden analizler yapılmıştır. Mazeret sınavlarından örneklem alınmamış evren üzerinde analizler gerçekleştirilmiştir. Veri temizleme işlemi yapıldıktan sonra Türkçe mazeret sınavında 2747 öğrencinin, matematik mazeret sınavında ise 2323 öğrencinin cevapları üzerinden analizler gerçekleştirilmiştir.

Veri Toplama Araçları

Araştırmada, TEOG Türkçe ve matematik alt testlerinin ortak ve mazeret sınavlarının verilerinin yanı sıra araştırmacılar tarafından geliştirilen ve aşağıda açıklanan formlar kullanılmıştır.

Soru-Kazanım Eşleştirme Formu

Uzman görüşlerine göre TEOG ortak ve mazeret sınavlarındaki soruların ölçtüğü kazanımları belirlemek ve karşılaştırmak amacıyla Soru ve Kazanım Eşleştirme Formu oluşturulmuştur. Türkçe ve matematik derslerinin ortak ve mazeret sınav soruları bir tablo halinde alan uzmanlarına verilmiş ve soruların karşılıklarına, sorunun hangi kazanımı ölçtüğünü belirtmeleri istenmiştir. Türkçe ve matematik derslerinin kazanım listesi ise uzmanlara farklı bir formda verilmiştir.

Soru Değerlendirme Formu

Maddelerin kapsam geçerlik oranı indeksini (KGO_i) hesaplamak amacıyla Lawshe (1975) tarafından önerilen yöntemle göre alan uzmanlarının maddeleri “gerekli/önemli”, “yararlı ama önemli değil” ve “gerekli” şeklinde sınıflandırabileceği Soru Değerlendirme Formu hazırlanmıştır.

Soru kazanım eşleştirme formu ve soru değerlendirme formu 14 matematik ve 10 Türkçe alan uzmanınca değerlendirilmiştir. Matematik alan uzmanlarından 7’si ilköğretim matematik öğretmenliği lisans mezunu; 1’i matematik eğitimi alanında yüksek lisans; 6’sı ise matematik öğretmenliği tezsiz yüksek lisans mezunudur. Türkçe alan uzmanlarından 4’ü Türkçe öğretmenliği lisans mezunu, 1’i Türkçe eğitimi yüksek lisans mezunu, 5’i ise Türk dili ve edebiyatı öğretmenliği tezsiz yüksek lisans mezunudur.

Verilerin Analizi

Türkçe ve matematik alt testlerinin ortak ve mazeret sınavları için test ve madde istatistikleri hesaplanarak; testlerin aritmetik ortalamaları arasında fark olup olmadığı bağımsız örneklem t-testi ile, madde ayırıcılık ortancaları arasında fark olup olmadığı Mann-Whitney U testi ile, testlerin ortama güçlükleri ve madde güçlükleri arasında fark olup olmadığı iki oran arasındaki farkı test etmeye yarayan z testi ile, varyansları arasında fark olup olmadığı F testi ile ve güvenilirlikleri arasında fark olup olmadığı korelasyonların test edilmesinde kullanılan z testi ile karşılaştırılmıştır. İç tutarlık anlamında güvenilirlik belirlenirken konjenerik ölçmelerde kullanılması önerilen (Yurdugül, 2006) McDonald’ın ω katsayısı hesaplanmıştır. Güvenirlik katsayısının hesaplanmasında Factor 10.3 programı, madde ayırıcılık indekslerinin hesaplanmasında ise ITEMAN 3.5 analiz programı diğer hesaplamalarda ise Office 2014 paket programı içerisinde bulunan Excel programı kullanılmıştır.

İstatistiksel sınavın yanında etki büyüklüklerinin de verilmesi araştırmaların pratikteki anlamlılığını da ortaya koyacağı ve daha nitelikli sonuçların elde edilmesini sağlayacağı için raporlanması bir gereklilik haline gelmiştir (American Educational Research Association, 2006;

American Psychological Association, 2010; Özsoy ve Özsoy, 2013). Bu çalışmada ortalamalar arasındaki etki büyüklüğünü belirlemek için Cohen's d hesaplanmıştır. Cohen's d istatistiği işaretine bakılmaksızın 0,20 küçük, 0,50 orta ve 0,80 ise büyük etki gücüne sahip olduğunu ifade etmektedir (Büyüköztürk, 2013; Cohen, 1988, 1992; Cumming, 2012). Oranlar arasındaki etki büyüklüğünü belirlemek için Cohen's h istatistiği kullanılmıştır (Cohen, 1988, 1992). Elde edilen h değerlerinin yorumu d değerlerinin yorumu gibidir. Varyanslar arasındaki farkın etki büyüklüğü yine Cohen's d istatistiği ile belirlenmiş ancak, varyansların anlamlılığı F testi ile yapıldığı için F testine yönelik olarak elde edilen formül kullanılmıştır (Thalheimer ve Cook, 2002). Bu istatistiğin yorumu da h istatistiğiyle ayıdır. Güvenirlik katsayıları arasındaki farkın etki büyüklüğünü belirlemek için Cohen's q istatistiği, madde ayırıcılık ortancaları arasındaki farkın etki büyüklüğünü belirlemek için ise r istatistiği kullanılmıştır (Cohen, 1988, 1992). Cohen's q ve Cohen's r istatistikleri 0,10 küçük, 0,30 orta ve 0,50 büyük etki büyüklüğüne sahip olduğu şeklinde yorumlanmaktadır (Cohen, 1992).

Soru kazanım eşleştirme formu verilerine göre kazanımların, öğrenme alanlarına göre dağılımı belirlenmiş ve öğrenme alanını temsil etme oranı hesaplanmıştır. Uzman görüşlerinden elde edilen Kapsam Geçerlik İndeksleri (KGİ) ve Kappa istatistiği arasında uyum olup olmadığını belirlemek için Soru Değerlendirme Formu verileri kullanılmıştır.

Maddelerin KGO_i değerlerini hesaplamak amacıyla Soru Değerlendirme Formu verileri kullanılmıştır. Alan uzmanlarının maddeleri sınıflandırması sonucunda her bir madde için kapsam geçerlik oranı elde edilmiştir (Yurdugül ve Bayrak, 2012). Kapsam geçerlik oranı her bir madde için;

$$KGO_i = \frac{N_G - \frac{N}{2}}{\frac{N}{2}}$$

şeklinde ifade edilmektedir. Burada KGO_i , i. maddenin kapsam geçerlik oranını, N_G i. madde için gerekli diyen uzman sayısını, N ise toplam uzman sayısını göstermektedir. KGO_i -1 ile +1 arasında değer almaktadır. Tüm maddeler için elde edilen KGO_i değerlerinin ortalaması ise KGİ değerini vermektedir.

Bu değerler arasında manidar bir fark olup olmadığı ise z testi ile belirlenmiştir. Bu hesaplamalarda Office 2014 paket programı içerisinde bulunan Excel programı kullanılmıştır.

Yapı geçerliğine yönelik kanıt toplamaya başlamadan önce verilerin analiz varsayımlarını sağlayıp sağlamadığı incelenmiştir. Öncelikle puan dağılımlarının normallığı incelemek amacıyla Kolmogorov Smirnov testi yapılmış bu testin sonucu manidar bulunmasına rağmen puan dağılımlarının histogram grafikleri incelenip dağılımların çarpıklık ve basıklık katsayıları gözden geçirildiğinde verilerin normal dağılıma uygun olduğu sonucuna ulaşılmıştır. K-S testinin manidar çıkmasının nedeninin örneklem büyüklüğü olduğu düşünülmüştür. Çarpıklık değerleri Türkçe testinin ortak sınavı için -0,63, mazeret sınavı için 0,35; matematik testinin ortak sınavı için 0,65, mazeret sınavı için 1,61 olduğu gözlenmiştir. Basıklık değerlerinin ise Türkçe testinin ortak sınavı için -0,58, mazeret sınavı için -0,89; matematik testinin ortak sınavı için -0,36, mazeret sınavı için 3,25 olduğu gözlenmiştir. Verilerin normal dağılımdan aşırı sapma göstermediği söylenebilir. Buna göre değişken dağılımlarının normallik sayılısını sağladığı iddia edilebilir (George ve Mallery, 2012; West, Finch, ve Curran, 1995).

Veri temizleme işlemi sonrası veri setlerinde kayıp veri olup olmadığı incelenmiş ve veri setlerinde kayıp veri bulunmadığı gözlenmiştir.

Veri setlerinde uç değerlerin bulunup bulunmadığı Mahalanobis uzaklıkları kullanılarak incelenmiş ve uç değerlerin bulunmadığı görülmüştür. Türkçe testi ortak sınavının Mahalanobis uzaklık değerlerinin 43,06 ile 2,13; mazeret sınavının ise 33,57 ile 4,59 arasında değiştiği ve ki-kare istatistiğinin anlamlı olmadığı gözlenmiştir ($p>0,001$). Benzer şekilde matematik testi ortak sınavının Mahalanobis uzaklık değerlerinin 36,24 ile 7,39; mazeret sınavının ise 36,42 ile 3,62 arasında değiştiği ve ki-kare istatistiğinin anlamlı olmadığı gözlenmiştir ($p>0,001$). Bu değerlere göre veri setlerinde uç değerlerin bulunmadığı yorumu yapılabilir.

Veri setlerinde çoklu bağlantı problemi olup olmadığını belirlemek için Tolerans değerleri, varyans şişkinlik faktörü ve koşul indeksleri incelenmiştir. İncelemeler sonucunda Tolerans değerinin 0,01'den büyük, varyans şişkinlik faktör değeri 10'dan küçük ve koşul indeksi 30'dan küçük olduğu gözlenmiştir. Buna göre çoklu bağlantı olmadığı söylenebilir (Kline, 2011; Tabachnik ve Fidell, 2012).

Yapı geçerliklerinin hangi düzeyde olduğunu araştırmak amacıyla Türkçe ve matematik testlerinin ortak ve mazeret sınavlarının faktör yapıları açıcı ve doğrulayıcı faktör analizleri ile araştırılmıştır. Faktör analizine başlamadan önce verilerin faktörleşmeye uygun olup olmadığı test edilmiştir. Bu amaçla Kaiser-Meyer-Olkin İstatistiği ve Bartlett Küresellik Testi sonuçları incelenmiştir. KMO değerinin yorumlanmasında Kaiser ve Rice (1974) tarafından belirtilen değerler dikkate alınmıştır.

KMO ve Bartlett Küresellik Testi sonuçları incelendiğinde Türkçe ortak sınavından elde edilen verilerin KMO değeri 0,96 ve Bartlett küresellik testi sonucu ise $\alpha=0,05$ düzeyinde anlamlı bulunmuştur. ($\chi^2=43512,92$; $sd=190$; $p=0,00$). Türkçe mazeret sınavından elde edilen verilerin KMO değeri 0,94 ve Bartlett küresellik testi sonucu ise $\alpha=0,05$ düzeyinde anlamlı bulunmuştur. ($\chi^2=10446,33$; $sd=190$; $p=0,00$).

Matematik testi ortak sınavından elde edilen verilerin KMO değeri 0,92 ve Bartlett küresellik testi sonucu ise $\alpha=0,05$ düzeyinde anlamlı bulunmuştur. ($\chi^2=12648,95$; $sd=190$; $p=0,00$). Matematik testi mazeret sınavından elde edilen verilerin KMO değeri 0,84 ve Bartlett küresellik testi sonucu ise $\alpha=0,05$ düzeyinde anlamlı bulunmuştur. ($\chi^2=1722,32$; $sd=190$; $p=0,00$). Bu sonuçlara göre tüm verilerin faktör analizine uygun olduğu söylenebilir.

DFA için bu araştırmada kullanılan uyum ölçütleri Schermelleh-Engel, Moosbrugger, ve Müller (2003)'den derlenerek aşağıdaki tabloda belirtilmiştir.

Tablo 1.Uyum Ölçütleri

Uyum ölçütleri	İyi Uyum	Kabul Edilebilir Uyum
χ^2/sd	$0 \leq \chi^2/sd \leq 2$	$2 \leq \chi^2/sd \leq 3$
p değeri	$0,05 < p \leq 1,00$	$0,01 \leq p \leq 0,05$
RMSEA	$0 \leq RMSEA \leq 0,05$	$0,05 \leq RMSEA \leq 0,10$
SRMR	$0 \leq SRMR \leq 0,05$	$0,05 \leq SRMR \leq 0,10$
NFI	$0,95 \leq NFI \leq 1,00$	$0,90 \leq NFI \leq 0,95$
NNFI	$0,97 \leq NNFI \leq 1,00$	$0,95 \leq NNFI \leq 0,97$
CFI	$0,95 \leq CFI \leq 1,00$	$0,90 \leq CFI \leq 0,95$
GFI	$0,95 \leq GFI \leq 1,00$	$0,90 \leq GFI \leq 0,95$
AGFI	$0,90 \leq AGFI \leq 1,00$	$0,85 \leq AGFI \leq 0,90$

Uyum ölçütleri incelendiğinde χ^2/sd oranının 3'ten küçük olduğu durumda mükemmel uyum, 5'ten küçük olduğu durumda orta düzey uyum olduğunu belirten kaynaklar bulunmaktadır (Kline, 2011; Meydan ve Şeşen, 2015). Fakat bu oranın büyük ya da küçük olması araştırmacının bakış açısıyla ilişkilidir. Kesin olarak tanımlanan bir standart bulunmamaktadır. Uygulamada, uyumlu olduğu bilinen modellerde bile, önerilen bu oran 3,4,5 ve hatta 5'ten büyük de çıkabilmektedir. χ^2/sd oranının yorumlanması χ^2 istatistiğinin büyük örneklerde büyük çıkma eğilimini ortadan kaldırmamaktadır. Büyük örnekler büyük χ^2 istatistiği üretmekte bu da χ^2/sd oranının büyümesine neden olmaktadır. Çünkü serbestlik derecesi örneklem büyüklüğü kullanılarak hesaplanmamaktadır (Mueller, 1996).

AFA ve DFA'yı gerçekleştirmek amacıyla veriler öncelikle rastgele iki gruba ayrılmıştır. Grubun birinde AFA için tetrakorik korelasyon katsayılarına dayalı olarak faktör analizi yapılmış diğer grupta ise DFA uygulanmıştır. AFA için Statistica 10 ve Factor 10.3, DFA için Factor 10.3 ve Lisrel 8.80 programları kullanılmıştır.

BULGULAR

Araştırma sonucunda elde edilen bulgular, sırasıyla test ve madde istatistiklerinin karşılaştırılması, kapsam geçerliklerinin karşılaştırılması ve yapı geçerlik düzeylerinin karşılaştırılması başlıklarıyla verilmiştir.

Türkçe Testlerinin Test İstatistiklerinin Karşılaştırılması

Türkçe testine ait test istatistiklerinin ortak ve mazeret sınavlarına göre farklılık gösterip göstermediği sınanmış ve sonuçlar Tablo 2’de belirtilmiştir.

Tablo 2. Türkçe Testi Ortak ve Mazeret Sınavlarının Test İstatistiklerinin Karşılaştırılması

Test Adı	\bar{X}	t	Cohen’s d	\bar{p}	z	Cohen’s h	S_x^2	F	Cohen’s d
Türkçe-Ortak	14,30			0,72			22,58		
Türkçe-Mazeret	9,93	41,12*	0,91	0,49	21,49*	0,48	24,73	12,05*	0,07

*p<0.05

Ortak ve mazeret sınavlarının aritmetik ortalamaları arasında fark ortak sınav lehine manidar bulunmuştur ($t(4257,661)=41,12$, $p<0,05$). Bu karşılaştırmaya ait etki büyüklüğü değeri ise 0,91 olarak elde edilmiştir. Buna göre yüksek derecede etki büyüklüğüne sahip olduğu ve bu farkın örneklemelerden kaynaklanmadığı söylenebilir. Aynı zamanda sınavların ortalama güçlükleri incelendiğinde Türkçe testinin ortak sınavının ortalama güçlüğü 0,72, mazeret sınavının ortalama güçlüğü ise 0,49 olduğu gözlenmiştir. Ortak sınav ortalamasının, mazeret sınavına göre daha yüksek ve ortalama güçlüklerinin manidar düzeyde farklı olduğu ($z(12519)=21,49$, $p<0,05$) ve ortalama güçlük farkının etki büyüklüğü orta düzeyde (0,47) olarak bulunmuştur. Bu bulgulara dayalı olarak mazeret sınavının öğrenciler için bir dezavantaj oluşturdu yorumu yapılabilir.

Ortak ve mazeret sınavlarının varyansları arasında mazeret sınavı lehine manidar düzeyde fark bulunmuştur ($F(1,12519)=12,05$, $p<0,05$). Bu durum ortak sınava katılan grubun mazeret sınavına katılan gruba göre daha homojen bir grup olduğu anlamına gelmektedir. Ancak etki büyüklüğü incelendiğinde Cohen’s d istatistiği 0,07 olarak bulunduğu görülmektedir. Buna göre varyanslar arasındaki farkın çok küçük bir etki büyüklüğüne sahip olduğu, yani bu farkın örneklem büyüklüğünden kaynaklanmış olabileceği söylenebilir.

Tablo 3. Türkçe Testi Ortak ve Mazeret Sınavlarının Test İstatistiklerinin Karşılaştırılması

Test Adı	Güvenirlilik	z	Cohen’s q	r_{jk} Ortancası	Mann-Whitney U	r
Türkçe-Ortak	0,94			0,57		
Türkçe-Mazeret	0,91	9,74*	0,21	0,64	193,00	0,03

*p<0.05

Tablo 3 incelendiğinde ortak ve mazeret sınavlarının McDonald ω değerlerinin karşılaştırılması sonucu ortak sınav lehine manidar düzeyinde fark bulunduğu görülmektedir ($z(12518)=9,74$, $p<0,05$). Bulunan bu farka yönelik elde edilen etki büyüklüğü incelendiğinde 0,21 olduğu görülmektedir. 0,10 küçük, 0,30 orta derecede etki büyüklüğüne sahip olduğu yönünde yorumlanmaktadır. Bu durumda güvenilirliklerin etki büyüklüğünün orta düzeyin altında olduğu söylenebilir. Ortak ve mazeret sınavlarının güvenilirliklerinin istatistiksel olarak farklı olması ölçme sonuçlarına karışan hata miktarlarının farklı olması anlamına gelebilir. Ancak bu sonuca analize alınan veri sayılarının farklı olması da neden olabilmektedir. Bu nedenle elde edilen etki büyüklüğü değerleri, güvenilirlik katsayıları arasındaki farkta örneklem büyüklüklerinin etkili olduğunu göstermektedir.

Ortak sınav ve mazeret sınavının madde ayırıcılık indeksleri ortancası (medyan) karşılaştırıldığında anlamlı fark gözlenmemiştir ($U=193$, $z=-0,19$, $p>0,05$, $r=0,03$). Etki büyüklüğünün çok düşük (0,03) olduğu görülmektedir. Ortak sınav ve mazeret sınavının bilen ve bilmeyen bireyleri birbirinden ayırma düzeylerinin benzer olduğu söylenebilir.

Türkçe ortak ve mazeret sınavlarının test istatistikleri incelendiğinde madde ayırıcılık indeksleri ortancaları dışındaki istatistiklerinin tümünün farklılaştığı görülmektedir. İki testin paralel olabilmesi için madde sayıları, güvenilirlikleri, varyansları, ortalamalarının eşit madde tiplerinin aynı olması gerekmektedir (Gulliksen, 1950). Ayrıca aynı kapsamı ya da yapıyı ölçmeli, her iki test de içerdiği maddelerle ölçülecek yapıyı eşit şekilde ölçmelidir (Bracken ve diğerleri., 2013). Burada testlerin paralel olup olmadığına karar verilirken testler aynı gruba uygulandıktan sonra adı geçen özellikleri incelenmektedir. Bu çalışma için iki farklı testi aynı gruba uygulama imkânı olmadığından ve testi alan grupların farklı olmasına rağmen tümünün sekizinci sınıfta olduğu göz önüne alınıp örneklem büyüklüklerinin de yeterli olduğu düşünüldüğünde testlerin özelliklerinin bu denli farklılaşmaması beklenebilir. Tüm sonuçlar birlikte değerlendirildiğinde Türkçe testi ortak ve mazeret sınavlarının psikometrik özelliklerinin farklılaşması nedeniyle birbirinin alternatifi olarak kullanılmayacağı ifade edilebilir.

Türkçe Testlerinin Madde İstatistiklerinin Karşılaştırılması

Türkçe testi ortak ve mazeret sınavlarında aynı kazanımı ölçen 11 madde bulunmaktadır (Tabloda 1-11. sıradaki maddeler). Kalan 9 maddenin 7'si aynı alt öğrenme alanında yer almaktadır (Tabloda 12-17. sıradaki maddeler). Hem aynı alt öğrenme alanından olmayan hem de aynı kazanımı ölçmeyen 2 madde ise diğer 2 madde ile eşleştirilerek karşılaştırılmıştır (Tabloda 19 ve 20. sıradaki maddeler). Eşleşen maddelerin madde istatistikleri arasında fark olup olmadığı araştırılmıştır. Türkçe ortak ve mazeret sınavında bulunan maddelerin madde güçlük indeksi ve madde ayırıcılık indeksleri açısından karşılaştırılması Tablo 4'te verilmiştir.

Tablo 4. Türkçe Testi Ortak ve Mazeret Sınavlarının Madde Güçlük ve Ayırıcılık İndekslerinin Karşılaştırılması

S. No	Ortak Sınav Madde No	Mazeret Sınavı Madde No	p_j^*	p_j^{**}	z	Cohen's h	r_{jx}^*	r_{jx}^{**}	z	Cohen's q
1	1	1	0,66	0,41	23,18*	0,51	0,69	0,46	16,23*	0,35
2	2	3	0,63	0,40	21,50*	0,46	0,68	0,54	10,41*	0,22
3	6	8	0,64	0,61	3,16*	0,06	0,68	0,63	4,06*	0,09
4	7	5	0,82	0,69	13,95*	0,30	0,53	0,65	-8,57*	-0,19
5	8	7	0,74	0,65	9,87*	0,20	0,56	0,66	-7,41*	-0,16
6	10	14	0,80	0,64	16,78*	0,36	0,53	0,68	-11,06*	-0,24
7	12	16	0,83	0,53	33,25*	0,66	0,47	0,67	-13,92*	-0,30
8	14	12	0,49	0,39	8,77*	0,20	0,68	0,5	12,95*	0,28
9	15	10	0,64	0,34	27,73*	0,61	0,63	0,44	12,46*	0,27
10	19	19	0,65	0,46	17,90*	0,38	0,67	0,53	10,21*	0,22
11	20	20	0,66	0,38	27,13*	0,57	0,46	0,44	1,16	0,03
12	5	4	0,92	0,54	47,65*	0,92	0,27	0,78	-35,57*	-0,77
13	18	17	0,50	0,34	14,98*	0,33	0,67	0,43	16,24*	0,35
14	4	2	0,81	0,65	18,15*	0,36	0,53	0,66	-9,38*	-0,20
15	9	9	0,73	0,48	24,01*	0,52	0,57	0,71	-11,09*	-0,24
16	11	6	0,84	0,32	54,83*	1,12	0,46	0,32	7,67*	0,17
17	13	13	0,84	0,57	29,94*	0,61	0,46	0,75	-22,02*	-0,48
18	16	11	0,65	0,63	2,04	0,04	0,77	0,72	5,22*	0,11
19	3	18	0,76	0,34	41,62*	0,87	0,49	0,52	-1,86	-0,04
20	17	15	0,69	0,55	13,45*	0,29	0,60	0,66	-4,61*	-0,10

* $p<0,05$, * Ortak Sınav, ** Mazeret Sınavı

Tablo 4 incelendiğinde madde güçlük indekslerinin ortak sınav için 0,49 ile 0,92; mazeret sınavı için 0,32 ile 0,69 arasında değiştiği görülmektedir. Ortak sınavda çok zor ve zor madde bulunmamaktadır. Mazeret sınavında ise çok zor ve çok kolay madde yoktur. Madde güçlüğü bakımından ortak sınavda orta güçlükte 2, kolay 12 ve çok kolay 6 madde bulunurken, mazeret sınavında zor 6, orta güçlükte 8 ve kolay 6 madde bulunmaktadır.

Mazeret sınavının gruba daha zor geldiği ve soruların güçlük düzeyi dağılımının ortak sınavdaki gibi olmadığı söylenebilir. Aynı ya da benzer kazanımı ölçen maddelerin madde güçlük düzeylerinin ortak sınavdaki 16. ve mazeret sınavındaki 11. maddeler için istatistiksel olarak anlamlı bir farklılık olmadığı diğer maddelerin ise istatistiksel olarak anlamlı düzeyde farklılaştığı görülmektedir. Bu durum, testi alan bireylerin farklı güçlükteki maddelere yanıt verdiğini göstermektedir. Birbirinin yerine kullanılan testlerin madde güçlüklerinin bu kadar farklılaşması testin adalet yönünü zedeleyeceği düşünülmektedir. Madde güçlük indeksi karşılaştırmalarının etki büyüklükleri incelendiğinde ise 0,04 ile 1,12 arasında değişen etki büyüklüklerine sahip oldukları görülmektedir. Bu karşılaştırmalara göre 2 karşılaştırmanın etki büyüklüğü çok küçük, 8 karşılaştırmanın etki büyüklüğü küçük, 7 karşılaştırmanın etki büyüklüğü orta ve 3 karşılaştırmanın ise etki büyüklüğü çok yüksektir. Buna göre maddelerin yarısının etki büyüklüklerinin orta ve üzerinde olduğu söylenebilir.

Tablo 4'te madde ayırıcılık indekslerinin ortak sınav için 0,27 ile 0,77; mazeret sınavı için ise 0,32 ile 0,78 arasında değiştiği görülmektedir. Herhangi bir maddenin, teste alınabilmesi için 0,30 ve üzerinde madde ayırıcılık indeksine sahip olması gerekmektedir (Erkuş, 2014). Ortak sınav bu açıdan incelendiğinde 5. maddenin madde ayırıcılık indeksinin 0,27 olduğu görülmektedir. Diğer 19 madde ise 0,41 ve üzerinde madde ayırıcılık indeksine sahiptir. Mazeret sınavında ise tüm maddeler 0,30'un üzerinde madde ayırıcılık değerine sahiptir.

Aynı ya da benzer kazanımı ölçen maddelerin, madde ayırıcılık indeksleri karşılaştırıldığında 2 madde haricindeki tüm maddelerde manidar fark bulunmuştur. Aynı kazanımı ölçen maddelerin büyük çoğunluğu madde ayırıcılık indeksi olarak farklılaşmasına rağmen testlerin madde ayırıcılık indeksi ortancaları farklılaşmamaktadır. Etki büyüklükleri işaretleri dikkate alınmadan incelendiğinde 0,03 ile 0,77 arasında değiştiği görülmektedir. 13 karşılaştırmanın küçük, 5'inin orta ve 2'sinin ise yüksek etki gücüne sahip olduğu söylenebilir.

Türkçe Testlerinin Kapsam Geçerlik Oranı İndekslerinin Karşılaştırılması

Tablo 5'te alt öğrenme alanlarına göre öğretim programında bulunan kazanım yüzdesi, ortak sınav ve mazeret sınavında bu alt öğrenme alanından kaç madde bulunduğu ve bu maddelerin yüzdesi verilmiştir.

Tablo 5. Türkçe Testi Ortak ve Mazeret Sınavında Bulunan Maddelerin Alt Öğrenme Alanına Göre Dağılımı

Alt Öğrenme Alanı	Öğretim Programındaki Kazanım Yüzdesi (%)	Ortak Sınavda Bulunan Madde Sayısı	Ortak Sınavda Bulunan Madde Yüzdesi	Mazeret Sınavında Bulunan Madde Sayısı	Mazeret Sınavında Bulunan Madde Yüzdesi
Anlatım Bozukluklarını Belirleme ve Düzeltme	3,57	0	0	0	0
Okuduğu Metni Değerlendirme	3,57	0	0	0	0
Yazım ve Noktalama Kurallarını Uygulama	3,57	0	0	0	0
Fiilimsiler İle İlgili Bilgi ve Kuralları Kavrama ve Uygulama	7,14	2	10	1	5
Söz Varlığını Zenginleştirme	7,14	3	15	2	10
Cümleyle İlgili Bilgi ve Kuralları Kavrama ve Uygulama	19,64	4	20	5	25
Okuduğu Metni Anlama ve Çözümleme	55,36	11	55	12	60
Toplam	100	20	100	20	100

Tablo 5 incelendiğinde en çok kazanıma sahip olan alt öğrenme alanı okuduğu metni anlama ve çözümlenme oluşu görülmektedir. Ortak sınavda toplam kazanımların yaklaşık %11'ine karşılık gelen anlatım bozukluklarını belirleme ve düzeltme, okuduğu metni değerlendirme ile yazım ve noktalama kurallarını uygulama alt öğrenme alanlarına ait madde bulunmamaktadır. En çok madde ise %55,36 kazanım yüzdesine sahip okuduğu metni anlama ve çözümlenme alt öğrenme alanında bulunmaktadır. Alt öğrenme alanlarında bulunan kazanımların sayısı ve yüzdeleriyle birlikte ortak sınavdaki madde sayısı da azalmaktadır. Fakat %7,14 ağırlığa sahip olan fiilimsiler ile ilgili bilgi ve kuralları kavrama ve uygulama, söz varlığını zenginleştirme alt öğrenme alanlarından eşit miktarda madde bulunmamaktadır. Burada maddelerin sadece tek bir kazanımı ölçmediği diğer kazanımlarla da ilişkili olabileceği göz önünde bulundurulmalıdır. Ortak sınavda, 7 alt öğrenme alanından 4'üne ait madde bulunduğundan ortak sınavın alt öğrenme alanlarını %57,1 oranında temsil ettiği söylenebilir.

Mazeret sınavında da ortak sınavla benzer şekilde toplam kazanımların yaklaşık %11'ine karşılık gelen anlatım bozukluklarını belirleme ve düzeltme, okuduğu metni değerlendirme ile yazım ve noktalama kurallarını uygulama alt öğrenme alanlarına ait madde bulunmamaktadır. En çok madde ise okuduğu metni anlama ve çözümlenme alt öğrenme alanından bulunmaktadır. Alt öğrenme alanlarında bulunan kazanımların sayısı ve yüzdelerine göre mazeret sınavındaki madde sayısı da azalmaktadır. Mazeret sınavında toplam 7 alt öğrenme alanından 4'üne ait madde bulunduğundan mazeret sınavının alt öğrenme alanlarını %57,1 oranında temsil ettiği söylenebilir.

Ortak ve mazeret sınavları için alt öğrenme alanlarında kazanım sayıları eşit olmasa da aynı alt öğrenme alanlarından maddeler bulunmakta farklı alt öğrenme alanlarından madde bulunmamaktadır. Bu açıdan bakıldığında benzer kapsamı ölçtükleri söylenebilir.

Ortak ve mazeret sınavlarının KGİ'leri karşılaştırılmış, ayrıca Yurdugül ve Bayrak'ın (2012) da belirttiği gibi Kappa istatistiği madde bazındaki KGİ'ye göre daha tutarlı sonuçlar ürettiğinden Kappa istatistiği de hesaplanmış ve sonuçlar Tablo 6'da verilmiştir.

Tablo 6. Türkçe Testi Ortak ve Mazeret Sınavları KGİ ve Kappa İstatistikleri Karşılaştırılması

Ortak Sınav KGİ	Mazeret Sınavı KGİ	z	Cohen's h	Ortak Sınav Kappa	Mazeret Sınavı Kappa	z	Cohen's h
0,50	0,49	0,05	0,02	0,57	0,58	-0,01	-0,02

* $\alpha = 0,05$ düzeyinde manidardır

Tablo 6 incelendiğinde KGİ ve Kappa istatistiklerinin karşılaştırılması sonucunda KGİ ($z=0,05$, $p>0,05$) ve Kappa ($z= -0,01$, $p>0,05$) istatistikleri manidar düzeyde farklılaşmamıştır. Etki büyüklükleri incelendiğinde ise sırasıyla 0,02 ve -0,02 olduğu görülmektedir. Buna göre etki büyüklüğü yok denecek kadar azdır. Türkçe ortak ve mazeret sınavlarının kapsamalarının KGİ ve Fleiss Kappa katsayısı sonuçlarına göre benzer olduğu söylenebilir.

Türkçe Testlerinin Yapı Geçerlik Düzeylerinin Araştırılması

Türkçe Testi Ortak Sınavının Yapı Geçerlik Düzeyinin Araştırılması

Türkçe testi ortak sınavının verileri üzerinde temel bileşenler faktör çıkarma yöntemi kullanılarak gerçekleştirilen faktör analizi sonucunda, özdeğeri 1'in üzerinde olan bir faktör bulunmuştur. Bu faktörün özdeğeri 10,88, açıklanan varyans oranı %54,40 olduğu gözlenmiştir. Birinci faktörün açıklanan varyans oranının %54,40 olduğu, öz değeri 1'in üzerinde sadece 1 faktör olması Türkçe testi ortak sınavının genel bir faktöre sahip olabileceğini düşündürmektedir. Tablo 7'de maddelerin faktör yükleri verilmiştir.

Tablo 7. Türkçe Testi Ortak Sınav Maddelerinin Faktör Yükleri

Madde Numarası	Faktör Yüğü
1	0,74
2	0,71
3	0,69
4	0,83
5	0,87
6	0,73
7	0,84
8	0,71
9	0,68
10	0,77
11	0,85
12	0,81
13	0,82
14	0,66
15	0,65
16	0,81
17	0,66
18	0,63
19	0,69
20	0,52
Özdeğer	10,88
Açıklanan Varyans (%)	54,40

Tablo 7 incelendiğinde faktör yük değerlerinin 0,52 (20. madde) ile 0,87 (5. madde) arasında değiştiği görülmektedir. 0,40'ın altında faktör yüküne sahip herhangi bir madde bulunmamaktadır. Açıklanan toplam varyans oranı ise %54,40'tır. Bu sonuçlar birlikte değerlendirildiğinde Türkçe testi ortak sınavının tek faktörlü bir yapıda olduğu söylenebilir. Bu sonuçlar literatürdeki diğer çalışmalarla da benzerlik göstermektedir (Anıl ve diğerleri., 2010; Anıl ve Güzeller, 2011; Örs, 2010).

20 maddeden oluşan Türkçe testi ortak sınavı için tek faktörlü bir ölçme modeli tanımlanmış ve bu model doğrulayıcı faktör analizi ile test edilmiştir. Analiz sonucunda göstergeler incelendiğinde faktör yük değerlerinin $\lambda=0,30$ 'un üzerinde olduğu ve 0,39 ile 0,77 arasında değiştiği, özgül varyansların ise tüm değerlerde $\epsilon<0,90$ olduğu görülmüştür.

Tanımlanan tek faktörlü ölçme modeli 0,05 anlamlılık düzeyinde maksimum olabilirlik yöntemi kullanılarak analiz edilmiştir. Model uyumu için elde edilen uyum indeksleri Tablo 8'de sunulmuştur.

Tablo 8. Türkçe Testi Ortak Sınavının DFA Sonucu Elde Edilen Uyum İndeksleri

Uyum İndeksleri	İstatistik Değeri
χ^2	5047,21 (p=0,000, sd=170)
χ^2/sd	29,68
GFI	0,906
AGFI	0,884
CFI	0,977
NFI	0,976
NNFI	0,974
RMSEA	0,077
RMR	0,038
Standardize RMR	0,038

Tablo 8 incelendiğinde $\chi^2/sd=29,68$ olduğu görülmektedir. Bu sonuca göre verinin modele uyum göstermediği söylenebilir. Ancak bu değer örneklem büyüklüğünden etkilendiği için diğer uyum indekslerinin de incelenmesi önerilmektedir (Mueller, 1996). Model uyumu için diğer istatistiklerde incelendiğinde; GFI, AGFI ve CFI istatistikleri sırasıyla 0,906, 0,884 ve 0,977 olduğu görülmektedir. NFI ve NNFI istatistikleri ise 0,976 ve 0,974 olarak bulunmuştur. RMSEA değeri 0,077, RMR ve standardize RMR değerleri ise 0,05'in altındadır. RMR ve SRMR indekslerinin 0,05'in altında olması mükemmel, 0,08'in altında olması iyi uyuma işaret etmektedir (Brown, 2015). Bu değerlere göre Türkçe testi ortak sınavının 20 maddelik tek faktörlü yapısı, DFA sonucunda doğrulandığı gözlenmiştir.

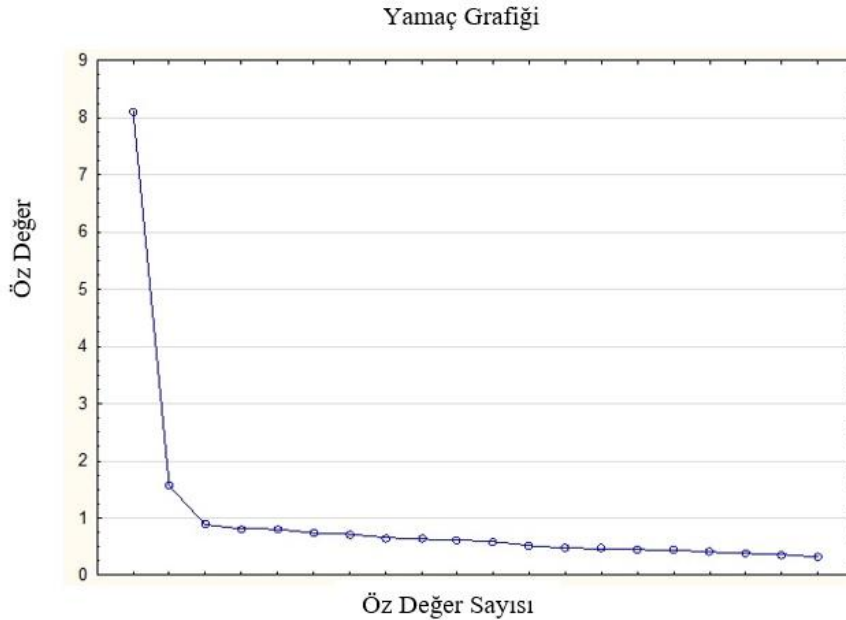
Türkçe Testi Mazeret Sınavının Yapı Geçerlik Düzeyinin Araştırılması

Türkçe testi mazeret sınavından elde edilen verilerin üzerinde temel bileşenler faktör çıkarma yöntemi kullanılarak yürütülen faktör analizi sonucunda, döndürme uygulanmadan elde edilen faktör sayısı, faktör öz değerleri ve açıklanan varyans oranlarını gösteren değerler Tablo 9'da verilmiştir.

Tablo 9. Türkçe Testi Mazeret Sınavının AFA Sonuçları

Faktörler	Faktör Özdeğeri	Açıklanan Varyans (%)	Toplam Varyans (%)
1	8,19	40,97	40,97
2	1,63	8,19	49,16

Tablo 9 incelendiğinde temel bileşenler analizi sonucunda elde edilen, 1 ve üzerinde faktör öz değerlerine sahip 2 faktör olduğu görülmektedir. Birinci faktörle açıklanan varyansın %40,97, ikinci faktörde açıklanan varyansın %8,19 olduğu görülmektedir. Ayrıca özdeğerlere ilişkin oluşturulan yamaç grafiği Şekil 1'de verilmiştir.



Şekil 1. Türkçe Testi Mazeret Sınavına Ait Yamaç Grafiği

Faktör yapısına karar vermek için Şekil 1'de verilen yamaç grafiği incelendiğinde birinci faktörden sonra ani bir düşüş yaşandığı sonrasında ise grafiğin plato yaptığı görülmektedir.

Faktör analizi sonuçları genel olarak değerlendirildiğinde; birinci faktörün açıklanan varyans oranının %40,97 olması, öz değer grafiğinde birinci faktörden sonra ani bir düşüşün yaşanması testin genel bir faktöre sahip olabileceğini düşündürmektedir. Veriler tek faktörle sınırlandırılıp analiz tekrarlandığında Tablo 10'daki sonuç elde edilmiştir.

Tablo 10. Türkçe Testi Mazeret Sınavının Tek Faktörlü AFA Sonuçları

Madde No	Faktör Yüğü
1	0,48
2	0,72
3	0,59
4	0,78
5	0,73
6	0,47
7	0,71
8	0,67
9	0,71
10	0,51
11	0,73
12	0,57
13	0,77
14	0,71
15	0,66
16	0,67
17	0,52
18	0,65
19	0,53
20	0,46
Özdeğer	8,19
Açıklanan Varyans (%)	40,97

Tablo 10 incelendiğinde tek faktörlü yapıda faktör yük değerlerinin 0,40 (6. madde) ile 0,78 (4. madde) arasında değiştiği görülmektedir. 0,30'un altında faktör yüküne sahip herhangi bir madde bulunmamaktadır. Ayrıca açıklanan toplam varyans %40,52 olduğundan Türkçe testi mazeret sınavının tek faktörlü bir yapıda olduğu düşünülmüştür. Bu sonuçlar literatürdeki diğer çalışmalarla da benzerlik göstermektedir (Anıl ve diğerleri., 2010; Anıl ve Güzeller, 2011; Örs, 2010).

Türkçe testi mazeret sınavının faktör yapısına ilişkin tek faktörlü bir ölçme modeli tanımlanmış ve bu model doğrulayıcı faktör analizi ile test edilmiştir. Analiz sonucunda oluşturulan ölçme modeli incelendiğinde, göstergelere ilişkin faktör yük değerlerinin $\lambda=0,30$ 'un üzerinde olduğu ve 0,34 ile 0,76 arasında değiştiği, özgül varyansların ise tüm değerlerde $\varepsilon<0,90$ olduğu görülmüştür.

Tanımlanan tek faktörlü ölçme modeli 0,05 anlamlılık düzeyinde maksimum olabilirlik yöntemi kullanılarak analiz edilmiştir. Model uyumu için elde edilen uyum indeksleri Tablo 11'de sunulmuştur.

Tablo 11. Türkçe Testi Mazeret Sınavının DFA Sonucu Elde Edilen Uyum İndeksleri

Uyum İndeksleri	İstatistik Değeri
χ^2	2087,47 (p=0,000, sd=170)
χ^2/sd	12,27
GFI	0,867
AGFI	0,851
CFI	0,956
NFI	0,951
NNFI	0,950
RMSEA	0,091
RMR	0,058
Standardize RMR	0,058

Tablo 11 incelendiğinde $\chi^2/sd=12,27$ olduğu görülmektedir. Bu sonuca göre verinin modele uyum göstermediği söylenebilir. Ancak bu değer örneklem büyüklüğünden etkilendiği için diğer uyum indekslerinin de incelenmesi önerilmektedir (Mueller, 1996). Model uyumu için diğer istatistiklerde incelendiğinde; GFI, AGFI ve CFI istatistikleri sırasıyla 0,867, 0,851 ve 0,956 olduğu görülmektedir. GFI 0,90 ve üzerinde olduğu durumda kabul edilebilir uyum olduğu düşünülmektedir. Fakat GFI değeri örneklem büyüklüğünden etkilenen bir istatistik olduğu için 0,867 olarak bulunmuştur. Buna rağmen diğer tüm değerler uyumlu olduğunu gösterdiği için diğer istatistiklerde incelenmiştir. NFI ve NNFI istatistikleri ise 0,951 ve 0,950 olarak bulunmuştur. RMSEA değeri 0,091, RMR ve standardize RMR değerleri ise 0,058 olarak bulunmuştur. Bu değerlere göre Türkçe testi mazeret sınavının 20 maddelik tek faktörlü yapısı DFA sonucunda doğrulanmıştır.

Türkçe testi ortak ve mazeret sınavlarının yapı geçerlik düzeylerini incelemek amacıyla AFA ve DFA gerçekleştirilmiş ve her iki sınav için de tek faktörlü bir yapı tanımlanmıştır. Bu yapılarda tüm maddeler faktörlerine yüklenmiş ve faktör yükü 0,30'un altında herhangi bir madde olmamıştır. Bu durumda yapı geçerliği açısından Türkçe testi ortak ve mazeret sınavlarının birbirine yakın düzeyde olduğunu söylemek mümkündür.

Matematik Testlerinin Test İstatistiklerinin Karşılaştırılması

Matematik testi ortak ve mazeret sınavları için test istatistiklerinin farklılık gösterip göstermediği sınımlanmış ve sonuçlar Tablo 12'de belirtilmiştir.

Tablo 12. Matematik Testi Ortak ve Mazeret Sınavlarının Test İstatistiklerinin Karşılaştırılması

Test Adı	\bar{X}	t	Cohen's d	\bar{p}	z	Cohen's h	S_x	F	Cohen's d
Matematik-Ortak	9,48			0,47			19,46		
Matematik-Mazeret	5,89	42,68*	1,28		15,40*	0,35		401,41*	0,46
				0,30			11,63		

*p<0.05

Matematik ortak ve mazeret sınavlarının aritmetik ortalamaları arasında fark, ortak sınav lehine manidar bulunmuştur ($t(4431,75)=42,68$, $p<0,05$). Etki büyüklüğü 1,28 olarak bulunduğu görülmektedir. Buna göre ortalamalar arasındaki farkın çok yüksek düzeyde etki büyüklüğüne sahip olduğu, bu farkın örneklem büyüklüğünden kaynaklanmadığı söylenebilir. Aynı zamanda sınavların ortalama güçlükleri incelendiğinde matematik testi ortak sınavının ortalama güçlüğü 0,47, mazeret sınavının ortalama güçlüğü ise 0,30 olduğu gözlenmiştir. Matematik testinin ortak sınavının ortalaması, mazeret sınavına göre daha yüksek olduğu ve ortalama güçlüklerinin manidar düzey farklılaşması ($z(11806)=15,40$, $p<0,05$), mazeret sınavının öğrenciler için bir dezavantaj oluşturduğunu düşündürmektedir. Ortalama güçlük karşılaştırmasına yönelik olarak elde edilen etki büyüklüğü 0,35 olarak bulunmuştur. Buna göre orta düzeye yakın bir etki büyüklüğüne sahip olduğu söylenebilir.

Ortak ve mazeret sınavlarının varyansları arasında ortak sınav lehine manidar düzeyde fark bulunmuştur ($F(1,11807)=401,41$, $p<0,05$). Bu durum ortak sınava katılan grubun mazeret sınavına katılan gruba göre daha homojen bir grup olduğu anlamına gelmektedir. Varyansların karşılaştırmasına yönelik olarak elde edilen etki büyüklüğü değeri 0,46 olarak bulunmuştur. Buna göre orta düzeyde bir etki büyüklüğüne sahip olduğu söylenebilir.

Tablo 13. Matematik Testi Ortak ve Mazeret Sınavlarının Test İstatistiklerinin Karşılaştırılması

Test Adı	Güvenirlilik	z	Cohen's q	r_{jx} Ortancası	Mann-Whitney U	r
Matematik-Ortak	0,88			0,57		
Matematik-Mazeret	0,81	11,94*	0,25	0,32	62,50*	0,58

*p<0.05

Tablo 13’de görüldüğü üzere sınavların McDonald ω değerleri karşılaştırıldığında ortak sınav lehine manidar fark bulunmuştur ($z(11807)=11,94, p<0,05$). Etki büyüklüğü incelendiğinde ise Cohen’s q istatistiği 0,25 olduğu görülmektedir. Buna göre orta düzeyde bir etki büyüklüğüne sahip olduğu, farkın gözlenmesinde örneklem büyüklüğünün etkisinin az olduğu söylenebilir.

Testlerin madde ayırıcılık indeksi ortancaları (medyan) kıyaslandığında ortak sınav lehine manidar fark bulunmuştur ($U=62,50, z=-3,72, p<0,05, r=0,58$). Etki büyüklüğü de incelendiğinde yüksek düzeyde etki büyüklüğüne sahip olduğu görülmektedir. Bu durum testlerin, testte ölçülen kazanımlara sahip olan ve olmayan bireyleri birbirinden ayırma düzeylerinin farklı olduğu anlamına gelmektedir.

Tüm bu karşılaştırmalar incelendiğinde matematik ortak ve mazeret sınavlarının birbirinin alternatifi olarak kullanılmamasının daha uygun olacağı söylenebilir.

Matematik Testlerinin Madde İstatistiklerinin Karşılaştırılması

Matematik ortak ve mazeret sınavlarındaki 13 maddenin (Tablo 14’te 1-13. sıradaki maddeler) aynı kazanımı ölçtüğü, kalan 7 maddenin ise farklı kazanımları ölçmekle beraber bu 7 maddenin 5’inin (Tablo 14’te 14-18. sıradaki maddeler) aynı alt öğrenme alanında bulunduğu ve konu olarak eşleştiği gözlenmiştir. Tabloda 14’te 19. sırada yer alan ortak sınavdaki madde üçgenler alt öğrenme alanındayken mazeret sınavındaki madde ise üçgenlerde ölçme alt öğrenme alanındadır. Birbirine çok yakın alt öğrenme alanlarından olduğu için bu maddeler birbiriyle eşleştirilmiştir. Kalan 1 madde (Tablo 14’te 20. sıradaki maddeler) ise farklı öğrenme alanlarındadır. Buna göre maddeler eşleştirilmiş ve eşleşen maddelerin madde istatistikleri arasında fark olup olmadığı araştırılmıştır. Matematik ortak ve mazeret sınavında bulunan maddelerin madde güçlükleri ve madde ayırıcılık indeksleri açısından karşılaştırılması Tablo 14’te verilmiştir.

Tablo 14. Matematik Testi Ortak ve Mazeret Sınavlarının Madde Güçlük ve Ayırıcılık İndekslerinin Karşılaştırılması

S.No	Ortak Sınav Madde No	Mazeret Sınavı Soru No	p_j^*	p_j^{**}	z	Cohen’s h	r_{jx}^*	r_{jx}^{**}	z	Cohen’s q
1	6	11	0,65	0,37	24,52*	0,57	0,62	0,46	9,83*	0,23
2	11	9	0,61	0,29	28,49*	0,66	0,57	0,39	10,18*	0,24
3	13	2	0,52	0,35	14,84*	0,34	0,68	0,28	23,38*	0,54
4	7	12	0,39	0,43	-3,20*	-0,08	0,42	0,30	5,97*	0,14
5	10	8	0,49	0,24	21,92*	0,53	0,56	0,28	14,90*	0,35
6	3	5	0,83	0,36	46,33*	1,00	0,36	0,51	-8,02*	-0,19
7	17	16	0,40	0,38	2,09	0,04	0,62	0,31	17,46*	0,40
8	8	14	0,50	0,25	21,64*	0,52	0,75	0,34	26,72*	0,62
9	18	18	0,41	0,29	11,29*	0,25	0,71	0,31	24,46*	0,57
10	19	19	0,32	0,32	0,22	0,00	0,63	0,42	12,68*	0,29
11	20	20	0,33	0,22	10,41*	0,25	0,20	0,17	1,34	0,03
12	4	6	0,44	0,28	14,44*	0,34	0,47	0,29	9,13*	0,21
13	5	15	0,71	0,33	34,16*	0,78	0,55	0,33	11,90*	0,28
14	1	3	0,45	0,23	19,01*	0,47	0,60	0,31	16,09*	0,37
15	9	10	0,37	0,15	20,11*	0,51	0,60	0,29	17,04*	0,39
16	12	13	0,55	0,37	15,67*	0,36	0,76	0,37	26,24*	0,61
17	15	7	0,33	0,27	5,53*	0,13	0,30	0,34	-1,92	-0,04
18	16	17	0,36	0,24	10,64*	0,26	0,46	0,23	11,36*	0,26
19	2	4	0,31	0,31	0,09	0,00	0,50	0,37	6,95*	0,16
20	14	1	0,51	0,24	23,18*	0,57	0,26	0,32	-2,83*	-0,07

* $p<0,05$, * Ortak Sınav, ** Mazeret Sınavı

Tablo 14 incelendiğinde madde güçlük indekslerinin ortak sınav için 0,31 ile 0,83; mazeret sınavı için 0,15 ile 0,43 arasında değiştiği görülmektedir. Ortak sınavda çok zor madde bulunmamaktadır. Mazeret sınavında ise kolay ve çok kolay madde yoktur. Madde güçlüğü bakımından ortak sınavda

zor 8, orta güçlükte 8, kolay 3 ve çok kolay 1 madde bulunurken, mazeret sınavında çok zor 1, zor 18, ve orta güçlükte 1 madde bulunmaktadır.

Mazeret sınavının gruba daha zor geldiği ve soruların güçlük düzeyi dağılımının ortak sınavdaki gibi olmadığı söylenebilir. Aynı ya da benzer kazanımı ölçen maddelerin madde güçlük düzeyleri bakımından 7., 10. ve 19. sıradaki maddeler için istatistiksel olarak anlamlı düzeyde farklılık olmadığı diğer maddelerin ise istatistiksel olarak anlamlı düzeyde farklılaştığı görülmektedir. Bu durum, testi alan bireylerin farklı güçlükteki maddelere yanıt verdiğini göstermektedir. Birbirinin yerine kullanılan testlerin madde güçlüklerinin bu kadar farklılaşması testin adalet yönünü zedeleyeceği düşünülmektedir. Madde güçlük indeksi karşılaştırmalarının etki büyüklükleri incelendiğinde ise -0,08 ile 1,00 arasında değiştiği görülmektedir. Buna göre; 4 karşılaştırmanın etki büyüklüğü çok küçük, 7 karşılaştırmanın etki büyüklüğü küçük, 7 karşılaştırmanın etki büyüklüğü orta ve 2 karşılaştırmanın ise etki büyüklüğü çok yüksektir. Karşılaştırmaların yaklaşık yarısının etki büyüklüklerinin orta ve üzerinde olduğu söylenebilir.

Tablo 14'te ortak sınavın madde ayırıcılık indekslerinin 0,20 ile 0,76 arasında değiştiği görülmektedir. 20. ve 14. maddelere ait madde ayırıcılık indeksleri sırasıyla 0,20 ve 0,26 olarak bulunmuştur. Diğer 18 madde ise 0,30 ve üzerinde madde ayırıcılık indeksine sahiptir. Mazeret sınavının madde ayırıcılık indeksleri incelendiğinde 0,17 ile 0,51 arasında değiştiği görülmektedir. Mazeret sınavında madde ayırıcılık indeksi 0,20'nin altında olan 1 madde (20. madde), 0,20-0,30 arasında olan 6 madde (2., 6., 8., 10., 12. ve 17. maddeler) bulunmaktadır.

Aynı ya da benzer kazanımı ölçen maddelerin madde ayırıcılık indeksleri karşılaştırıldığında, 2 madde haricindeki tüm maddelerde $\alpha=0,05$ düzeyinde manidar bulunmuştur. Ortak sınavın madde ayırıcılık indeksleri daha yüksek olduğu için ortak sınavın mazeret sınavında göre bu psikometrik özellik açısından daha uygun bir test olduğu söylenebilir. Etki büyüklükleri incelendiğinde etki büyüklüğünün -0,19 ile 0,62 arasında değiştiği Tablo 14'te görülmektedir. Buna göre karşılaştırmalarının büyük çoğunluğunun etki büyüklüğü orta ve altında iken 4 adet karşılaştırmanın yüksek düzeyde etki büyüklüğüne sahip olduğu gözlenmiştir.

Matematik Testlerinin Kapsam Geçerlik Oranı İndekslerinin Karşılaştırılması

Tablo 15'te alt öğrenme alanlarına göre öğretim programında bulunan kazanım yüzdesi, ortak sınav ve mazeret sınavında bu alt öğrenme alanından kaç madde bulunduğu ve bu maddelerin yüzdesi verilmiştir.

Tablo 15. Matematik Ortak ve Mazeret Sınavlarında Bulunan Maddelerin Alt Öğrenme Alanına Göre Dağılımı

Alt Öğrenme Alanı	Öğretim Programındaki Kazanım Yüzdesi	Ortak Sınavda Bulunan Madde Sayısı	Ortak Sınavda Bulunan Madde Yüzdesi	Mazeret Sınavında Bulunan Madde Sayısı	Mazeret Sınavında Bulunan Madde Yüzdesi
Örüntü ve Süslemeler	1,92	0	0	0	0
Tablo ve Grafikler	1,92	0	0	0	0
Olasılık Çeşitleri	1,92	0	0	0	0
Gerçek sayılar	1,92	0	0	1	5
Merkezi eğilim ve yayılma ölçüleri	3,85	0	0	0	0
Örüntüler ve İlişkiler	3,85	1	5	1	5
Dönüşüm Geometrisi	3,85	1	5	1	5
Olası Durumları Belirleme	3,85	1	5	1	5
Olay Çeşitleri	3,85	1	5	1	5
Üslü Sayılar	5,77	1	5	1	5
Köklü Sayılar	7,69	1	5	1	5
Üçgenlerde Ölçme	7,69	2	10	3	15
Cebirsel İfadeler	9,62	2	10	2	10
Geometrik Cisimler	11,54	2	10	2	10
Denklemler	13,46	4	20	4	20
Üçgenler	17,31	4	20	2	10
Toplam	100	20	100	20	100

Tablo 15 incelendiğinde en çok kazanıma ve en çok ders saatine sahip olan alt öğrenme alanı üçgenlerdir. Toplam kazanımın %13,5'ine karşılık gelen örüntü ve süslemeler, tablo ve grafikler, olasılık çeşitleri, gerçek sayılar ve merkezi eğilim ve yayılma ölçüleri alt öğrenme alanlarına ait ortak sınavda madde bulunmamaktadır. En çok madde ise %17,31 kazanım yüzdesine sahip üçgenler alt öğrenme alanında bulunmaktadır. Denklemler alt öğrenme alanında toplam kazanımın %9,62'si bulunmasına rağmen ortak sınavda bu öğrenme alanına ait 4 madde bulunmaktadır. Geometrik cisimler alt öğrenme alanında ise 2 madde vardır. Oysaki üçgenler alt öğrenme alanından sonra en çok kazanım bulunan öğrenme alanı geometrik cisimler alt öğrenme alanıdır. Burada maddelerin sadece tek bir kazanımı ölçmediği diğer kazanımlarla da ilişkili olabileceği göz önünde bulundurulmalıdır. Ortak sınavda toplam 16 alt öğrenme alanından 11'ine ait madde bulunduğu ortak sınavın alt öğrenme alanlarını %68,75 oranında temsil ettiği söylenebilir.

Toplam kazanımın %9,61'ine karşılık gelen örüntü ve süslemeler, tablo ve grafikler, olasılık çeşitleri ve merkezi eğilim ve yayılma ölçüleri alt öğrenme alanlarından mazeret sınavında madde bulunmamaktadır. Mazeret sınavında ortak sınava benzer şekilde denklemler alt öğrenme alanından 4 madde bulunmaktadır. Fakat ortak sınavda üçgenler alt öğrenme alanından 4 madde varken mazeret sınavında 2 madde bulunmaktadır. Mazeret sınavında toplam 16 alt öğrenme alanından 12'sine ait madde bulunduğu mazeret sınavının alt öğrenme alanlarını %75 oranında temsil ettiği söylenebilir.

Mazeret sınavında, gerçek sayılar alt öğrenme alanından 1 madde bulunmasına rağmen ortak sınavda bu alt öğrenme alanından madde bulunmamaktadır. Üçgenlerde ölçme alt öğrenme alanından ortak sınavda 2, mazeret sınavında ise 3 madde bulunmaktadır. Üçgenler alt öğrenme alanından ise ortak sınavda 4, mazeret sınavında ise 2 madde vardır. Diğer alt öğrenme alanlarından eşit sayıda madde bulunmaktadır.

Ortak ve mazeret sınavlarının KGİ'leri ve Kappa istatistikleri karşılaştırılmış ve sonuçlar Tablo 16'da verilmiştir.

Tablo 16. Matematik Ortak ve Mazeret Sınavları KGİ ve Kappa İstatistikleri Karşılaştırılması

Ortak Sınav KGİ	Mazeret Sınavı KGİ	z	Cohen's h	Ortak Sınav Kappa	Mazeret Sınavı Kappa	z	Cohen's h
0,46	0,54	0,42	0,16	0,69	0,75	-0,29	-0,13

* $\alpha = 0,05$ düzeyinde manidardır.

Matematik ortak ve mazeret sınavlarına ilişkin hesaplanan KGİ ($z=0,42$, $p>0,05$) ve Kappa istatistikleri ($z= -0,29$, $p>0,05$) arasında anlamlı farklılıklar gözlenmemiştir. Etki büyüklükleri de incelendiğinde çok küçük oldukları görülmektedir. Matematik ortak ve matematik mazeret sınavlarının KGİ ve Fleiss Kappa katsayısı açısından benzer olduğu söylenebilir. Yılmaz-Koğar ve Aygün, (2015)'de 2013-2014 eğitim öğretim yılında uygulanan TEOG sınavlarından matematik alt testinin ortak sınavını iki dönem içinde incelemiş ve kapsam geçerliği için gerekli ölçütü sağladığını belirtmiştir.

Matematik ortak ve mazeret sınavlarının benzer kapsamı ölçmesine rağmen madde güçlük ve ayrıricılık indeksleri gibi madde istatistikleri bakımından farklılaştığı görülmektedir. Benzer kapsamı ölçen testlerin, madde istatistikleri açısından farklılaşması testin uygulandığı grupların birbirinden farklı özelliklerde olmasından kaynaklanabileceği düşünülmektedir. Madde ve test istatistikleri gruba bağımlı istatistikler olduğu için aynı kapsamı ölçen maddelerde de farklılaşabilmektedir.

Matematik Testlerinin Yapı Geçerlik Düzeylerinin Araştırılması

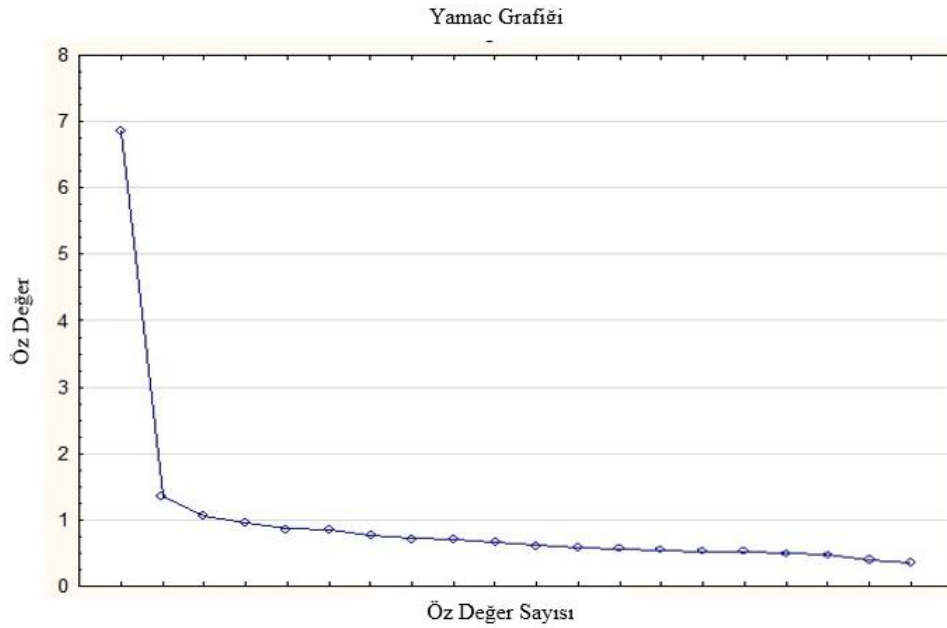
Matematik Testi Ortak Sınavının Yapı Geçerlik Düzeyinin Araştırılması

Matematik testi ortak sınavının verileri üzerinde temel bileşenler faktör çıkarma yöntemi kullanılarak gerçekleştirilen faktör analizi sonucunda, elde edilen faktör sayısı, faktör öz değerleri ve açıklanan varyans oranlarını gösteren değerler Tablo 17'de verilmiştir.

Tablo 17. Matematik Ortak Sınavının AFA Sonuçları

Faktörler	Faktör Özdeğerleri	Açıklanan Varyans (%)	Toplam Varyans (%)
1	6,95	34,75	34,75
2	1,35	6,79	41,54
3	1,10	5,50	47,04

Tablo 17 incelendiğinde temel bileşenler analizi sonucunda elde edilen, 1 ve üzerinde öz değere sahip 3 faktör olduğu görülmektedir. Birinci faktörle açıklanan varyansın %34,75 olduğu ve üç faktörün tamamının %47,04 oranında toplam varyansı açıkladığı görülmektedir. Faktörlerin öz değerleri birinci faktör için 6,95, ikinci faktör için 1,35 ve üçüncü faktör için 1,10 olarak bulunmuştur. Ayrıca özdeğerlere ilişkin oluşturulan yamaç grafiği Şekil 2’de verilmiştir.



Şekil 2. Matematik Testi Ortak Sınavına Ait Yamaç Grafiği

Faktör yapısına karar vermeden önce Şekil 2’de verilen yamaç grafiği de incelenmiş birinci faktörden sonra ani ve ivmeli bir düşüş olduğu fakat diğer faktörlerden sonra grafiğin yatay olarak devam ettiği ve ani bir düşüş olmadığı gözlenmiştir.

Faktör analizi sonuçları uzman görüşlerine göre de değerlendirilmiş ve matematik testinin cebir ve geometri olarak iki yapıda olabileceği sonucuna varılmıştır. Matematik ortak sınavı için faktör analizi iki faktörle sınırlandırılarak faktör analizi tekrarlanmış ve eğik (oblique) döndürme yöntemlerinden promax yöntemi kullanılarak faktörler döndürülmüştür.

Uzman görüşleri doğrultusunda cebir ve geometri olarak iki faktörlü bir yapıda olduğu düşünülen Matematik ortak sınavındaki maddelerin iki faktöre göre dağılımı Tablo 18’de verilmiştir.

Tablo 18. Matematik Ortak Sınavı 2 Faktörlü, Döndürülmüş AFA için Maddelerin Faktörlere Göre Dağılımı

Madde No	Cebir	Geometri
1	0,27	0,43*
2	0,09	0,61**
3	0,74**	-0,23
4	0,40*	0,11
5	0,82*	-0,23
6	0,82*	-0,17
7	-0,14	0,71**
8	0,70**	0,09
9	0,30	0,45*
10	0,30	0,31**
11	0,70**	-0,15
12	0,74*	0,03
13	0,61*	0,10
14	0,25*	-0,08
15	-0,24	0,67*
16	0,05	0,56**
17	0,41**	0,35
18	0,55**	0,25
19	0,45**	0,36
20	-0,33	0,60**
Özdeğer	6,75	1,36
Açıklanan Varyans (%)	34,75	6,79

*Uzman görüşüne göre diğer faktörde yer alan fakat analiz sonucunda mevcut faktöre yük veren madde.

** Faktör yükü ve uzman görüşüne göre mevcut faktörde yer alan madde.

Tablo 18 incelendiğinde ortak sınavda bulunan 20 maddeden 12'si cebir faktöründe, 8'i geometri faktöründe yer almaktadır. Bu şekilde oluşturulan yapı toplam varyansın %41,54'ünü açıklamaktadır. Matematik ortak sınavı uzman görüşleri doğrultusunda incelendiğinde cebir konu alanıyla ilgili olan 9 madde, (1., 3., 8., 9., 11., 15., 17., 18. ve 19. maddeler) geometri konu alanıyla ilgili 11 madde (2., 4., 5., 6., 7., 10., 12., 13., 14., 16. ve 20. madde) olduğu görüşüne ulaşılmıştır. Bu duruma göre FA sonuçları incelendiğinde 1., 9. ve 15. maddenin cebir konu alanıyla ilgili olmasına rağmen geometri alanını ölçüyor gibi görünmektedir. Benzer şekilde 4., 5., 6., 12., 13., 14. maddeler geometri konu alanıyla ilgili olmasına rağmen cebir konu alanını ölçüyor gibi görünmektedir. Bu durum testin yapı geçerliğini zedeleyici bir etki yapmaktadır. Literatürdeki diğer çalışmalar incelendiğinde genellikle bu tür sınavlarda tek faktörlü yapılar elde edilmiştir. Ancak bu çalışmalarda faktörlere yük vermeyen maddeler analiz dışı bırakılmıştır (Anıl ve diğerleri., 2010; Anıl ve Güzeller, 2011; Örs, 2010). Fakat burada uzman görüşleri ve faktör yükleri dikkate alınarak 2 faktörlü bir yapı tanımlanmıştır. Tanımlanan bu yapı DFA ile test edilmiştir. Analiz sonucu tanımlanan bu ölçme modeli incelendiğinde göstergelere ilişkin faktör yük değerlerinin 14. ve 20. maddeler hariç $\lambda=0,30$ 'un üzerinde olduğu ve 0,31 ile 0,74 arasında değiştiği, özgül varyanslara bakıldığında genellikle $\varepsilon<0,90$ olduğu ancak 14. ve 20. maddeler için 0,90'ın üzerinde olduğu görülmüştür. Tanımlanan 2 faktörlü ölçme modeli 0,05 anlamlılık düzeyinde maksimum olabilirlik yöntemi kullanılarak analiz edilmiştir. Model uyumu için elde edilen uyum indeksleri Tablo 19'da sunulmuştur.

Tablo 19. Matematik Ortak Sınavının DFA Sonucu Elde Edilen Uyum İndeksleri

Uyum İndeksleri	İstatistik Değeri
χ^2	2772,38 (p=0,000, s _d =169)
χ^2/sd	16,404
GFI	0,944
AGFI	0,931
CFI	0,970
NFI	0,968
NNFI	0,966
RMSEA	0,057
RMR	0,038
Standardize RMR	0,038

Tablo 19 incelendiğinde $\chi^2/sd=16,404$ olarak bulunmuştur. Bu sonuca göre verinin modele uyum göstermediği söylenebilir. Ancak bu değer örneklem büyüklüğünden etkilendiği için diğer uyum indekslerinin de incelenmesi önerilmektedir (Mueller, 1996). Model uyumu için diğer istatistiklerde incelendiğinde; GFI, AGFI ve CFI istatistiklerinin sırasıyla 0,944, 0,931, 0,970 olduğu görülmektedir. NFI ve NNFI istatistikleri ise 0,968 ve 0,966 olarak bulunmuştur. RMSEA, RMR ve standardize RMR değerleri 0,057 ve 0,038'dir. Bu değerlere göre matematik ortak sınavının 20 maddelik iki faktörlü yapısı uyumlu görülmüştür. Fakat 2 maddenin ilişki katsayısının 0,30'un altında olması maddelerin faktörlerle ilişkilerinin çok düşük düzeyde olduğu anlamına gelmektedir. Bu durum da yapı geçerlik düzeyini düşürücü bir etki yapmaktadır.

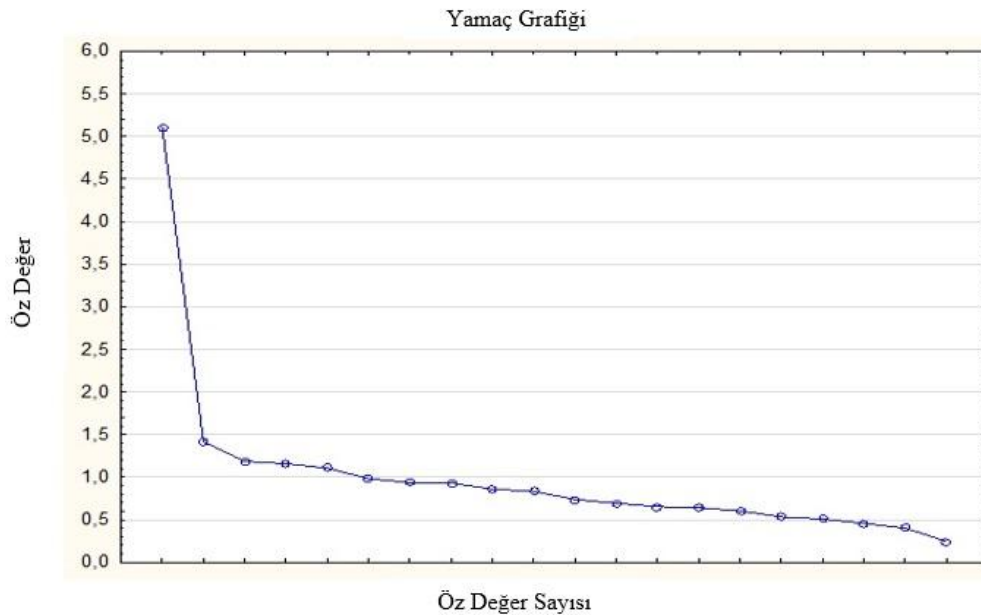
Matematik Mazeret Sınavının Yapı Geçerlik Düzeyinin Araştırılması

Matematik mazeret sınavının faktör yapısını belirleyebilmek amacıyla temel bileşenler faktör çıkarma yöntemi kullanılarak AFA gerçekleştirilmiş ve döndürme uygulanmadan elde edilen faktör sayısı, faktör öz değerleri ve açıklanan varyans oranlarını gösteren değerler Tablo 20'de verilmiştir.

Tablo 20. Matematik Mazeret Sınavının AFA Sonuçları

Faktörler	Faktör Özdeğerleri	Açıklanan Varyans (%)	Toplam Varyans (%)
1	5,11	25,55	25,55
2	1,42	7,08	32,63
3	1,19	5,93	38,55
4	1,16	5,79	44,34
5	1,11	5,53	49,87

Tablo 20 incelendiğinde temel bileşenler analizi sonucunda elde edilen, 1 ve üzerinde faktör öz değerlerine sahip 5 faktör olduğu görülmektedir. Birinci faktörle açıklanan varyansın %25,55 olduğu ve beş faktörün tamamının %49,87 oranında toplam varyansı açıkladığı görülmektedir. Faktörlerin öz değerleri birinci faktör için 5,11, ikinci faktör için 1,42, üçüncü faktör için 1,19, dördüncü faktör için 1,16, beşinci faktör için 1,11 olarak bulunmuştur. Özdeğerlere ilişkin oluşturulan yamaç grafiği Şekil 3'te verilmiştir.



Şekil 3. Matematik Testi Mazeret Sınavına Ait Yamaç Grafiği

Şekil 3 incelendiğinde birinci faktörden sonra ani ve ivmeli bir düşüş olduğu fakat diğer faktörlerden sonra grafiğin yatay olarak devam ettiği ve ani bir düşüş olmadığı görülmektedir.

Faktör analizi sonuçları uzman görüşlerine göre de değerlendirilmiş ve matematik testinin cebir ve geometri gibi iki faktörlü bir yapıda olabileceği sonucuna varılmıştır. Matematik mazeret sınavı için faktör analizi iki faktörle sınırlandırılarak faktör analizi tekrarlanmış ve eğik (oblique) döndürme yöntemlerinden promax döndürme yöntemi kullanılarak faktörler döndürülmüştür.

Uzman görüşleri doğrultusunda cebir ve geometri olarak iki faktörlü bir yapıda olduğu düşünülen Matematik mazeret sınavındaki maddelerin iki faktöre göre dağılımı Tablo 21’de verilmiştir

Tablo 21. Matematik Mazeret Sınavının 2 Faktörlü ve Döndürülmüş AFA Sonuçları

Madde No	Cebir	Geometri
1	0,58	0,15
2	-0,06	0,34
3	0,55	0,01
4	0,38	0,17
5	0,28	0,45
6	0,07	0,47
7	0,63	-0,08
8	0,44	-0,03
9	0,60	0,15
10	0,70	0,17
11	0,23	0,48
12	-0,12	0,56
13	-0,09	0,60
14	0,43	0,23
15	-0,13	0,62
16	0,45	-0,07
17	0,68	-0,45
18	0,19	0,24
19	0,49	0,13
20	0,47	-0,35
Özdeğer	5,11	1,42
Açıklanan Varyans (%)	25,55	7,08

*Uzman görüşüne göre diğer faktörde yer alan fakat analiz sonucunda mevcut faktöre yük veren madde.

** Faktör yükü ve uzman görüşüne göre mevcut faktörde yer alan madde.

Tablo 21 incelendiğinde 7 maddenin geometri, 13 maddenin ise cebir faktöründe yer aldığı görülmektedir. Matematik mazeret sınavı uzman görüşleri doğrultusunda incelendiğinde cebir konu alanıyla ilgili olan 12 madde (1., 3., 5., 7., 9., 10., 14., 16., 17., 18., 19. ve 20. maddeler) geometri konu alanıyla ilgili 8 madde (2., 4., 6., 8., 11., 12., 13., ve 15. madde) olduğu sonucuna ulaşılmıştır. Bu duruma göre FA sonuçları incelendiğinde 4. ve 8. maddeler geometri konu alanıyla ilgili olmasına rağmen cebir konu alanıyla ilişki ölçüm yapıyormuş gibi görünmektedir. Benzer şekilde cebir konu alanına ait olan 5. madde ise geometri konusu gibi algılandığı görülmektedir. Bu durum mazeret sınavının yapı geçerliğini zedeleyici bir etki yapmaktadır.

Literatürdeki diğer çalışmalar incelendiğinde genellikle tek faktörlü yapılar elde edilmiştir. (Anıl ve diğerleri., 2010; Anıl ve Güzeller, 2011; Örs, 2010). Fakat bu çalışmada madde yükleri, yamaç grafiği, açıklanan varyans oranları ve uzman görüşleri dikkate alınarak 2 faktörlü bir yapı tanımlanmıştır. Tanımlanan bu yapı DFA ile test edilmiş, faktör yük değerleri ile maddeler arasındaki ilişki katsayılarının 0,04 ile 0,73 arasında olduğu görülmüştür. 2. ,6. 12., 16. ve 20. maddelerin faktörleriyle ilişki katsayıları 0,30’un altında olduğu, özgül varyansların ise genellikle 0,90 altında olduğu fakat 2. ,6. 12., 16. 17. ve 20. maddelerin 0,90’ın üzerinde olduğu gözlenmiştir.

Tanımlanan 2 faktörlü ölçme modeli 0,05 anlamlılık düzeyinde maksimum olabilirlik yöntemi kullanılarak analiz edilmiştir. Model uyumu için elde edilen uyum indeksleri Tablo 22’de sunulmuştur.

Tablo 22. Matematik Mazeret Sınavının DFA Sonucu Elde Edilen Uyum İndeksleri

Uyum İndeksleri	İstatistik Değeri
χ^2	892,01 (p=0,000, s _d =169)
χ^2/sd	5,278
GFI	0,929
AGFI	0,912
CFI	0,905
NFI	0,890
NNFI	0,893
RMSEA	0,060
RMR	0,048
Standardize RMR	0,048

Tablo 22 incelendiğinde $\chi^2/sd=5,278$ olduğu görülmektedir. Bu sonuca göre verinin modele uyum göstermediği söylenebilir. Ancak bu değer örneklem büyüklüğünden etkilendiği için diğer uyum indekslerinin de incelenmesi önerilmektedir (Mueller, 1996). Bunun yanında model uyumu için diğer istatistiklerde incelendiğinde; GFI, AGFI ve CFI istatistiklerinin sırasıyla 0,929, 0,912 ve 0,905 olduğu gözlenmiştir. NFI ve NNFI istatistikleri ise 0,890 ve 0,893 olarak bulunmuştur. RMSEA 0,060 olup, RMR ve standardize RMR değerleri ise 0,048 olarak bulunmuştur. Uyum katsayıları genel olarak incelendiğinde kurulan modelin uyumlu olduğu söylenebilir.

Matematik ortak ve mazeret sınavlarının yapı geçerlik düzeylerini ortaya koymak için AFA ve DFA sonucunda elde edilen yapılar, testlerin birbirinin alternatif olarak kullanılmamasının daha uygun olacağını göstermektedir. Testlerde uzman görüşleri ile cebir ve geometri olmak üzere iki faktörlü bir yapı oluşturulmuştur. Fakat her iki sınav içinde geometri ve cebirle ilişkili olup faktör analizi sonucu kendi faktörlerine yüklenmeyen maddeler olduğu gözlenmiştir. Geometri öğrenme alanıyla ilişki olan maddenin cebir öğrenme alanında bir maddeymiş gibi yüklenmesi, maddenin farklı bir yapıyı ölçtüğüne işaret etmektedir. Bu durum yapı geçerlik düzeylerinin düşmesine neden olmaktadır.

SONUÇLAR ve TARTIŞMA

Bu araştırma sonucunda aşağıdaki sonuçlara ulaşılmıştır.

1. 2013-2014 eğitim öğretim yılında uygulanan TEOG Türkçe ve matematik testleri ortak ve mazeret sınavlarının madde ve test istatistikleri karşılaştırılmış ve bu istatistiklerin manidar düzeyde farklı olduğu sonucuna ulaşılmıştır ($\alpha=0,05$). Literatür incelendiğinde Özkan (2015) tarafından 2013-2014 eğitim öğretim yılı I. döneminde gerçekleştirilen matematik ortak ve mazeret sınavlarının test eşitleme çalışması yapılmış ve her iki sınavın güvenilirlik ve ortalama güçlüklerinin eşit olduğu, ortalamaları arasında manidar farklılık bulunmasına karşın, testlerin varyanslarının eşit olduğu belirtilmiştir. Bu çalışmada ise ortak ve mazeret sınavlarının test ve madde istatistiklerinin farklılaştığı sonucuna ulaşılmıştır. Testlerin uygulanan dönemlerinin farklı olmasının böyle bir sonucu doğurmuş olabileceği düşünülmektedir. Bu araştırma sonucuna göre test ve madde istatistiklerinin farklılaşması durumunun sınavların fırsat eşitliği yönünü zedelediği söylenebilir.

2. 2013-2014 eğitim öğretim yılında uygulanan TEOG Türkçe ve matematik testleri ortak ve mazeret sınavlarının KGİ ve Kappa istatistikleri arasında anlamlı fark bulunmamıştır. Bu nedenle ortak ve mazeret sınavlarının uzman görüşlerine dayalı olarak değerlendirilen kapsam geçerliklerinin benzer düzeyde olduğu sonucuna ulaşılmıştır. Genel olarak bu sonuç TEOG alt testleri için literatürdeki diğer bulgularla da benzerlik göstermektedir (Karadeniz ve diğerleri., 2015; Kardeş-Birinci, 2014;

Yılmaz-Koğar ve Aygün, 2015). Benzer kapsamı ölçtüğü gözlenen testlerin, test ve madde istatistiklerinin farklılaştığı da göz önüne alındığında testi alan grupların farklı olmasının bu duruma neden olabileceği düşünülmüştür. Ancak etki büyüklükleri incelendiğinde etki büyüklüğü yüksek olan farklılıkların da bulunduğu gözlenmiştir.

3. Türkçe ortak ve mazeret sınavları için AFA ve DFA sonucunda tanımlanan modeller doğrulanmıştır. Türkçe testi ortak ve mazeret sınavlarının yapı geçerlik düzeyine yönelik toplanan ipuçları değerlendirildiğinde bu sınavların yapı geçerlik düzeylerinin bu amaçla kullanılacak bir sınav için yeterli olduğu düşünülmektedir.

Matematik ortak ve mazeret sınavları için AFA sonucunda, uzman görüşleri de değerlendirilerek cebir ve geometri olmak üzere 2 faktörlü bir yapı tanımlanmıştır. DFA ile tanımlanan bu ölçme modeli test edilmiş ve tanımlanan ölçme modelinin uyumlu olduğu gözlenmiştir. Ancak ortak sınavda ve mazeret sınavında da yapı geçerliğini zedeleyici durumlar gözlenmiştir. Uzman görüşlerine göre cebir faktöründe olması gereken bazı maddeler geometri faktörüne yüklenmiştir. Bu durumun ortaya çıkmasında testi alan bireylerin testteki bu maddeleri tam olarak anlayamamış olmasının da rolü olabileceği düşünülmektedir. Bu durum ortak sınav ve mazeret sınavının yapı geçerlik düzeyini zedeleyici bir etki yapmıştır. Literatürdeki benzer araştırmalarda tek faktörlü yapılar elde edilmiş ve faktöre yeterince yüklenmeyen maddeler analizlerden çıkarılmıştır (Anıl ve diğerleri., 2010; Örs, 2010). Ayrıca matematik ortak sınavına bulunan 14. maddenin ve matematik mazeret sınavında bulunan 18. maddenin herhangi bir faktöre yük vermemesi maddenin tam olarak anlaşılabilir olabileceğini düşündürmektedir. Madde incelendiğinde öğrencilerin maddeyi anlamakta güçlük çekebileceği sonucuna ulaşılmıştır. Matematik ortak ve mazeret sınavları üzerinde gerçekleştirilen AFA ve DFA sonucunda bu testlerin birbiri yerine kullanılmamasının daha uygun olacağı görüşü benimsenmiştir.

Araştırma sonuçları değerlendirildiğinde ortak ve mazeret sınavlarının psikometrik özelliklerinin benzer düzeyde hazırlanması için, bu sınavların test ve madde istatistikleri uzman görüşleri doğrultusunda tahmin edilerek düzenlenmesi, ortak ve mazeret sınavlarının sonuçlarının birbirinin yerine kullanılabilmesi için test eşitleme çalışmasının yapılması önerilmektedir. Bunun için sınavların uygulama süreci, test eşitlemeye olanak verecek biçimde yürütülmelidir. Testler arasındaki süre birkaç hafta olacak şekilde düzenlenerek sorular açıklanmayıp ortak maddeler teste eklenebilir. Böylece test eşitleme çalışmalarına imkân verilebilir. Faktör analizi sonucunda matematik testi ortak ve mazeret sınavları için farklı faktörlere yük veren maddeler olduğu gözlenmiştir. Örneğin faktör analizi sonucunda cebir faktöründe çıkan bir madde uzman görüşüne göre geometri faktöründe yer almakta ya da tam tersi bir durum söz konusu olmaktadır. Maddeler, öğrenciler tarafından tam olarak anlaşılmadığı için bu durumun oluşabileceği düşünülmektedir. Bu nedenle öğrencilerin seviyeleri de göz önünde bulundurularak maddeler yazılırken öğrencilerin maddeyi tam olarak anlayabilecekleri şekilde yazılmasına dikkat edilmelidir.

KAYNAKÇA

- AERA, APA, ve NCME. (2014). *Standarts for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Alıcı, D., Başol, G., Çakan, M., Kan, A., Karaca, E., Özbek, Ö. Y., ve Yaşar, M. (2011). *Eğitimde ölçme ve değerlendirme*. (S. Tekindal, Ed.). Ankara: Pegem Akademi.
- American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications: American Educational Research Association. *Educational Researcher*, 35(6), 33–40. doi.org/10.3102/0013189X035006033
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6. Baskı). Washington, DC: American Psychological Association. doi.org/10.1017/CBO9781107415324.004
- Anıl, D., ve Güzeller, C. O. (2011). Seviye belirleme sınavı fen ve teknoloji alt testi ile diğer alt testler arasındaki ilişkinin yol analizi ile incelenmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 11(1), 1–10.

- Anıl, D., Güzeller, C. O., Çokluk, Ö., ve Şekercioğlu, G. (2010). Level determination exam (SBS-2008) the determination of the validity and reliability of 7th grade mathematics sub- test. *Procedia - Social and Behavioral Sciences*, 2(2), 5292–5298. doi.org/10.1016/j.sbspro.2010.03.863
- Atıla, M. E., ve Özeken, Ö. F. (2015). Temel eğitimden ortaöğretime geçiş sınavı: Fen bilimleri öğretmenleri ne düşünüyor? *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 34(1), 124–140. doi.org/10.7822/omuefd.34.1.7
- Bracken, B. A., Carlson, J. F., Hansen, J.-I. C., Kuncel, N. R., Reise, S. P., ve Rodriguez, M. C. (2013). *APA handbook of testing and assesment in psychology*. (K. F. Geisinger, Ed.) (1. Cilt). Washington, DC: American Psychological Association.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2. Baskı). New York: The Guilford Press.
- Büyüköztürk, Ş. (2013). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum* (18. Baskı). Ankara: Pegem Akademi.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., ve Demirel, F. (2013). *Bilimsel araştırma yöntemleri* (14. Baskı). Ankara: Pegem Akademi.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. *Journal of Chemical Information and Modeling* (2. Baskı, Vol. 53). New York: Lawrence Erlbaum Associates, Inc. http://doi.org/10.1017/CBO9781107415324.004
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. http://doi.org/10.1037/0033-2909.112.1.155
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge. doi.org/10.1037/a0028079
- Dalak, O. (2015). *TEOG sınav soruları ile 8. sınıf öğretim programlarındaki ilgili kazanımları yenilenmiş Bloom Taksonomisine göre incelenmesi*. (Yüksek lisans tezi, Gaziantep Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Gaziantep). http://tez2.yok.gov.tr/ adresinden edinilmiştir.
- Dinç, E., Dere, İ., ve Koluman, S. (2014). Kademeler arası geçiş uygulamalarına yönelik görüş ve deneyimler. *Adıyaman Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 7(17), 397–423. dx.doi.org/10.14520/adyusbd.761
- Erkuş, A. (2014). *Psikolojide ölçme ve ölçek geliştirme-I: Temel kavramlar ve işlemler* (2. Baskı). Ankara: Pegem Akademi.
- George, D., ve Mallery, P. (2012). *IBM SPSS Statistics 19 Step by Step: A simple guide and reference*. *Journal of Chemical Information and Modeling* (12. Ed., Vol. 53). Boston: Pearson.
- Görmez, M., ve Coşkun, İ. (2015). *1. yılında temel eğitimden ortaöğretime geçiş reformunun değerlendirilmesi*. Ankara:SETA.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley ve Sons Inc.
- Kaiser, H. F., ve Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, 34(1), 111–117. doi.org/10.1177/001316447403400115
- Karaca, M. (2015). *TEOG sınavlarında soru sorulan ve sorulmayan fen ve teknoloji dersi konularına ilişkin öğrenci ve öğretmen bakış açıları*. (Yüksek lisans tezi, Erciyes Üniversitesi, İlköğretim Anabilim Dalı, Kayseri). http://tez2.yok.gov.tr/ adresinden edinilmiştir.
- Karaca, M., Bektaş, O., ve Armağan, F. Ö. (2015). 8. sınıf öğrencilerinin merkezi sınavlarda sorulmayan fen bilimleri konularına yönelik görüşleri. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 35(1), 63–86. http://doi.org/10.17152/gefad.99679
- Karadeniz, O., Eker, C., ve Ulusoy, M. (2015). TEOG sınavlarındaki T.C. İnkılap Tarihi ve Atatürkçülük dersine ait soruların kazanım temelli olarak değerlendirilmesi. *Uluslararası Avrasya Sosyal Bilimler Dergisi*, 6(18), 115–134.
- Karasar, N. (2014). *Bilimsel araştırma yöntemi* (26. Baskı). Ankara: Nobel Yayınları.
- Kardeş-Birinci, D. (2014). Merkezi sistem ortak sınavlarda ilk deneyim: Matematik dersi. *Eğitim ve Öğretim Araştırmaları Dergisi*, 3(2), 8–16.
- Kline, R. B. (2011). *Principles and practise of structural equating modeling*. (T. G. Press, Ed.) (3. Baskı). New York.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575.
- MEB. (2015). 2014-2015 öğretim yılı ortak sınavlar e-kılavuzu, [Çevrim-İçi: http://oges.meb.gov.tr/meb_iys_dosyalar/2014_11/04053521_ortaksnavlareklavuz20142015.pdf], erişim tarihi: 20 Mart 2015
- Meydan, C. H., ve Şeşen, H. (2015). *Yapısal eşitlik modellemesi* (2. Baskı). Ankara: Detay Yayıncılık.
- Mueller, R. O. (1996). *Basic principles of structural equation modeling: An introduction to LISREL and EQS. Design* (Vol. 102). New York: Springer Science ve Business Media, LLC. http://doi.org/10.1016/j.peva.2007.06.006

- Örs, S. (2010). *İlköğretim 6., 7. ve 8. sınıf Seviye Belirleme Sınavı fen ve teknoloji alt testlerinin faktör yapılarının belirlenmesi*. (Yüksek lisans tezi, Ankara Üniversitesi, Ölçme ve Değerlendirme Anabilim Dalı, Ankara). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Özgüven, İ. E. (2011). *Psikolojik testler*. Ankara: PDREM Yayınları.
- Özkan, M. (2015). *TEOG kapsamında uygulanan matematik alt testi ile matematik mazeret alt testinin istatistiksel eşitliğinin sınanması*. (Yayınlanmamış yüksek lisans tezi). Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Özsoy, S., ve Özsoy, G. (2013). Eğitim araştırmalarında etki büyüklüğünün raporlanması. *İlköğretim Online*, 12(2), 334–346.
- Schermelleh-Engel, K., Moosbrugger, H., ve Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- Şad, S. N., ve Şahiner, Y. K. (2016). Temel eğitimden ortaöğretime geçiş (TEOG) sistemine ilişkin öğrenci, öğretmen ve veli görüşleri. *İlköğretim Online*, 15(1), 53–76.
- Tabachnik, B. G., ve Fidell, L. S. (2012). *Using multivariate statistics* (6. Baskı). Boston: Pearson.
- Thalheimer, W., ve Cook, S. (2002). How to calculate effect sizes from published research: A simplified methodology. <http://doi.org/10.1113/jphysiol.2004.078915>
- Turgut, M. F., ve Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme* (4. Baskı). Ankara: Pegem Akademi.
- West, S. G., Finch, J. F., ve Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: SAGE.
- Yılmaz, G. (2014). *SBS ve TEOG sınavındaki söz varlığımızla ilgili Türkçe sorularının 2005 Türkçe Öğretim Programı'ndaki sözcük varlığımızla ilgili kazanımları ölçme yeterlilik düzeyinin öğretmen görüşlerine göre incelenmesi*. (Yüksek lisans tezi, Erzincan Üniversitesi, Türkçe Eğitimi Anabilim Dalı, Erzincan). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Yılmaz-Koçar, E., ve Aygün, B. (2015). Temel eğitimden orta öğretime geçiş sınavı (TEOG)'nın matematik temel alanına ait testlerin kapsam geçerliğinin incelenmesi. *Pegem Eğitim ve Öğretim Dergisi*, 5(5), 667–680. <http://doi.org/10.14527/pegegog.2015.036>
- Yorgancı, O. K. (2015). *Sekizinci sınıf Türkçe dersi ortak sınavı sorularının öğretim programına göre değerlendirilmesi*. (Yüksek lisans tezi, Gazi Üniversitesi, Türkçe Eğitimi Anabilim Dalı, Ankara). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Yurdugül, H. (2006). Parelel, eşdeğer ve konjenetik ölçmelerde güvenilirlik katsayılarının karşılaştırılması. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 39(1), 15–37.
- Yurdugül, H., ve Bayrak, F. (2012). Ölçek geliştirme çalışmalarında kapsam geçerlik ölçüleri: Kapsam geçerlik indeksi ve Kappa istatistiğinin karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, (Özel Sayı 2), 264–271.
- Zayimoğlu-Öztürk, F., ve Aksoy, H. (2014). Temel eğitimden ortaöğretime geçiş modelinin 8 . sınıf öğrenci görüşlerine göre değerlendirilmesi (Ordu ili örneği). *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 33(2), 439–454. <http://doi.org/10.7822/omuefd.33.2.8>

EXTENDED ABSTRACT

Introduction

Tests are applied almost at all levels of education system in Turkey and as a result of these tests decisions made about individuals. Transition from basic to secondary education exam (TEOG) at lower secondary level, transition to higher education exam at upper secondary level (YGS), undergraduate placement exam (LYS) can be given as example for the exams applied in nation-wide. Common characteristics of these exams are being highly determinant for the life of individuals. For the tests having significant power of influence, it is necessary to collect evidence for validity and reliability.

Analyzing the validity and reliability of a test gives off important evidence so collecting these data especially for the national exams are necessary. Transition from basic to secondary education exam which has been applied since 2013-2014 school year is one of these exams. This examination system aims to evaluate the achievements of individuals throughout the education duration (MEB, 2015). Transition from basic to secondary education exam contains Turkish, mathematics, science and

technology, education of religion and ethics, TR History of Revolution and Kemalism and foreign language course. This exam is applied by Ministry of National Education.

Common and make up exams are carried out within the scope of TEOG exams. Individuals who couldn't attend common exam and having a valid excuse can attend make up exams instead of common exams on a date determined by the Ministry of National Education. On that case it is come out as a matter whether tests applied in both exams and their items are identical or not. In these two tests individuals are evaluated in the same way and decisions made through the same criterion, if the items aren't identical this cause disadvantage for one of the groups. To be able overcome this risk psychometric properties of both tests should be investigated, equivalence of both tests should be provided and necessary precautions should be taken for similar test to be conducted.

In this study the psychometric properties of Turkish and math subtests in common and make-up examination of TEOG have been compared.

Method

Sample of the research is consisted of randomly selected 10.000 students attended common exams. Data obtained from the individuals in the sample analyzed and data with systematically repeated answers and same answers to all questions was omitted. Research was carried out via item scores matrix of 9773 students for Turkish common exam, 9485 students for math common exam, 2747 students for Turkish make-up exam and 2323 students for math make-up exam.

It was consulted to expert opinion for the content validity study. According to this, in order to determine and compare the acquisitions that is measured in common and make-up examinations of TEOG, question-acquisition matching form and question assessment form was prepared by researchers.

To conduct analysis Statistica 10, Factor 10.3, Lisrel 8.80, ITEMAN 3.5 and Excel 2016 package programs were used.

Results and Discussion

As a result of research, item and test statistics of the common and make up exams have been compared and it has been observed that test and item statistics are different in significantly ($\alpha=0,05$). Because of this situation, it can be said that fairness of exams has been damaged. Common and make up exams content validity which based on expert opinions is in similar level.

To evaluate clues related construct validity, it is believed that Turkish common and make up exams construct validity level is adequate for using this purpose. But the results of the study revealed that math's common and make-up exams items don't measure the same things in terms of construct validity.

Therefore, common and make-up examinations are needed to be prepared in similar level by investigating the test and item statistics of them through expert predictions. Besides, it can be conducted test equating studies for this exams. For this purpose, common exam questions shouldn't announce until make up exams being applied.