

T.C.
AKDENİZ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANA BİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
TEZLİ YÜKSEK LİSANS PROGRAMI

**TIMSS 2015 SEKİZİNCİ SINIF MATEMATİK BAŞARI TESTİNİN OECD
ÜLKELERİNE GÖRE ÖLÇME DEĞİŞMEZLİĞİNİN İNCELENMESİ**

YÜKSEK LİSANS TEZİ

Özge ÖNCÜ

Antalya

2019

T.C.
AKDENİZ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANA BİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
TEZLİ YÜKSEK LİSANS PROGRAMI

**TIMSS 2015 SEKİZİNCİ SINIF MATEMATİK BAŞARI TESTİNİN OECD
ÜLKELERİNE GÖRE ÖLÇME DEĞİŞMEZLİĞİNİN İNCELENMESİ**

YÜKSEK LİSANS TEZİ

Özge ÖNCÜ

Danışman:

Dr. Öğr. Üy. Hakan KOĞAR

Antalya

2019

İMZA SAYFASI

T.C.

AKDENİZ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

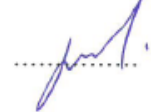
..dege... ..'nın bu çalışması ..05..06..2019... tarihinde jürimiz tarafından Eğitim Bilimleri Anabilim Dalı Eğitimde İktisadi ve Sosyal Bilimler Tezli Yüksek Lisans Programında **Yüksek Lisans Tezi** olarak oy birliği/oy çokluğu ile kabul edilmiştir

İMZA

Başkan : Dr. Öğr. Üyesi...Güçlü SEKERCİOĞLU
(..Akdeniz.. Üniversitesi.....)



Üye : Dr. Öğr. Üyesi Sema DİYAR
(..Mehmet Akif Ersoy Üniversitesi.....)



Üye (Danışman) : ..Dr. Öğr. Üyesi...Hakan KOBAR
(..Akdeniz.. Üniversitesi.....)



YÜKSEK LİSANS TEZİNİN ADI:

ONAY: Bu tez, Enstitü Yönetim Kurulunca belirlenen yukarıdaki jüri üyeleri tarafından uygun görülmüş ve Enstitü Yönetim Kurulunun tarihli ve sayılı kararıyla kabul edilmiştir.

Doç. Dr. Ramazan KARATAŞ

Enstitü Müdürü

DOĞRULUK BEYANI

Yüksek lisans tezi olarak sunduğum bu çalışmayı, bilimsel etik ve geleneklere aykırı düşecek bir yol ve yardıma başvurmaksızın yazdığımı, yararlandığım araştırmaların kaynakçada gösterilenlerden oluştuğunu ve bu kaynakları her kullanımında alıntı yaparak yararlandığımı belirtir; bunu onurumla doğrularım. Enstitü tarafından belli bir zamana bağlı olmaksızın, tezimle ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara katlanacağımı bildiririm.

02 / 05 / 2019

Özge ÖNCÜ

TEŞEKKÜR

Tez çalışmamı yürüttüğüm bu zorlu süreçte, hayatımda önemli yerlere sahip ve benim için çok değerli olan kişilerin katkıları olmuştur.

Öncelikle çalışmam süresince, her konuda iyi niyetini ve desteğini hiçbir zaman benden esirgemeyen ve beni her zaman yüreklendiren saygıdeğer danışman hocam Dr.Öğr. Üy. Hakan KOĞAR'a,

Çalışmalarım sırasında bıkmadan, usanmadan tüm sorularımı cevaplayan, bilgi ve deneyimleriyle gelişimime çok büyük katkısı olan ve her konuda bana destek olan değerli hocam Dr. Öğr. Üy. Güçlü ŞEKERCİOĞLU'na, ayrıca bilgi ve önerileriyle çalışmama katkı sağlayan hocam Doç. Dr. Bayram BIÇAK'a,

Her zaman yanımda olan ve gülen gözleri ile içimi doldurup bu zorlu süreçte beni hiç yalnız bırakmayan sevgili eşim Salih ÖNCÜ'ye,

Hayatımın her anında yanımda olan, bana her zaman güvenen ve arkamda duran, beni bugünlere getiren, beni ben yapan en büyük destekçilerim canım annem Şerife GÖÇER'e, canım babam Mümtaz GÖÇER'e, canım biricik kardeşim Öykü GÖÇER'e sonsuz teşekkürler...

ÖZET

TIMSS 2015 SEKİZİNCİ SINIF MATEMATİK BAŞARI TESTİNİN OECD ÜLKELERİNE GÖRE ÖLÇME DEĞİŞMEZLİĞİNİN İNCELENMESİ

ÖNCÜ, Özge

Yüksek Lisans, Eğitim Bilimleri Ana Bilim Dalı

Tez Yöneticisi: Dr. Öğr. Üy. Hakan Koğar

Mayıs 2019, 132 Sayfa

Bu çalışmada, TIMSS 2015 verileri kullanılarak, matematik başarı testine katılan sekizinci sınıf öğrencilerinin matematik başarılarının ülkelere göre ölçme değişmezliği incelenerek, ülkeler bakımından farklılık gösterip göstermediği, değişmezliği sağlayıp sağlamadığı, sağlıyor ise en iyi hangi değişmezlik türünde sağlandığı tespit edilmiş, önceden bu alanda yapılmış çalışmalarda ortak ve farklı yönleri bakılarak, uygulayıcıya ve araştırmaya yönelik önerilere önemli noktalarda değinilmiştir.

İlk aşamada, ilişkisel tarama modelinde yürütülen araştırmaya TIMSS 2015 değerlendirmesinde uygulanan bilişsel test kitapçıklarından tüm ülkelerde ortak olarak uygulanmış olan 11 no'lu kitapçık seçilmiştir ve araştırma TIMSS'e katılan ülkelerde 11 no'lu kitapçığı almış olan 301.472 öğrenciye ait veri seti üzerinden yürütülmüştür. Ülkelerin sekizinci sınıf matematik başarı testi puanları için test istatistikleri, normallik testleri ve güvenilirlik katsayılarına bakılmış, merkezi eğilim ve dağılım ölçülerinin birbirine yakın olduğu sonucuna varılmıştır. Normallik için basıklık ve çarpıklık katsayıları incelenerek, veri setinin normale yakın bir dağılım gösterdiğini söylemek mümkündür. Test maddelerinin güvenilirlik katsayılarının ve iç tutarlılığının genel olarak iyi olduğu sonucuna ulaşılmıştır. Daha sonraki aşamada OECD ülkeleri seçilerek bu ülkelerde bulunan sekizinci sınıfa giden 7.223 öğrenci için ayrı ayrı doğrulayıcı faktör analizi yapılarak χ^2/sd oranının oldukça iyi olduğu sonucuna ulaşılmıştır. Uyum indekslerine bakıldığında, CFI değerlerinin örnekleme

yer alan bütün ülkelerde iyi uyum gösterdiği, GFI değerlerinin Hong Kong, İsveç, İtalya, Japonya, Kore ve Türkiye için iyi uyuma yakın, diğer ülkeler için oldukça iyi uyum gösterdiği, SRMR değerlerinin Hong Kong, İsveç, İtalya için iyi uyuma yakın, geri kalan ülkeler içinse çok iyi uyum gösterdiği tespit edilerek uyum indeksleri genel olarak değerlendirildiğinde tüm ülkelere ait ölçme modellerinin tüm gruplar için ayrı ayrı doğrulandığına karar verilmiştir.

Daha sonra tüm gruplar için kovaryans matrislerinin eşitliği testi yapılarak ölçme değişmezliğinin test edilmesi aşamasına geçilmiştir. OECD üyesi ülkelerin matematik başarı testi puanları için kovaryans matrislerinin eşitliği testi sonuçlarını ele aldığımızda, uyum indekslerinin oldukça iyi bir uyum gösterdiği ifade edilebilir.

H_0 (daha kısıtlı model ile daha az kısıtlı model arasında uyum açısından anlamlı bir fark yoktur) ve H_1 (daha kısıtlı model ile daha az kısıtlı model arasında uyum açısından anlamlı bir fark vardır) hipotezleri kurularak, ölçme değişmezliği çoklu-grup doğrulayıcı faktör analizi ile test edilmiştir. Yapısal değişmezlik için uyum indekslerinin değerlerinin kabul edilebilir düzeylerde olduğu ve modelin doğrulandığı söylenebilir. Daha sonra Model 1'e alternatif olarak kurulan Model 2, Model 3 ve Model 4'ün yuvalanmış model (nested) temel alınarak karşılaştırılması sonucunda, sırasıyla yapısal değişmezlik (Model 1) ve zayıf faktöriyel değişmezlik (Model 2), yapısal değişmezlik (Model 1) ve güçlü faktöriyel değişmezlik (Model 3), zayıf faktöriyel değişmezlik (Model 2) ve güçlü faktöriyel değişmezlik (Model 3) arasında manidar bir fark olması sebebi ile H_0 ret edilmiştir (H_1 kabul edilmiştir). Yapısal değişmezlik (Model 1) ve katı faktöriyel değişmezlik (Model 4), güçlü faktöriyel değişmezlik (Model 3) ve katı faktöriyel değişmezlik (Model 4) arasında manidar bir farkın olmaması sebebi ile H_0 kabul edilmiştir (H_1 ret edilmiştir) şeklinde ifade edilebilir. Sonuç olarak, araştırma sonuçları OECD üyesi ülkeler için ölçme değişmezliğinin sağlanamadığını, ölçme değişmezliğine ilişkin modeller arasında en iyi çalışan modelin güçlü faktöriyel değişmezlik modeli olduğunu göstermiştir.

Literatürde bazı araştırma sonuçları, geniş ölçekli çalışmalarda farklı ülkeler arasında ölçme değişmezliğinin sağlandığını, bunun aksine bazı araştırma sonuçları da farklı dil ve farklı kültürlere sahip bireyler arasında kullanılan ölçme araçlarında ölçme değişmezliğinin sağlanamadığını raporlaştırmıştır. Bu sonuçlar bağlamında, madde yazımında her ülkenin kendine has yaşanmışlığı, tarihi ve kültürel değer ve

gelenekleri olması sebebi ile farklı kültürden ve milletlerden insanların komisyonda yer alması gerekliliği, sınav sorusu hazırlanmasından, sınavın uygulanması hatta sınavın sonucunun açıklanıp analiz edilmesine kadar eğitimde karar alıcılar ve politikacıların bu sınavlarda çok önemli işlevleri olması sebebi ile grup karşılaştırmalarına çok da güvenmemek gerektiği, ölçülen özellik ve uygulama türüne göre ölçme değişmezliğinin yorumlama, karşılaştırma çalışmalarının değişiklik gösterebileceği ve bu yüzden çalışmanın titizlikle yapılması gerekliliği uygulayıcıya yönelik önerileri oluşturur. Araştırmaya yönelik öneriler olarak da gelecek çalışmalarda farklı ülkeler ve demografik değişkenlerin de dahil edilmesi, farklı kitapçık türlerinin ve başarı alanlarının da analiz edilmesi, ölçme değişmezliğinin daha farklı yöntemlerle test edilerek, hangi yöntemin daha etkili olabileceğinin bulunması, farklı gruplar arasında dil, cinsiyet, kültür gibi değişkenlerle karşılaştırmalar yapılmadan önce yapılan yorumların daha anlamlı ve sağlıklı olabilmesi için ölçme aracına ilişkin ölçme değişmezliği çalışmalarının büyük bir titizlikle yapılması oldukça önemlidir.

Tüm sonuçlar değerlendirildiğinde, TIMSS 2015 uygulamasında matematik başarısına ait elde edilen sonuçların farklı ülkeler, diller, cinsiyetler ve kültürlere sahip katılımcılardan elde edilen sonuçların karşılaştırılabilirliği ve buna bağlı olarak yapılan yorumlar ölçme değişmezliğinin sağlanamaması yani uygulanan ölçme aracının tüm katılımcılar için aynı anlamı ifade etmemesi nedeniyle tartışmaya açık hale gelmektedir. Bununla birlikte, geniş ölçekli sınav uygulamalarının sonuçları değerlendirilirken ölçme değişmezliğinin sağlanamıyor oluşu, özellikle ülkeler arası karşılaştırmalar ve yorumlamalar yapılırken titizlikle göz önünde bulundurulması gereken bir durumdur.

Anahtar Sözcükler: TIMSS 2015, matematik başarısı, kovaryans matrislerinin eşitliği, yapısal değişmezlik, zayıf faktöriyel değişmezlik, güçlü faktöriyel değişmezlik, katı faktöriyel değişmezlik, ölçme değişmezliği, çoklu-grup doğrulayıcı faktör analizi.

ABSTRACT

AN INVESTIGATION INTO THE MEASUREMENT INVARIANCE ACCORDING TO OECD COUNTRIES OF TIMSS 2015 EIGHTH GRADE MATH ACHIEVEMENT TEST

ÖNCÜ, Özge

Master of Arts, Department of Educational Sciences

Supervisor: Assist. Prof. Dr. Hakan Koğar

May 2019, 132 Pages

In this study, by using TIMSS 2015 data, the mathematical achievement of the eighth grade students who participated in the mathematics achievement test is analysed. It is tried to determined what kind of invariance is the best provided if this success is different in terms of countries, whether it provides invariance or not. Considering common and different aspects in the studies conducted previously in this field, the recommendations for the practitioner and the research are mentioned in important points.

In first step, among cognitive test booklets implemented in TIMSS 2015, booklet 11 which was used commonly by all countries was selected for this correlational survey study and the study was conducted using dataset belonging to 301.472 students that took booklet 11 in the countries that attend TIMSS. Test statistics, normality tests and reliability coefficients were examined for the eighth grade mathematic achievement test scores of the countries and it was concluded that the measures of central tendency and distribution were similar. For simplicity, it is possible to say that the data set has a close distribution to normal. The reliability coefficients and internal consistency of test items were generally good. In the following stage, OECD countries were selected and 7.223 students attending the eighth grade in these countries were subjected to confirmatory factor analysis and the measurement models were validated for all groups. The χ^2 / sd ratio was found to be quite good.

When considered to the compliance indices, it is seen that CFI datas are compatible with the patterns in all countries. It is determined GFI datas are delivering compatible results in Hong Kong, Swedish, Italy, Japan, Korea and Turkey , for the other countries it extremely shows the compatible results, SRMR datas are closed to harmony in Hong Kong, Swedish, Italy, and for the rest of the other countries it is estimated that they show the best compatible. When compatible index evaluated in general, it is decided that evaluation models of all countries are verified for all groups.

Then, for all groups, covariance matrices were tested for equality. The other step is testing of measurement invariance. Considering the results of the covariance matrices equality test for the mathematics achievement test scores of OECD member countries, it can be stated that the fit indexes are in good agreement. H_0 (there is no significant difference between the more restricted model and the less restricted model) and H_1 (there is a significant difference between the more restricted model and the less restricted model) hypotheses are established, and the measurement invariance is tested by multi-group confirmatory factor analysis. It can be said that the values of the fit indices for the structural invariance are acceptable and the model is confirmed. Then, H_0 has been rejected because of a significant difference between the models based on the nested model based on Model 2, Model 3 and Model 4, which were established as an alternative to Model 1, respectively the configural (structural) invariance (Model 1) and the weak factorial invariance (Model 2), configural (structural) invariance (Model 1) and strong factor invariance (Model 3), weak factorial invariance (Model 2) and strong factorial invariance (Model 3). (H_1 was accepted.) H_0 was accepted because of a significant difference between the configural (structural) invariance (Model 1) and the strict factorial invariance (Model 4), the strong factorial invariance (Model 3) and the strict factorial invariance (Model 4) (H_1 was rejected). As a result, the results of the study showed that the measurement invariance was not achieved for OECD member countries, and the best model for measurement invariance was the strong factorial invariance model.

In the literature, some research results reported that the measurement invariance between different countries was ensured in large scale studies. On the contrary, some research results reported that the measurement invariance could not be achieved in the measurement tools used among individuals with different languages and different

cultures. In the context of these results, the necessity of the presence of people from different cultures and nationalities in the article writing because each country has its own unique experience, historical, cultural values and traditions. From the preparation of the exam question, to the implementation of the exam and even to the examination and analysis of the exam result, the decision-makers and politicians in the education should not be too confident in the group fixtures because they have very important functions in these exams. The interpretation of the measurement invariance according to the measured characteristics and the type of application, and the efforts to be meticulously performed, create suggestions for the practitioner. Including different countries and demographic variables in future studies as suggestions for research, analysis of different booklet types and success areas, determining which method can be more effective by testing the measurement invariance with different methods, In order to make the interpretations made before different comparisons between different groups such as language, gender and culture, to be more meaningful and healthy, it is very important to carry out the measurement invariance of the measurement tool with great care.

When all the results are evaluated, the comparability of the results obtained from the results of mathematics achievement in the TIMSS 2015 application against the participants with different countries, languages, genders and cultures, and the comments made accordingly, cannot be ensured by the measurement invariance is becoming. However, measurement of the results of large-scale exam applications can not be achieved in measuring the results, especially when making comparisons between countries and must be taken into consideration diligently.

Keywords: TIMSS 2015, mathematics achievement, equality of covariance matrices, configural (structural) invariance, weak factorial invariance, strong factor invariance, strict factorial invariance, measurement invariance, multi-group confirmatory factor analysis.

İÇİNDEKİLER

İMZA SAYFASI	i
DOĞRULUK BEYANI	ii
TEŞEKKÜR	1
ÖZET	2
ABSTRACT	5
AN INVESTIGATION INTO THE MEASUREMENT INVARIANCE ACCORDING TO OECD COUNTRIES OF TIMSS 2015 EIGHTH GRADE MATH ACHIEVEMENT TEST	5
TABLolar LİSTESİ	10
BÖLÜM I	13
GİRİŞ	13
1.1. Problem Durumu.....	13
1.2. Araştırmanın Amacı.....	20
1.2.1. Alt Problemler.....	20
1.3. Araştırmanın Önemi.....	21
1.4. Sınırlılıklar.....	21
1.5. Tanımlar.....	22
1.6. Kısaltmalar.....	22
BÖLÜM II	24
KAVRAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR	24
2.1. Geçerlilik.....	24
2.2. Faktör Analizi.....	27
2.2.1. Açımlayıcı Faktör Analizi (Exploratory Factor Analysis).....	28
2.2.2. Doğrulamalı Faktör Analizi (Confirmatory Factor Analysis).....	29
2.4. Ölçme Değişmezliği.....	37
2.4.1. Yapısal Değişmezlik (Configural Invariance).....	41
2.4.2. Metrik Değişmezlik (Metric Invariance).....	41

2.4.3.Güçlü Faktöriyel Değişmezlik (Strong Factorial Invariance)	42
2.4.4.Katı Faktöriyel Değişmezlik (Strict Factorial Invariance)	43
2.5. Ölçme Değişmezliği Modellerinin Karşılaştırılması	44
2.6. İlgili Araştırmalar.....	45
BÖLÜM III	51
YÖNTEM.....	51
3.1. Araştırma Modeli	51
3.2. Evren ve Örneklem	51
3.3. Araştırma Verileri	54
3.4. Verilerin Analizi	69
BÖLÜM IV	80
BULGULAR	80
4.1. OECD Üyesi Ülkeler İçin Ölçme Değişmezliğine İlişkin Bulgular	80
BÖLÜM V	87
SONUÇ, TARTIŞMA VE ÖNERİLER.....	87
5.1. Sonuç ve Tartışma.....	87
5.2. Öneriler	89
KAYNAKÇA	91
EKLER.....	104

TABLolar LİSTESİ

Tablo 1: Ölçme Değişmezliği Modelleri	40
Tablo 2: OECD Ülkelerine Göre Örneklem Dağılımı	53
Tablo 3: Analize Dâhil Edilen Matematik Başarı Testi Maddeleri.....	56
Tablo 4: TIMSS 2015 Matematik 8. Sınıf İngilizce – Türkçe Madde Başlıkları ve Cevap Anahtarı	57
Tablo 4'ün devamı: TIMSS 2015 Matematik 8. Sınıf İngilizce – Türkçe Madde Başlıkları ve Cevap Anahtarı.....	58
Tablo 4'ün devamı: TIMSS 2015 Matematik 8. Sınıf İngilizce – Türkçe Madde Başlıkları ve Cevap Anahtarı.....	59
Tablo 4'ün devamı: TIMSS 2015 Matematik 8. Sınıf İngilizce – Türkçe Madde Başlıkları ve Cevap Anahtarı.....	60
Tablo 5: TIMSS Matematik Maddelerinin 8. Sınıf Bilişsel Alanlara Göre Yüzdesele Dağılımı.....	61
Tablo 6: TIMSS Matematik Maddelerinin Bilişsel Alanlara Göre Yüzdesi	61
Tablo 7: TIMSS Matematik Maddelerinin Sekizinci Sınıf Matematik Öğrenme Alanlarına Göre Yüzdesele Dağılımı.....	62
Tablo 8: TIMSS 8. Sınıf Matematik Maddelerinin Matematik Öğrenme Alanlarına Göre Yüzdesi	63
Tablo 9: TIMSS'e Ait Yıllara Göre Temel Bilgiler ve Katılım Durumu	65
Tablo 10: Türkiye'de TIMSS 2011 Uygulamasına Katılan Kişilere Dair Sayısal Veriler.....	65
Tablo 11: 1999, 2007 ve 2011'de TIMSS 8. Sınıf Matematik Başarısının Ülke Sayısı Bağlamında Sıralaması.....	66

Tablo 12: 1999, 2007 ve 2011’de TIMSS 8. Sınıf Cinsiyete Göre Matematik Alanındaki Öğrenci Başarısı	67
Tablo 13: 1999, 2007 ve 2011’de TIMSS Uygulamasının Yeterlik Düzeyine Göre Matematik Başarı Yüzdesi	68
Tablo 14: 2007 Yılından 2011 Yılına 8. Sınıf Matematik Başarısında En Fazla Artış Gösteren 10 Ülke.....	68
Tablo 15: OECD Üyesi Ülkelerin 8. Sınıf Matematik Başarı Testi Puanları için Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları	71
Tablo 16: OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Kovaryans Matrislerinin Eşitliği Testi Sonuçları	72
Tablo 17: OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Doğrulayıcı Faktör Analizi Sonuçları	73
Tablo 18: OECD Ülkelerinin Pearson Korelasyon Değer Aralığı.....	75
Tablo 19: OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları.....	81
Tablo 20: OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları.....	82
Tablo 21: OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları.....	83
Tablo 22: OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları.....	84
Tablo 23: OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları.....	85

Şekil 1. TIMSS 2015 Sekizinci Sınıf Matematik 2. Madde Şekli	104
Şekil 2. TIMSS 2015 Sekizinci Sınıf Matematik 7. Madde Şekli	106
Şekil 3. TIMSS 2015 Sekizinci Sınıf Matematik 9. Madde Şekli	107
Şekil 4. TIMSS 2015 Sekizinci Sınıf Matematik 10. Madde Şekli	108
Şekil 5. TIMSS 2015 Sekizinci Sınıf Matematik 11. Madde Şekli	109
Şekil 6. TIMSS 2015 Sekizinci Sınıf Matematik 11. Madde Şekli	109
Şekil 7. TIMSS 2015 Sekizinci Sınıf Matematik 12. Madde Şekli	110
Şekil 8. TIMSS 2015 Sekizinci Sınıf Matematik 13. Madde Şekli	111
Şekil 9. TIMSS 2015 Sekizinci Sınıf Matematik 14. Madde Şekli	112
Şekil 10. TIMSS 2011 Sekizinci Sınıf Matematik 2. Madde Şekli	113
Şekil 11. TIMSS 2011 Sekizinci Sınıf Matematik 3. Madde Şekli	114
Şekil 12. TIMSS 2011 Sekizinci Sınıf Matematik 4. Madde Şekli	115
Şekil 13. TIMSS 2011 Sekizinci Sınıf Matematik 5. Madde Şekli	116
Şekil 14. TIMSS 2011 Sekizinci Sınıf Matematik 7. Madde Şekli	117
Şekil 15. TIMSS 2007 Sekizinci Sınıf Matematik 1. Madde Şekli	119
Şekil 16. TIMSS 2007 Sekizinci Sınıf Matematik 3. Madde Şekli	120
Şekil 17. TIMSS 2007 Sekizinci Sınıf Matematik 5. Madde Şekli	121
Şekil 18. TIMSS 2007 Sekizinci Sınıf Matematik 7. Madde Şekli	122
Şekil 19. TIMSS 2007 Sekizinci Sınıf Matematik 11. Madde Şekli	124

BÖLÜM I

GİRİŞ

Bu bölümde, TIMSS 2015 sekizinci sınıf matematik başarısının ülkelere göre ölçme değişmezliğinin incelenmesi çalışmasının problem durumu, ilgili çalışmanın amacı, sonraki aşamada çalışmanın alt problemleri, araştırmanın önemi, sınırlılıkları ve tanımlarına ilişkin bilgilere sırasıyla değinilmiştir.

1.1. Problem Durumu

Çağımızda her geçen gün teknoloji başta olmak üzere bütün alanlarda kendisini iyice hissettiren rekabet ortamı, gelişmiş ve gelişmekte olan ülkelerin en önemli hedefini, sürdürülebilir gelişmeyi gerçekleştirmeye zorlamaktadır. Dünya üzerinde gerçekleşen bu gelişmelerin aynı zamanda eğitim sistemini de derinden etkilediği söylenebilir.

Sovyetlerin uzaya ilk kozmonotlarını yollamaya başlaması ile dünyanın birçok ülkesinde teknoloji hızlı bir şekilde ilerlemiştir. Bulduğumuz çağa damgasını vuran bilgi teknolojileri, özellikle 1900'lü yılların ikinci yarısından itibaren başta Amerika Birleşik Devletleri olmak üzere dünyanın birçok ülkesinde fen ve matematik bilimlerine ve bunların eğitime verilen önemin giderek artmasına sebep olmuştur ve eğitim hedefleri zaman içerisinde büyük değişimlere uğramıştır (Korkmaz, 2004). Ülkelerin statüsünü belirleyen en önemli unsurlardan biri olan bilim ve teknoloji toplumsal ilerlemenin başlangıç noktası olarak kabul edilmektedir (Bayır, Çakıcı ve Atalay, 2016). Özellikle teknolojik yarışta geri kalmak istemeyen bu ülkeler çareyi fen bilimleri ve ilgili alanlara önem vermekte ve bu önem doğrultusunda öğretimi geliştirmekte bulmuşlardır (Çepni, Ayas, Johnson ve Turgut, 1997). Teknolojik gelişmeler, insan kaynakları ve bu alanda verilen eğitim, ülkelerin gelişmesini, yaşam standartlarının yükselmesini, daha bilgili nesiller yetişmesini sağlamasından dolayı, kalite ve verimin merkezinde bulunması sebebiyle de bilgi toplumlarında oldukça önemli bir konumdadır (Ergül, 1999). Bu duruma günümüzde de oldukça önem verilmektedir ve artan bir şekilde devam etmektedir.

İnsan gn kaynađının çağın gereklerine uygun bilgi ve becerilere sahip olarak yetiřtirilmesi lkelerin sosyal, ekonomik ve kltrel aıdan ilerlemesinin vazgeilmez gereklerindedir. Bilgi ve teknolojilerin srekli geliřmesi nedeniyle insan kaynađının gerekli donanımlarla yetiřtirilmesi iin eđitim ve đretimin srekli gncellenmesi gerekmektedir. Bu nedenle nemli bazı uluslararası kuruluřlar eđitim alanında arařtırmalar yapmaktadır. Sonularının lkelerin eđitim politikalarına yn verdiđi uluslararası dzeyde yapılan bu arařtırmalar, lkelerin eđitim sistemleri zerine getirebilecekleri nemli birer eleřtiri ve geri bildirim niteliđindedir.

Kreselleřen dnyada, eđitim politikalarının belirlenmesinde ve yeniden dzenlenmesinde etkili olan ulusal lme ve deđerlendirme alıřmalarının yanı sıra uluslararası alıřmalar da nem kazanmaktadır. Uluslararası dzeyde lkelerin eđitim sistemi aısından hangi konumda olduđunun, diđer lkelerle olan farklılıklarının ve eđitim eksiklerinin belirlenebilmesi iin birok lke tarafından referans olarak kabul edilen ve uluslararası dzeyde uygulanan sınavlar bulunmaktadır. đrenci bařarıları gz nnde bulundurularak pek ok lkede eđitim sistemi ve đelerine ynelik eksikliklerin belirlenmesi ve gereksinim duyulan alanların geliřtirilmesi alıřmaları yrtlmektedir (MEB, 2014). Bunların yksek nem tařıyan arařtırmalar olduđu dřnlmektedir nk deđerlendiriciler, đrenci performansını karřılařtırmak ve katılımcı lkelerdeki eđitim politikalarının ve uygulamalarının etkililiđini deđerlendirmek iin sonuları kullanmaktadır (Gierl, 2000).

Geliřmiř lkelerin, zamanın řartlarına uygun fen bilimleri eđitim programlarını geliřtirme abası iinde olmaları, lkemizin de bu dođrultuda eđitim sistemini srekli olarak gzden geirmesi gerektiđini dřnmesine yol amıřtır. Avrupa ve Orta Asya blgelerinde 29 lkeyi ieren yakın zamanda yapılan bir arařtırmaya gre eđitim, Trk vatandařlarının endiřelendiđi en nemli konudur. Bu blgelerde devlet yatırımlarında ncelikli alanın eđitim olması gerektiđini dřnen lkelerde, en yksek oran Tacikistan'dan sonra Trkiye'ye aittir (elebi, Gner, Kaya ve Korumaz, 2014). Bu sebeple Trkiye'de, fen bilimleri ve matematik eđitimindeki geliřimlerini daha iyi grebilmek iin TIMSS (Trends in International Mathematics and Science Study), PISA (Programme for International Student Assessment) ve PIRLS (Progress in International Reading Literacy Study) gibi uluslararası karřılařtırma sınavlarına katılmıřtır.

Türkiye'nin de katıldığı uluslararası düzeyde öğrenci başarılarını belirlemek amacıyla eğitim araştırmaları yürütülmektedir. Bu araştırmalardan biri de Uluslararası Matematik ve Fen Eğilimleri Araştırmasıdır (TIMSS-Trends in International Mathematics and Science Study). Öğrencilerin matematik ve fen alanlarında kazandıkları bilgi ve becerilerin değerlendirilmesine yönelik bir tarama araştırması olan TIMSS, Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (International Association for the Evaluation of Educational Achievement) IEA'nın bir projesidir ve dört yıllık aralıklarla dördüncü ve sekizinci sınıf düzeyindeki öğrencilere uygulanmaktadır.

TIMSS'de yer alan matematik ve fen soruları daha önceden belirlenen kazanımlar çerçevesinde ülke temsilcileri tarafından ortak olarak hazırlanır. Hazırlanan bu sorular IEA'nın Fen ve Matematik maddeleri inceleme komitesi tarafından incelenir ve açık uçlu soruların puanlama anahtarları hazırlanır. Daha sonra oluşturulan yedek ve esas sorular, taslak bloklar halinde incelenir ve sorulara son şekli verilir. Hazırlanan sorular katılımcı ülkelerde çeviri, uyarlama işlemlerinden sonra pilot uygulama ile test edilir. İşleyen sorular önceki uygulama soruları ile birleştirilerek pilot uygulamadan bir yıl sonra nihai uygulamada yer alır.

Uluslararası Eğitim Başarılarını Değerlendirme Derneğine (IEA) göre TIMSS, genel olarak fenden çok uluslararası bir matematik değerlendirmesi olarak kabul görmektedir ve dördüncü ve sekizinci sınıflarda 1995'ten bu yana her dört yılda bir yapılmaktadır (Mullis, Martin, Foy ve Arora, 2012).

Ülkeler 4. ve 8. sınıf düzeyinde uygulanan sınava sadece 4. sınıf veya sadece 8. sınıf düzeyinde katılma hakkına sahip olmasının yanında, her iki düzeyde de katılabilme şansına sahiptir. Ayrıca TIMSS'in dördüncü sınıf öğrencileri için çok zor olabileceği ülkelerde TIMSS, beşinci veya altıncı sınıftaki öğrencilere ve sekizinci sınıf öğrencileri yerine dokuzuncu sınıfta da benzer biçimde verilebilir. Yani buradan anlaşıldığı üzere TIMSS bazı ülkelerde öğrencilere oldukça zor gelmektedir ve bu yüzden bilişsel farkları gözeterek 4., 5., 6. sınıf ve 8., 9. sınıflar arasında öğrencilerin bilişsel düzeylerine bağlı olarak farklı sınıf düzeyinde sınava katılabilmektedirler. Öğrencilerin bulunduğu okul ve sınıflar büyük bir titizlikle her bir ülkeyi temsil etmesi sağlanacak şekilde seçilmektedir. Toplam 28 blok şeklinde ayrılan parçalardan 14 bloğu matematik ve geriye kalan 14 bloğu fen bilimleridir. PISA'nın

matematik ve fen okuryazarlığını ölçtüğü bir sınav olmasının aksine TIMSS'in genel olarak matematik ve fen başarısını ölçtüğü bir sınav olması sebebi ile her bir bloktaki sorular 4. ve 8. sınıf seviyesinde ve bilgiyi anlamadan çok bilme ve uygulama basamakları yoğun bir şekilde işlenen bir sınav olması ile tanınır.

TIMSS şablonu bir bütün olarak değerlendirilirse, her ülkenin farklı eğitim sistemi, farklı kültürü, farklı demografik yapısı, farklı düşünce tarzı ve farklı eğitim sistemi olduğu aşikârdır. Bir ülke eğitimde ahlâğa ve çalışkanlığa önem verirken, başka bir ülke bilgiyi bilmeye, daha başka bir ülke ise bilgiyi yorumlama ve anlamlandırmaya ve daha birçok ülke birçok kritere göre eğitimine o alanda önem vermektedir. Mesela bu kriterleri ülkeler ile ilişkilendirildiğinde, eğitimde ahlak ve çalışkanlığı genel olarak Japonya'nın yansıttığını, eğitimde bilgiyi bilmeye Türkiye'nin önem verdiğini ve verilen bilgiyi yorumlama ve anlamlandırmaya da Finlandiya'nın önem verdiği söylenebilir. TIMSS uygulamasına katılan 4. ve 8. sınıf öğrencilerinin bulunduğu okullarda bu sınavın nasıl bir süreçten geçtiği, her ülkenin eğitim sisteminde nasıl ele alındığı, okul uygulamalarındaki benzer ve farklı yönleri, ne gibi dönütler aldığı, ne kadar verimle yordama yapabildiği vb. bilgileri TIMSS havuzunda karşılaştırması bilgi düzeyinde toplanır ve bu bilgiler değerlendirilerek sınav sonu raporu hazırlanır. Tüm bu bilgiler TIMSS'in amacını ifade etmektedir ve matematik ve fen bilimleri başarısını yordayan bu proje belirtilen karşılaştırmalı bilgilerden yola çıkarak her bir ülkenin eğitim sisteminde bulunan eksiklikleri, farklılıkları ve olumlu yanları analiz ederek, öğrencilerin bu sınavlardaki başarılarının hangi kritere daha yaklaştığını göstermektedir. Bununla birlikte bu proje uygulamasında öğrencilere çeşitli anketler ve başarı testleri uygulanmış, bu anket ve başarı testlerine bağlı olarak, okulların yönetim, eğitim-öğretim düzeyi şeması, öğrencilerin karakteristik özellik ve bilgi düzeyi şeması, eğitim ve öğretim programları, bu programlara dayalı olarak işlenen ders kazanımları, sınava katılan öğrencilerin matematik ve fen başarıları ölçülmüştür. Bu bilgiler ışığında tüm veriler TIMSS havuzunda toplanarak her bir ülkenin eğitim durumuna değinilmiştir. (EARGED, 2014; TIMSS, 2015.).

Bu projenin bir diğere amacı, matematik ve fen alanında öğretilimi ve öğrenmeyi geliřtirmek için kullanılabilircek önemli bir arka plan bilgisi sağlamaktır. Örneğinin TIMSS, müfredat ve müfredat uygulaması, öğretim uygulamaları ve okul kaynakları hakkında ayrıntılı bilgi toplar. Bu amaç doğrultusunda başarı testleri ve çeşitli anketler kullanılarak öğrencilerin fen ve matematik alanındaki performansları, eğitim sistemleri, öğretim programları, öğrenci özellikleri, öğretmen ve okulların karakteristik özellikleri ile ilgili bilgiler toplanmaktadır. TIMSS, düzenli aralıklarla başarıdaki değişiklikleri izlediğinden, yeni veya gözden geçirilmiş eğitim politikalarının başarıyı etkileyip etkilemediğini arařtırmak için değerli bir araçtır.

TIMSS 2015 uygulaması, eğitim sistemindeki performansların daha iyi anlaşılması ve bu konudaki uzmanlara yön vermesi amacıyla yapılmakta olup, matematik ve fen alanında iki ayrı eğitim seviyesindeki (4. ve 8. sınıf) öğrencilerin bilgilerini karşılařtırmaktadır. Bu karşılařtırmalar yapılırken bir takım veriler de toplanır. Bunlar: eğitim sistemi, idareciler, öğretmen, okul bilgileri ve öğrenci anketleri gibi verilerdir. Böylece, matematik ile fen başarısında eğitim sisteminin nasıl işlediğii konusunda bilgiler sağlanırken, eğitimde yapılacak reform ve gelişmeler için de değerli birer kaynak oluşturulmaktadır (Mullis, Martin, Foy, ve Arora, 2012).

Böylesine detaylı olan TIMSS uygulamasının merkezini Boston College ve TIMSS, PIRLS Uluslararası Çalışma Merkezi oluşturur. Bu birimler tarafından ele alınan her ülkeye dair matematik ve fen başarısı, anketler vb. datalar bu birimler bünyesinde depolanır ve yorumlanır. Ayrıca bu ulusal kurum uluslar arası kurumlarla bağlantı halinde olup gerekli olan verileri işbirliğı içerisinde devam ettirmektedir. Türkiye’de TIMSS projesinin başında yer alan Milli Eğitim Bakanlığına bağılı Ölçme ve Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü (ÖDSGM) gelmektedir. Bu genel müdürlük Türkiye’de bulunan lise giriş sınavından, üniversite giriş sınavına, içişleri bakanlığı sınavlarına kadar daha birçok alanda sınavı bünyesinde duyurmakta ve yapmaktadır. Her bir dersten kazanım kavrama testleri ve değerlendirme soruları yayımlayarak öğrencilerin önem vermesi gereken kazanımlara ve müfredata vurgu yapmaktadır. Tüm bu bilgiler doğrultusunda bu birim eğitimin temel birimini oluşturmaktadır.

Uluslararası sınavlar kadar bu sınavların ölçme kriterlerine uygunluğu da çok önemlidir. Kùltürler arası karşılaştırmaların geçerliliđi, psikoloji ve eğitimle ilgili araştırmalardaki birçok uygulama için öneme sahiptir. Kùltürler arası araştırmanın geçerlilik kanıtı farklı ùlkelerden elde edilen test puanlarının aynı metriđe ait aynı yapıyı ölçmesidir (Wu, Li and Zumbo, 2007). TIMSS ùlkelerin fen ve matematik başarılarını tek bir ölçme aracıyla deđerlendirip karşılaştıran bir uygulamadır. Yapılan karşılaştırmaların anlamlılıđından bahsedebilmek için kullanılan ölçme aracının farklı gruplarda aynı özelliđi ölçtüđünün dođrulanması gerekmektedir (Uzun ve Öğretmen, 2010).

Klasik Test Kuramı'nda güvenilirlik ve geçerlilik, ölçme araçlarının aksine ölçme sonuçlarına göre deđerlendirilip sonuçlandırılır. Bundan dolayı geçerlilik ve güvenilirlik sonuçları, madde ve test istatistikleri grubun yapısına bađlı olarak hesaplandıđından her grubun yapısına göre farklı farklı sonuçlar vermektedir (Crocker ve Algina, 1986). Klasik Test Kuramı'ndaki bu sınırlılık ölçme sonuçlarının ne tarz gruplarda ve durumlarda geçerli ve genellenebilir olduđu sorununu dođurmuştur.

Vandenberg ve Lance'a (2000) göre, bu sınırlılıđa bađlı olarak araştırmacıların elde ettiđi ölçme sonuçlarını yorumlayabilme ve karşılaştırabilmeleri ile ilgili aşıđıdaki sorular ortaya çıkmaktadır.

- 1) Farklı kùltürlere sahip bireylere ait bir ölçme sonucu benzer biçiminde yorumlanabilir mi?
- 2) Derecelendirme kaynakları hedef grubun performansını derecelendirirken aynı performans tanımını mı kullanmaktadırlar?
- 3) Test araçlardaki maddelerde cinsiyet, etnik yapı gibi farklılıklar var mıdır?
- 4) Önemli olan bir durum veya süreç zamanla belirli bir ölçme aracındaki maddelere verilen tepkiyi ve kavramsal yapısını deđiştirmekte midir?

Bu sorulardan yola çıkarak, ölçmenin eşdeđerliđi yani diđer bir deyişle ölçme deđişmezliđi çalışmalarında yukarıdaki varsayımlar dikkate alınarak araştırma sürecine yön verilmektedir.

Bu bilgilerden hareketle biraz da ölçme değişmezliğinin ne zamandan beri popülerleşmeye başladığından bahsedilirse, son zamanlarda ölçme değişmezliğinin sosyal bilimlerde giderek daha önemli ve tartışmalı bir konu haline gelmesinden başlamak gerekir. Örneğin 2014 yılı Nisan ayında “ölçme değişmezliği” terimi bir Google Akademik araştırmasında yaklaşık 239.000 sonuç vermiştir. Bu bilimsel makalelerin bolluğu ise üç kategoriye ayrılır. İlk kategori, ölçme değişmezliğine ilişkin çeşitli yöntem ve yaklaşımları tanıtan, tartışan ve değerlendiren metodolojik makalelerin bulunduğu bir bölümdür. İkinci kategori, daha fazla karşılaştırmalı analiz için ön koşul olarak verilen bir temelin ölçme değişmezliğini test eden makaleleri içermektedir. Bu makaleler ölçme değişmezliğini, temel hipotezlerin anlamlı bir şekilde test edilmesini sağlayan bir ön analiz olarak değerlendirmektedir. Üçüncü kategori ise, belirli gizli yapıların ölçülmesi için geliştirilen belirli ölçeklerin ölçme değişmezliği özelliklerini bildirir. Ayrıca bir çalışmada ölçme değişmezliği oluşturmak, bir anketin her zaman ölçme değişmezliği olduğu anlamına gelmez. Ölçme değişmezliği gruplar arasında tekrar tekrar test edilmelidir çünkü değişkenlik, cihazın özelliklerine ek olarak çalışmanın dış etkenlerinden de kaynaklanabilir (Cieciuch, Davidov, Schmidt, Algesheimer ve Schwartz, 2014).

Gruplar arasında ölçme değişmezliğinin sağlanması, çapraz grup karşılaştırmaları yapmak için mantıksal bir gerekliliktir ancak ölçme değişmezliği örgütsel araştırmalarda nadiren test edilmiştir (Vandenberg ve Lance, 2000). Mark ve Wan'a (2005) göre ölçme değişmezliğinin varlığının kanıtlanmadığı durumlarda yapılan çıkarımların bilimsel bir nitelik taşımasının azalması gibi, gruplar arası farklılıkları yorumlayabilmek giderek azalmaktadır.

Araştırmalarda kullanılan ölçme aracıyla yapılan analizler ışığında eşdeğer psikometrik özellikte kriterler taşıdığı varsayılmakla birlikte, farklı gruplarda belirli bir yapıyı ölçerken benzer biçimde çalışacağı da kabul edilmektedir. Farklı gruplar arasında karşılaştırma yapılabilmesi için öncelikle ölçmelerin değişmezliğinin incelenmesi gerekmektedir (Horn ve McArdle, 1992). Yani ölçme değişmezliği çalışması, ölçme modellerinin gruplar arasında ne ölçüde genellenebileceğini belirlemek için yapılır. Eğer bir ölçme aracı heterojen bir gruba uygulanmak için geliştirildiyse, testin ölçme özelliklerinin alt gruplarda da benzer olduğunun kanıtlanması gerekmektedir. Ölçme değişmezliği diğer bir ifadeyle ölçme eşdeğerliliği ölçme işlemi uygulanan evrendeki bireylere ait nitel ve nicel özelliklerle

ilgili olmaktan çok ölçme aracının kendisi ile ilgilidir. Ölçme değişmezliği çalışmalarında amaç yeni bir ölçme aracı geliştirmek değil, farklı gruplar arasında karşılaştırma yapabilmektir (Cheung ve Rensvold, 2000).

Bu doğrultuda araştırmanın problemini, TIMSS 2015 uygulamasındaki matematik başarısının OECD üyesi ülkeler için ölçme değişmezliğinin test edilmesi oluşturmaktadır.

1.2. Araştırmanın Amacı

Bu araştırmanın amacı, TIMSS 2015 matematik testinin sekizinci sınıf uygulamasının OECD üyesi ülkeler için ölçme değişmezliğinin test edilmesidir. Ve bu genel amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır.

1.2.1. Alt Problemler

- 1) TIMSS 2015 uygulaması sonucunda elde edilen veri seti için sekizinci sınıf matematik başarı testi, OECD ülkeleri arasında ölçme değişmezliğini sağlamakta mıdır?
- a) TIMSS 2015 uygulaması sonucunda elde edilen veri seti için sekizinci sınıf matematik başarı testi, OECD ülkeleri arasında şekil değişmezliği sağlamakta mıdır?
- b) TIMSS 2015 uygulaması sonucunda elde edilen veri seti için sekizinci sınıf matematik başarı testi, OECD ülkeleri arasında metrik değişmezliği sağlamakta mıdır?
- c) TIMSS 2015 uygulaması sonucunda elde edilen veri seti için sekizinci sınıf matematik başarı testi, OECD ülkeleri arasında ölçek değişmezliğini sağlamakta mıdır?
- d) TIMSS 2015 uygulaması sonucunda elde edilen veri seti için sekizinci sınıf matematik başarı testi, OECD ülkeleri arasında katı değişmezliği sağlamakta mıdır?

1.3. Araştırmanın Önemi

Belirli gruplar arası karşılaştırmaların yapıldığı birçok çalışmada araştırmacılar genellikle ölçme araçlarının, ölçme değişmezliğini/eşitliğini sağladığı varsayımına dayanarak söz konusu karşılaştırmaları yaparlar. Ancak bilişsel özelliklerin analiz edilip ölçülmesi sürecinde, en çok da farklı dil, kültür, cinsiyet ve geleneğe sahip gruplar için bilişsel özellikleri ölçen testlerin aynı sonucu verip vermediği, aynı anlama gelip gelmediği gibi sorulara yanıt bulmadan bu karşılaştırmaları yaparak sonuçlardan bazı çıkarımlar elde etmek yanlış bulgulara ve yorumlamalara sebep olabilir (Vandenberg ve Lance, 2000).

TIMSS ile ilgili yapılan çalışmalar incelendiğinde, duyuşsal özellikleri ölçen araçların ilişkin ölçme değişmezliğinin incelendiği literatür taraması sonucu ulaşılabilen çalışma sayısı oldukça fazladır. Buna bağlı olarak çalışmalarda daha çok cinsiyetler, bölgeler, gelenekler gibi değişkenler çalışılmış olup, bilişsel alanların ölçülmesinde kullanılan çoktan seçmeli ve açık uçlu testler ve öğrenci ölçekleri üzerinden yürütülen çalışmaların özellikle de fene oranla matematik alanında yapılan araştırmaların azlığı saptanmıştır. Ayrıca değişmezlik çalışmalarında OECD ülkelerinin dâhil edildiği çalışmalara daha az ulaşılmış olup OECD ülkeleri arası matematik başarı performanslarında farklılıklar tespit edilmiştir (MEB, 2015). Buradan yola çıkarak araştırmada farklı gruplarda farklı bireysel özelliklere sahip kişiler arasında ölçme değişmezliği çalışmasının yapılması, geniş ölçekli testler ve bu testlerin sonuçlarının doğru yorumlanabilmesi gerekmektedir.

1.4. Sınırlılıklar

Araştırma kapsamında:

1. TIMSS 2015 uygulamasında yer alan ülkeler arasından OECD ülkeleri (17 ülke) ile sınırlandırılmıştır.
2. TIMSS 2015 uygulamasında yer alan testlerden sekizinci sınıflara uygulanan matematik başarı testi ele alınmıştır.
3. TIMSS 2015 uygulamasında yer alan 11 no'lu kitapçıktaki matematik başarısına ilişkin 13'ü açık uçlu, 16'sı çoktan seçmeli olmak üzere 29 adet madde ele alınmıştır.

Araştırma sınırlılıkları genel olarak ifade edilirse, veri setindeki tüm ülkelerin 8. sınıf matematik başarıları ölçme değişmezliği sonucunu vermemesi sebebi ile 17 OECD üyesi ülke, 11 no'lu kitapçığı alan 7223 katılımcı ve 29 madde ile sınırlandırılmıştır.

1.5. Tanımlar

Ölçme Değişmezliği: Genel bir ifadeyle ölçme değişmezliği, bir durumdan etkilenmeden ölçmek istediğimiz özelliğin ya da kaçınmak istediğimiz durumun saf, yalın halinin elde edilmesi olarak ifade edilebilir.

Doğrulamalı Faktör Analizi: Önceden doğrulanmış ve tanımlanmış olan bir yapının bir model olarak kanıtlanmasında ve geçerlilik çalışmalarının yapılmasında kullanılan eşdeğer analiz türü pek bulunmayan literatürde çok kullanılan bir analiz türüdür.

1.6. Kısaltmalar

OECD: Ekonomik İşbirliği ve Kalkınma Örgütüdür. İngilizcesi 'Organisation for Economic Co-operation and Development' şeklinde ifade edilir.

TIMSS: Uluslararası Matematik ve Fen Çalışmalarında Yer Alan Eğilimlerdir. İngilizcesi 'Trends in International Mathematics and Science Study' şeklinde ifade edilir.

IEA: Uluslararası Eğitim Başarılarını Değerlendirme Birliğidir. İngilizcesi 'International Association for the Evaluation of Educational Achievement' şeklinde ifade edilir.

ÖDSGM: Ölçme ve Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğüdür. İngilizcesi 'The General Directorate of Measurement, Evaluation and Examination Services' şeklinde ifade edilir.

RMSEA: Yaklaşık hataların ortalama kareköküdür. İngilizcesi 'root mean square error of approximation' şeklinde ifade edilir.

CFI: Karşılaştırmalı uyum indeksidir. İngilizcesi 'comparative fit index' şeklinde ifade edilir.

NFI: Normlaştırılmış uyum indeksidir. İngilizcesi 'normed fit index' şeklinde ifade edilir.

NNFI: Normlaştırılmamış uyum indeksidir. İngilizcesi 'non-normed fit indeks' şeklinde ifade edilir.

GFI: Uyum iyiliği indeksidir. İngilizcesi 'goodness of fit' şeklinde ifade edilir.

AGFI: Düzeltilmiş uyum iyiliği indeksidir. İngilizcesi 'adjusted goodness of fit' şeklinde ifade edilebilir.

SRMR: Standardize edilmiş artık ortalamaların kareköküdür. İngilizcesi 'standardized root mean square residual' şeklinde ifade edilir.

Sd: Serbestlik derecesidir. İngilizcesi 'degree of freedom' şeklinde ifade edilir.

AFA: Açımlayıcı Faktör Analizidir. İngilizcesi 'Exploratory factor analysis' şeklinde ifade edilir.

DFA: Doğrulayıcı Faktör Analizidir. İngilizcesi 'Confirmatory Factor Analysis' şeklinde ifade edilir.

ÇGDFA: Çoklu Grup Doğrulayıcı Faktör Analizidir. İngilizcesi 'Multiple Group Confirmatory Factor Analysis' şeklinde ifade edilir.

YEM: Yapısal Eşitlik Modellemesidir. İngilizcesi 'Structural Equation Modeling' şeklinde ifade edilir.

BÖLÜM II

KAVRAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde geçerlilik, faktör analizi, hipotez testleri, ölçme değişmezliği ve ölçme değişmezliği ile ilgili alan yazın taraması sonucunda elde edilen araştırmalar şeffaflıkla açıklanacaktır.

2.1. Geçerlilik

Geçerlilik terimi için Büyüköztürk, Çakmak, Akgün, Karadeniz & Demirel (2008)'in fikirlerine bakıldığında testin ölçülmek istenilen ve amaçlanan özelliği ne kadar doğru ve iyi ölçtüğüyle ilgili olduğu şeklinde ifade etmişlerdir. Bir başka ifadeyle geçerlilik, ölçmeyi gerçekleştirebilme derecesidir. Anastasi(1968), Nunnally ve Bernstein (1994), Murphy ve Davidshofer'a (2005) göre geçerlilik ifadesi ölçme aracının bilimsel yararı olarak da anlatılabilir. LoBiondo-Wood ve Haber'e (2014) göre geçerlilik, bir ölçme aracının bir kavramın niteliklerini ne kadar doğru ölçtüğünün bir ölçütüdür ve bu yüzden çok önemlidir. Benzer şekilde Joppe (2000) geçerliliği, araştırmanın ölçmek istediği şeyi gerçekten ölçüp ölçmediğinin ve sonuçlarının doğruluğunun göstergesi olarak tanımlamıştır. Kelley'e (1927) göre bir test ancak ölçmeyi iddia ettiği özelliği ölçebiliyorsa geçerlidir o yüzden ölçmeyi iddia ettiği özellik bu bağlamda çok önemlidir. Cronbach (1984) geçerliliği, ölçme aracını geliştiren kişinin, elde ettiği test puanlarının yorumlamasında ve anlamlandırmasında kanıt topladığı bir süreç olarak ifade etmiştir. Messick (1989) ise geçerliliği, farklı değerlendirmeler ya da test sonuçlarını temel alan uygulama ve fiillerin, yeterliliği ve uygunluğunu destekleyen deneysel kanıtların ve kuramsal gerekçelerin derecesini belirleyen genel bir değerlendirme olarak tanımlamıştır. Aynı zamanda geçerlilik Şencan'a (2005) göre, ölçmeden elde edilen sonuçların evrene genellenebilir olmasıdır. Bu bağlamda, bir evreni temsil eden örneklem üzerinden yapılan çalışmalarda ve evren hakkında genellemelere ulaşma amacındaki araştırmalarda geçerliliğin sağlam kanıtlara dayandırılması zorunluluğu ve önemi ortaya çıkmaktadır.

En genel ve geçerli ifadeyi kullanan Sireci'ye (2007) göre geçerlilik ise:

- (i) testin tek bir özelliğine inmekten çok, belirli bir amaç için kullanılabilir olmasını ifade etmektedir,
- (ii) testin belirli bir amaca uygun olup olmadığının değerlendirebilmesi için farklı kaynaklardan edinilen kanıtların olması gerekir,
- (iii) bir testin belirli bir amaç için kullanılabilir olması bekleniyorsa, bu amaç için testin kullanımını savunacak yeterli ve gerekli kanıtlar ortaya koyulmalıdır,
- (iv) bir testin geçerliliğinin değerlendirilmesi sabit bir şey değil, devam eden bir süreçtir. Şeklinde dört madde ile genel olarak ifade etmiştir.

Ayrıca Sireci (1998), geçerliliğin yalnızca yapı geçerliliği çatısı altında kavramsallaştırılmasının, özellikle eğitim ortamlarında kullanılan testlerin amaca uygunluğunu değerlendirmede kapsam geçerliliğinin önemini zayıflattığını düşünmektedir. Tüm bu geçerlilik tanımlarından yola çıkarak en genel ifadeyle geçerlilik, bir ölçme aracının belirli bir yapıya ilişkin ölçülmesi amaçlanan özelliğin, başka etkenler dâhil olmadan ölçülebilir derecesidir. 1966 yılında Amerikan Psikoloji Birliği tarafından geçerlilik, kapsam geçerliliği, ölçüt dayanaklı geçerlilik ve yapı geçerliliği olmak üzere üç farklı kategoriye ayrılmıştır. Messick (1995) ise, ilk kez Cronbach ve Meehl (1955) tarafından ortaya atılmış olan tüm geçerlilik türlerinin yapı geçerliliğinin altında birleştirilmiş bir kavram olduğunu ve diğer geçerlilik türlerinin yapı geçerliliğine kanıt oluşturduğu iddiasını destekleyerek geçerliliğin alt dallarını birleştirerek geçerlilik terimine yeni bir boyut kazandırmıştır. Tüm bu kanıtlardan yola çıkarak bu bağlamda ölçülecek yapılara ilişkin geliştirilen ölçme araçlarının yapı geçerliliğine ilişkin kanıt ortaya koymak için genellikle aşağıda verilen yöntemler kullanılmaktadır:

- a) Gruplar Arası Farklılaşma: Crocker ve Algina'ya (1986) göre gruplar arası farklılaşma, farklı gruplardan elde edilen puan ortalamalarının test edilen hipotez doğrultusunda farklılaşmasının beklenmesidir. Böylece beklenen farkın bulunması durumunda ölçülen yapıya ait geçerlilik kanıtları elde edilmiş olur.

b) Var Olan Testlerle Ölçülen Yapı Arasındaki Korelasyonlar: Anastasi'ye (1968) göre var olan testlerle ölçülen yapı arasındaki korelasyonlar, yeni oluşturulan bir testin daha önceden mevcut ve geçerliliği kanıtlanmış benzer bir testle olan yüksek korelasyonu yine yapı geçerliliğine kanıt oluşturan durumlardandır, şeklinde ifade edilmiştir.

c) İç Tutarlılık: Anastasi'ye (1968) göre iç tutarlılık, temel özelliği ölçme aracının toplam puanının testin kendisine kriter oluşturmasıdır, şeklinde ifade etmiştir. Üst kriter grubu ile alt kriter grubunun toplam puanları temel alınarak karşılaştırma yapılmıştır ve üst gruptakiler alt gruptakileri anlamlı bir oranda geçtiyse maddeler geçerlidir biçiminde yorumlanmaktadır. Bu kanıtlama yöntemiyle ilgili diğer bir geçerlilik uygulaması ise alt testlere ait puanların toplam puanla olan ilişkileri saptanarak geçerliliğe kanıt oluşturulmasıdır.

d) Test Puanları Üzerinde Deneysel Değişkenlerin Etkisi: Anastasi'ye (1968) göre test puanları üzerinde deneysel değişkenlerin etkisinde, seçilen değişkenlerin test puanları üzerindeki etkisini araştıran deneylerle yapı geçerliliğinin sağlanmasına ilişkin kanıtlar elde edilir. Bu kanıtlar tamamen ölçülen alana ilişkin belirli bir testin uygulamaya elverişli olup olmamasıyla ilgilidir.

e) Yakınsak ve Ayırt Edici Geçerlilik: Campbell ve Fiske (1959) yakınsak ve ayırt edici geçerliliği, bu yöntemin yapı geçerliliğini değerlendirmede sık kullanılan yöntemlerden biri olmakla beraber, birçok özellik veya yapıyı ölçmek için birden fazla yöntem kullanılırsa bu ölçümler arasındaki korelasyonların çoklu özellik gösterdiği ve çoklu yöntem matrisi olacak şekilde (multitrait-multimethod matrix) dönüştüğünü söylemişlerdir. Bu yönetime bağlı olarak belirli bir yapıyı ölçen iyi bir testin:

- (i) test sonuçlarının aynı yapıya ait diğer ölçümlerle elde edilen çıktılarla tutarlı olması,
- (ii) testin sonuçlarının, ölçülen yapıyla kuramsal açıdan ilgisi olmayan ölçümlerle ilişkisiz sonuçlar vermesi,
- (iii) testle kullanılan ölçme yöntemlerinin yanlılık için kanıt oluşturması, çoklu özellik ve çoklu yöntem matrisi yönteminin ileri sürdüğü karakterlerdir.

f) Faktör Analizi: Ölçme araçlarının yapı geçerliliğine ilişkin güçlü kanıtlar elde edilmesine yardımcı olan ve alanyazında en çok kullanılan yöntemlerden biri olan faktör analizine, yapılan çalışmanın temelini oluşturması sebebi ile aşağıda oldukça detaylı biçimde yer verilmiştir.

2.2. Faktör Analizi

Faktör analizi, ölçeklerin yapı geçerliliği kanıtlarını ortaya koymak için araştırmacıların en sık başvurduğu tekniklerden biridir (Çokluk, Şekercioğlu ve Büyüköztürk, 2010). Ölçeklerin faktör yapısını ortaya koyan bu teknik diğer analizlere kıyasla daha geniş kapsamlı bilgi vermektedir. Faktör analizi, birbiriyle ilişkili bir diğer ifadeyle benzer yapıları ölçen değişkenleri bir araya getirerek gruplama işlemi yapan ve daha az sayıdaki bu gruplarla ölçme işlemini açıklamaya çalışan çok değişkenli bir istatistiksel yöntemdir (Büyüköztürk, 2002). Byrne (1998) veri analizinde faktör analizi kullanımını, altta yatan gizli yapılar hakkında bilgi toplamak ve gözlenen değişkenler arasındaki kovaryansı incelemek için kullanılan eski tekniklerden biri olarak tanımlamıştır. Faktör analizi, çok değişkenli bir veri kümesinin karmaşıklığını azaltmayı sağlayan bir teknik olup araştırmacılara uygulama yapma ve teori geliştirmede kolaylıklar sağlamaktadır (Huck, 2007). Crocker ve Algina'ya (1986) göre faktör analizinin üç temel amacı bulunmaktadır:

- (i) Bir modele ait bir dizi testin tüm çiftleri arasındaki korelasyonları için hesaplanması gereken ortak faktörlerin sayısına karar vermektedir.
- (ii) Test iç korelasyonları için hesaplanan ortak faktörlerin niteliğine karar vermektedir.
- (iii) Ortak faktör varyansı ile ilişkili bir gözlenen değişken için varyans oranını belirlemektedir.

Jöreskog ve Sörbom'a (1993) göre, faktör analizinde açımlayıcı (exploratory) ve doğrulayıcı (confirmatory) olmak üzere iki çeşit yaklaşım vardır. Açımlayıcı faktör analizinde veriler üzerinde kesin bir model oluşturmadan ölçme aracının karakteristik özelliklerini tespit etmek amaçlanırken, doğrulayıcı faktör analizinde göreceli olarak az sayıda parametre açısından deneysel verileri açıklamak ve tanımlamak için bir model oluşturulması amaçlanmaktadır. Hem açımlayıcı hem de doğrulayıcı faktör analizi bir grup gizil değişken ile çıktılar arasındaki gözlemlenen

ilişkiyi yeniden ortaya koymayı hedefler. Fakat açımlayıcı ve doğrulayıcı faktör analizleri temelde, faktör modeli üzerindeki bazı kısıtlamalar, giriş şartlarının sayısı ve niteliği konusunda farklılık gösterir (Brown, 2006).

2.2.1. Açımlayıcı Faktör Analizi (Exploratory Factor Analysis)

Psikolojik özelliklerin belirlenmesi için alanyazında çokça kullanılan çok değişkenli bir istatistik olan faktör analizinin tanımı ve asıl amacı Brown'a (2006) göre, belirli bir yapıya ait gözlenen değişkenlerle gizil değişkenlerin arasındaki ilişkiyi ortaya koymayı amaçlayan istatistiksel bir tekniktir ve bu teknik gözlenen değişkenler arasındaki kovaryans ya da korelasyonlara dayalı olarak örtük değişken ya da değişkenlerin sayı ve türlerini belirlemektir. Büyüköztürk (2002) ise faktör analizini birbirleriyle ilişkili ve aynı yapıyı oluşturan değişkenlerin bir araya gelerek kavramsal açıdan daha az sayıda değişkenler olarak tanımlanmasına yardımcı olan çok değişkenli istatistiksel bir yöntem olarak tanımlamıştır.

Açımlayıcı Faktör Analizi (AFA) sosyal bilimlerde ölçeklerin yapı geçerliliğini incelemede yaygın olarak kullanılan ve uygulanan istatistiksel bir tekniktir. Temelde faktör analizi maddelerin faktör adı verilen yapılar altında gösterdikleri ilişkiye dair bilgi verir (Çokluk, Şekercioğlu ve Büyüköztürk, 2010). Bu faktör analizi gözlenen değişkenlerle gizil değişkenler arasındaki bağlantıların bilinmediği durumlar için tasarlanmıştır. Analizin amacı, gözlenen değişkenlerin altında yatan faktörlerle nasıl ve ne derecede ilişkili olduğunu ortaya koyma ve gözlenen değişkenlerin ortaklaştığı asgari faktör sayısını belirlemektir (Byrne, 1998). Açımlayıcı faktör analizi bir grup değişken arasındaki ortak kovaryansı daha anlaşılır kılar çünkü faktörler ölçülen değişkenlerden daha az sayıdadır (Brown, 2006). Keşfedici faktör analizi olarak da adlandırılan bu yöntem araştırmanın başında karşımıza çıkan, gözlenen değişkenlerin belirli yapılar altında toplanıp toplanmadığı, kaç adet yapının ortaya çıkacağı ve bu yapıların niteliklerinin nasıl olacağı gibi sorulara yanıtlar bulmamıza yardımcı olmaktadır.

Açımlayıcı faktör analizinin iki genel kullanım alanı vardır: Bunlardan birincisi psikolojik yapıların ölçülmesinde açıklama ve ikincisi psikolojik yapıların ölçülmesinde veri indirgemedir. Bunlardan birinci kullanım alanı olan açıklama, belirli bir ölçme aracı ile belirli bir alana ilişkin birçok boyutun belirlenmesini ve gizil değişkenlerin keşfedilmesini sağlar. Bu amaca ulaşmak için, ölçülen değişkenler

arasındaki korelasyon ya da kovaryanslar matrisini kullanır. İkinci kullanım alanı olan veri indirgemedede ise ölçülen değişkenler kümesini özetlenmiş indeksler olarak birleştirir ve devamı da temel bileşenler analizi kullanılarak yapılır (Floyd ve Widaman, 1995). Ayrıca açımlayıcı faktör analizinin bazı önemli ve temel adımları bulunmaktadır. Bu adımlar Tabachnick ve Fidell'e (2013) göre aşağıdaki şekilde sıralanabilir:

- a) bir değişken setini seçme ve bu seti ölçme adımı,
- b) seçilen veri setine ait korelasyon matrisini hazırlama adımı,
- c) hazırlanan korelasyon matrisinden belirli faktörleri çıkarma adımı,
- d) faktör sayısını tespit etme adımı,
- e) analiz sonuçlarının yorumlanmasını kolaylaştırmak için faktörlere döndürme işlemi uygulama adımı,
- f) ve son olarak da analiz sonucunda elde edilen verileri çözümlenme ve yorumlama adımı şeklindedir.

2.2.2. Doğrulayıcı Faktör Analizi (Confirmatory Factor Analysis)

Diğer analiz türlerinden farklı olarak, yapısal eşitlik modelinin özel türü olan Doğrulayıcı Faktör Analizi (DFA), gözlenen değişkenler ile gizil değişkenler arasındaki ilişkileri test eden kapsamlı ve nedensel bir analizdir (Yılmaz, 2004). Bu doğrultuda, Doğrulayıcı Faktör Analizi (DFA), gözlemlenen bir kümenin faktör yapısını doğrulamak için kullanılan istatistiksel bir tekniktir. Ayrıca, gözlenen değişkenler ile altta yatan gizli yapılar arasında ilişki olduğuna dair hipotezin test edilmesini sağlar (Suhr, 2006). Jöreskog ve Sörbom'un düşünce yapısına göre (1993), Açımlayıcı Faktör Analizi (AFA) veride açıkça görülemeyen değişkenler arasındaki korelasyonların bile görülmesini sağlar, bu yönüyle bir keşif analizidir. Diğer yandan, Doğrulayıcı Faktör Analizi (DFA) teoride olması beklenen ilişkilerin gerçekte de verilerde görünüp görünmediğini test etmeyi sağlar. Doğrulayıcı Faktör Analizi, modelin tüm özelliklerinin önceden bilinmesini gerektiğinden Açımlayıcı Faktör Analizinden farklıdır, dolayısıyla faktör modelinin değerlendirilmesi için güçlü bir ampirik temel gereklidir. Bu nedenle, çoğunlukla Açımlayıcı Faktör Analizi uygulaması yapı geçerliliği doğrulaması sürecinde önceki aşamalarda kullanılır, doğrulayıcı faktör analizi ise daha önce kurulan deneysel ve teorik temeller üzerine uygulanır (Brown, 2006). Doğrulayıcı Faktör Analizinde, güçlü

deneysel alt yapıya sahip olan, önceden yapılandırılmış olan modeldeki faktör yükleri, faktör sayıları, faktör sıralamaları gibi özelliklere bakılır. Bu özelliklerden yola çıkılarak tasarlanan modelin yapı geçerliliği de böylece test edilmiş olur (Maruyama, 1998).

Kline'in düşünce yapısına göre (2005) Doğrulayıcı Faktör Analizi aşağıdaki koşulları ve adımları içermelidir. Doğrulayıcı faktör analizi:

- a) modelin betimlenmesi,
- b) modelin tanımlanabilmesi için uygun olan koşulların belirlenmesi,
- c) modelde gösterilen değişkenlerden yola çıkarak uygun verilerin toplanması,
- d) modelin analizi,
- e) veri seti ve modelin uyum düzeylerinin değerlendirilmesi,
- f) modelin iyi uyum göstermediği durumlarda modelin yeniden tanımlanması ve
- g) tartışma ve yorumlama adımlarını içerir.

Yukarıdaki adımlardan hareketle Brown'a (2006) göre doğrulayıcı faktör analizinin en yaygın kullanım alanlarından ve aşamalarından bazıları şunlardır:

- a) Test Araçlarının Değerlendirilmesi: Genellikle doğrulayıcı faktör analizinin kullanım alanı ölçme aracının gizli, saklanmış özelliklerini araştırmak ve incelemek içindir, şeklinde ifade edilebilir. Ayrıca testin nasıl puanlanacağına da yardımcı olmaktadır. Tek faktörlü yapının yanı sıra çok faktörlü yani iki ya da daha fazla faktörlü yapılarda alt ölçeklerin nasıl puanlanması gerektiğini gösterir. Doğrulayıcı faktör analizi ölçek geçerliliği hakkında da bilgi vererek bu geçerliliğin psikometrik değerlendirmenin birçok alanında kullanılabilmesini sağlamaktadır.
- b) Yapı Doğrulama: Doğrulayıcı faktör analizi sosyal bilimlerde yapı doğrulama için vazgeçilmez bir tekniktir. Analiz, yapı geçerliliğini değerlendirmek için kullanılan yöntemlerden biri olarak kabul edilen birleşim (convergent) ve ayrışım (discriminant) geçerliliği adına sağlam kanıtlar sağlayabilmektedir. Yapı doğrulamada doğrulayıcı faktör analizinin temel önemi, ölçüm hatalarının geçerlilik sonuçlarında hesaba katılmasıdır. Bu yönüyle

doğrulamalı faktör analizi geleneksel yöntemlerden daha güçlü bir kanıt sağlar.

c) Yöntem Etkileri: Çoğu zaman, gözlenen ölçümlerin bazı kovaryansları gizil değişkenlerden başka kaynaklar sebebiyle ortaya çıkmaktadır. Ölçme yaklaşımları ile göstergeler arasında ilave bazı değişkenler görüldüğü zaman yöntem etkisinden bahsedilebilir. Yöntem etkisi genellikle olumlu ve olumsuz olarak ifade edilen cümlelerin kombinasyonlarını içeren anketlerde görülmektedir. Açımlayıcı faktör analizi, yöntem etkilerini ortaya çıkarmada yetersiz kalırken, doğrulamalı faktör analizi, yöntem etkilerini ölçme modeline ait hata teorisinin bir parçası olarak ele alır.

d) Ölçme Değişmezliği Değerlendirmesi: Ölçme değişmezliği test geliştirmenin ve uygulamanın önemli bir yönüdür. Ölçüm özelliklerinin değişmezliğinin, cinsiyet, ırk vs. gibi etkenler için tespit edilmesi gereklidir. Değişmezliğin tespit edilememesi durumunda o testin yanlılığından şüphelenilir. Bu tip sorunlar çoklu grup doğrulamalı faktör analizinde ele alınabilir. Bu analizde çeşitli alt gruplarda ölçme modeli aynı anda tahminlenebilmektedir.

Tüm bu bilgilerden hareketle, betimlenen modele ilişkin uyumun değerlendirilmesinde bazı uyum indekslerinden faydalanılır. Bu uyum indeksleri: (i) mutlak uyum indeksleri ve (ii) artan (fazlalık) uyum indeksleri olmak üzere ikiye ayrılmaktadır. Mutlak uyum indeksleri McDonald ve Ho'ya (2002) göre önsel modelin, örneklem modele ne kadar iyi biçimde uyduğunu belirleyerek önerilen modellerden en üst düzey uyuma sahip olanı gösterir. Bu ölçümler önerilen teorisinin veriye ne kadar iyi uyduğunu ve verinin en temel göstergesini elde etmeyi sağlar. Artan uyum indekslerinden farklı olarak mutlak uyum indekslerinin hesaplanması temel (baseline) bir modelle karşılaştırmaya dayandırılmaz. Bunun yerine, hiçbir modelle karşılaştırılmadan modelin ne kadar uyumlu olduğu ölçülür (Jöreskog ve Sörbom, 1993). Aynı zamanda karşılaştırmalı (Miles ve Shevlin, 2007) veya göreceli (bağıl/nispi) uyum indeksleri olarak da bilinen artan uyum indeksleri ise temel formunda ki-kare modelini kullanmayan ancak ki-kare değerini temel model ile kıyaslayan bir grup indeks olarak belirtilir (McDonald ve Ho, 2002).

Araştırmada kullanılacak uyum indekslerine bakacak olursak Ki-Kare uyum iyiliği testi, GFI, AGFI, RMR, SRMR, NFI, CFI ve RMSEA genel olarak kullanılmaktadır. Aşağıda bu indekslerin daha detaylı açıklamalarına yer verilmiştir:

Ki-kare değeri (χ^2): Ki-kare değeri genel olarak model uyumunun genelini değerlendirmek için kullanılan geleneksel bir ölçümdür ve örneklem ile uyumlu kovaryans matrisleri arasındaki uyumsuzluk kısmını inceler (Hu ve Bentler, 1999: 2). İyi bir model uyumu .05 düzeyinde sonucu manidar değildir (Barrett, 2007). Böylece ki-kare istatistik testi literatürde çoğu zaman “kötülük uyumu” (Kline, 2005) veya “uyum eksikliği” (Mulaik ve diğerleri, 1989) şeklinde tanımlanır.

Örneklem büyüklüğüne göre ki-kare değerine bakıldığında ki-kare testi uyum istatistiğinin kullanımına ilişkin bazı kısıtlamalar mevcuttur. Bu kısıtlamalardan ilki: Küçük örneklem kullanıldığında ki-kare istatistiği uyumu düşüren bir etki yarattığından, iyi uyumlu modeller ile zayıf uyumlu modeller arasında ayrımı bazen yapamadığı gözlemlenmiştir.(Kenny ve McCoach, 2003). Ki-kare istatistiğinin bu durumu ve küçük örneklem için kısıtlayıcı olması nedeniyle model uyumlarını değerlendirmek için araştırmacılar tarafından alternatif indeksler önerilmiştir. Wheaton, Muthen Alwin ve Summers (1977) tarafından önerilen örneklem büyüklüğünün etkisini azaltan normlaştırılmış ki-kare istatistiği (χ^2/sd) bulunmaktadır ve bu değer kabul edilebilir düzeyinin en fazla 5, en az 2 (Tabachnick ve Fidell, 2013) kabul edilebileceği belirtilmiştir. Kısıtlamalardan ikincisi ise: Çok değişkenli normallik varsayımdır ve normallikten kaynaklı yüksek sapmalar, model uygun biçimde ayarlansa bile modelin reddedilmesine yol açabilmektedir (McIntosh, 2006). Kısıtlamaların diğer bir sebebi de: Ki-kare istatistiği, temelde istatistiksel manidarlık testi olduğundan, örneklem büyüklüğüne karşı duyarlıdır. Bu anlamdan yola çıkarak ki-kare istatistiği büyük örneklem kullanıldığı şartlarda genellikle modeli reddeder (Bentler ve Bonnet, 1980; Jöreskog ve Sörbom, 1993).

RMSEA: Modelde en uygun değeri veren fakat modelin bilinmeyen parametre tahminleriyle birlikte evrenlerin kovaryans matrisine ne oranda uyduğunu gösterir (Byrne, 1998). RMSEA istatistiğinde uyum indeksini belirlemek için için .06'ya yakın bir kesme noktası iyi uyum olarak değerlendirilirken (Hu ve Bentler, 1999) .05

ile 0 arasındaki deęerler modelin mükemmel uyum gösterdiği anlamına gelir (Jöreskog ve Sörbom, 1993; Brown, 2006; Raykov ve Marcoulides, 2008).

GFI: Geliştirilmesi Jöreskog'la beraber bu alanda çalışan Sörbom tarafından yapılan GFI, tahmini evren kovaryansı ile açıklanan varyans oranını hesaplamakta yaygın olarak kullanılan bir uyum indeksidir (Tabachnick ve Fidell, 2013). GFI 0 ile 1 arasında deęerler alır ve örneklem büyüdükçe GFI deęeri artar (Bollen, 1990; Miles ve Shevlin, 1998). Genel olarak GFI için kesme noktası olarak .90 önerilmektedir ancak elde edilen sonuçlar simülasyon çalışmaları, faktör yükleri ve örneklem büyüklüğü küçük olduğunda .95 deęerindeki kesme noktasının daha uygun olduğunu göstermektedir (Miles ve Shevlin, 1998). Bu genel ifadeler ışığında, AGFI da örneklem büyüklüğünün artmasıyla birlikte artış gösterir ve 0 ile 1 arasında deęerler alır. AGFI, .90 ve üzerinde deęer alıyorsa bu modelin mükemmel uyum gösterdiğini ifade etmektedir.

SRMR ve RMR: Veri seti içerisindeki örnekleme bulunan kovaryans matrisi ve varsayımda bulunulan kovaryans modelinin artık deęerlerinin arasındaki farkın karekökü alınarak SRMR deęeri bulunur. Bir ölçek, deęişik düzeyde maddeler içeriyorsa (bazı maddeler 1 ile 5 arasında deęişirken, dięerleri 1 ile 7 arasında deęişiklik gösteriyorsa) SRMR'nin yorumlanması gittikçe zorlaşır (Kline, 2005). Çünkü SRMR deęer aralığı her bir göstergenin deęer ölçeğine baęlıdır Standardize edilmiş RMR (SRMR) bu sorunu çözdüğü için yorumlanması daha kolaydır. SRMR deęerlerine genel olarak bakacak olursak 0 ve 1 düzeylerinde deęiştiğini ve .05'ten küçük deęerler model uyumunun iyi olduğunu göstermektedir (Byrne, 2013). Buna ilaveten .08'e kadar olan SRMR deęerleri de kabul edilebilir düzeyde bulunmaktadır (Hu ve Bentler, 1999). SRMR'nin 0 deęerine yaklaşması uyumun iyileştiğini, 0 deęerini alması mükemmel uyumu işaret etmektedir ancak SRMR modelde çok sayıda parametre olduğunda ve büyük örneklem büyüklüklerine sahip modellerde dięer uyum indekslerinde olduğu gibi düşük deęer göstermeye eğilimlidir.

NFI: Yokluk hipoteziyle oluşturulmuş modelin χ^2 deęeriyle, normal modelin χ^2 deęerini, karşılaştırarak ölçülür ve NFI deęeri bulunur. Yokluk hipotezine dayalı oluşturan model, deęeri bulunan tüm deęişkenlerin korelasyonsuz olduğunu belirttiği için en kötü durumları açıklayıp gösterdiği söylenebilir. Bu istatistik sonucu için çıkan deęerler 0 ile 1 arasında yer alır. Bentler ve Bonnet (1980) .90'dan büyük

değere sahip modellerin iyi uyum sergilediğini söylemişlerdir. Küçük örneklemelerin kullanıldığı durumlarda, diğer istatistikler iyi uyum değerleri gösterirken NNFI zayıf uyum değeri gösterebilir (Bentler, 1990; Kline, 2005; Tabachnick ve Fidell, 2007). Önemli araştırmacılardan Bentler ve Hu (1999), NNFI için kesme noktası kriterinin .95 ve üzerinde olması gerektiği önerisinde bulunmuşlardır. Bu öneri genel olarak kabul görmüştür. Bu uyum indeksi örneklem büyüklüğüne karşı hassasiyet gösterir (Mulaik, James, Van Alstine, Bennet, Lind ve Stilwell, 1989; Bentler, 1990) ve bu sebepten ötürü NFI indeksine analiz yorumlamalarında tek başına güvenilmemesi gerekmektedir (Kline, 2005). Bu sorun, daha basit modelleri tercih eden Normlaştırılmamış Uyum İndeksi (NNFI, aynı zamanda Tucker-Lewis indeksi olarak da bilinir) ile iyileştirilmiştir ve günümüzde çokça kullanılmaktadır.

CFI: NFI'nın örneklem boyutu küçükken dahi iyi sonuç verdiği, örneklem boyutunu göz önünde bulundurduğu Byrne (1998) tarafından literatüre kazandırılmıştır. Ayrıca Tabachnick ve Fidell (2007), NFI'nın revize edilmiş halinin CFI olduğunu da belirtmiştir. NFI gibi CFI istatistiği de örneklem kovaryansını, yokluk hipotezine dayalı olarak oluşturulan model ile kıyaslar ve tüm örtük değişkenlerin korelasyonsuz olduğunu kabul eder. NFI'da olduğu gibi bu istatistik için de değerler 0 ve 1 kesme değerleri arasında değişkenlik gösterir ve değerler 1'e yaklaştıkça model uyumu artar yani model benzerlik kazanır. Literatüre bakıldığında yapılan çalışmalarda göre mükemmel uyum değeri sayılabilmesi için .95'den büyük olması gerektiği kabul edilmektedir (Hu ve Bentler, 1999). Günümüzde bu indeks örneklem boyutundan en az etkilenen uyum indekslerinden biri olduğundan en çok rapor edilen uyum indekslerinden biri olmuştur ve böylece tüm YEM programlarına dâhil edilmiştir ve en sık kullanılan uyum indeksi sayılabilir (Fan, Thompson, ve Wang, 1999).

Uyum İndekslerini Raporlaştırma: Yapılan bir çalışmada alanyazında en çok rastlanan ve genel olarak kullanılan uyum indekslerinin SRMR, GFI, CFI, RMSEA, NNFI ve NFI olduğu bulunmuştur (McDonald ve Ho, 2000). Örneklem büyüklüğünden etkilendiği için Ki-kare istatistiğinden yola çıkarak yapılan değerlendirmeler bu istatistiğin, serbestlik derecesi ve ona bağlı p değeriyle birlikte her zaman kullanılması gerektiği söylenmektedir (Kline, 2005; Hayduk, Cummings, Boadu, Pazderka-Robinson ve Boulianne, 2007). Bunun yanında, normal dağılımdan sapma gösteren ve örneklem büyüklüğünün çok olduğu koşullarda, Satorra-Bentler

istatistiği ($S-B_{\chi^2}$) normallik gösteren ve küçük örnekleme sahip durumlarda hesaplanan değerine yakın değerler üretir ve böylece bu bilgiler ışığında hesaplanır. $S-B_{\chi^2}$ yani Satorra-Bentler düzeltmesi farklı puan dağılımına ve örneklem büyüklüğüne ve sahip veri setlerinde, örneklem büyüklüğünün üzerindeki etkisini azaltmak için kullanılan güvenilir bir istatistiktir (Byrne, 2013).

Hu ve Bentler (1999) ise her zaman için SRMR kullanımını RMSEA, NNFI (TLI) veya CFI ile birlikte olması gerektiğini ifade etmişlerdir. Literatür tarandığında birçok araştırmacı araştırmalarında sonucu raporlaştırmada kesinlikle ki-kare istatistiğinin kullanılması gerekliliğine değinmiştir (Kline, 2005). Genel olarak bakılırsa, yapılan araştırmalara dayanarak çalışmalarda ki-kare istatistiğini, ki-kare serbestlik derecesini ve p değerini, RMSEA, SRMR, CFI ve PNFI gibi uyum indeksleri örneklem boyutu, modelin yanlış belirlenmesi ve parametre tahminlerinde bulunma bakış açısı açısından en az duyarlılığa sahip oldukları için modellerin değerlendirilmesinde kullanılabilir.

2.2.2.1. Çoklu Grup Doğrulayıcı Faktör Analizi

Doğrulayıcı faktör analizi, belirli bir teoriye dayalı olarak oluşturulan bir modelin ya da psikolojik bir yapının geçerliliğini kanıtlamak için yapılan bir analizdir. Aynı zamanda bir ölçme aracının alt grupları olan bir evrene uygulanması durumunda, söz konusu psikolojik yapının bu alt gruplarda da değişmez olduğunu kanıtlamak için yine doğrulayıcı faktör analizi şeması altında yer alan ve en çok kullanılan alt gruplarından biri olan çoklu grup doğrulayıcı faktör analizi yapılır. Çoklu grup doğrulayıcı faktör analizi en az iki grupta, eş zamanlı olarak yapılır ve bu analiz ölçme değişmezliğinin incelenerek psikolojik değişkenlerin farklı şemalar içi genellenmesini sağlar ve bu bağlamda çok da önemlidir. Çoklu-grup doğrulayıcı faktör analizinde oluşturulan ölçme modeline ait parametreler, analize dâhil edilen bütün gruplarda eş zamanlı olarak kestirilerek bu parametrelerin birbirlerinden anlamlı bir biçimde farklılaşıp farklılaşmadığı test edilir (Brown, 2006).

2.3. Hipotez Testleri

Hipotezin genel olarak ifadesi, parametre ile ilgili aklımıza gelen bir düşüncedir, şeklindedir. Parametre ile alakadar olan düşünceyi inceleyip, kullanıp

kullanmayacağımıza karar verebilmek için hipotez testi yapılır. Böylece hipotez testleri aracılığıyla örnekten elde edilen istatistikler hesaplanır ve anakütle değerleri hakkında karar verme aşamasına geçilir. Anakütle parametreleri hakkında karar verirken doğru ya da yanlış olması muhtemel yargılardan hareket edilir ve bu yol hipotez testlerinin daha güvenilir sonuçlar vermesini sağlar.

Ki-kare testi, gözlenen frekanslar ile beklenen frekansların farkının istatistiki açıdan anlamlı olup olmadığını gösterme temeline dayanır. Ki-kare testinde niteliksel olarak belirtilen verilere yer verilir. İlaveten ölçüm sonucunda çıkan sürekli değişkenler belli bir değerden az veya çok olarak kıyaslanarak ki-kare testi uygulanabilir. Veriler, oranlar veya yüzdelikler şeklinde ifade edilmişse testin uygulanması olanak dışıdır. Ki-kare testi, serbestlik derecesi (sd) ile belirtilir. Dağılımın ortalaması sd'ye ve varyans ise sd'nin iki katına eşittir. Ki-kare değerleri, sıfır ile artı sonsuz arasında değerler alır. Dağılım, küçük sd'lerinde basık olmasına rağmen sd arttıkça normal dağılıma yaklaşır. Ki-kare dağılımı, sürekli bir dağılımdır.

Ki-kare dağılımı, genellikle iki bağımsız niteliksel kriteri test etmek için kullanılır. Daha önce doğru olduğu ispatlanan veya ortak kabul görmüş H_0 hipotezi kabul (doğru) ise iki kriterin arasında manidar bir fark olmadığını, iddia edilen durum olan H_1 hipotezi kabul (doğru) ise iki kriterin arasında manidar bir fark

olduğunu ifade eder. H_0 'ın doğru olduğu ihtimali altında B'ler, $B_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$

ifadesiyle hesaplanır. B'ler, deneyde belli bir koşula göre gerçekleştirilmesi olası olan frekanslardır. İkinci sınıflama kriterinde kategorilere ayrılan G değerleri ile hesaplanan B'ler karşılaştırılır. Eğer B ile G arasındaki farklar küçük ise hesaplanacak olan ki-kare değeri küçük olacak ve H_0 red edilmeyecektir. Eğer çıkan farklar büyük ise kriterler arasında bağımsızlığı ifade eden H_0 red edilecektir.

Hesapla elde edilen ki-kare değeri (χ_{hes}^2), ilgili sd'de ki-kare tablosunda bulunan ki-kare değeri (χ_{tab}^2) ile karşılaştırılır. Eğer, $\chi_{hes}^2 \geq \chi_{tab}^2$ ise H_0 red edilecektir. Bunun tam tersi ise H_0 kabul edilecektir. χ_{tab}^2 değeri, saptanan yanılma olasılığı (α) ve sd'ye göre ki-kare tablolarından bulunur. Burada:

$$\chi^2_{hes} = \sum_{j=1}^c \sum_{i=1}^r \frac{(G_{ij} - B_{ij})^2}{B_{ij}} = \sum_{j=1}^c \sum_{i=1}^r \frac{G_{ij}^2}{B_{ij}} - n \quad \text{ve sd} = (r-1)(c-1) \text{ eşitlikleri ile verilir.}$$

Ki-kare testinde red bölgesi genellikle sağ kuyrukta ve tek taraflı olarak ele alınır. Ancak varyans için ki-kare testinde eğer çift taraflı veya sol kuyrukta ve tek taraflı H_1 söz konusu ise çift taraflı veya sol kuyrukta ve tek taraflı red bölgesi uygulanır. Varyansla ilgili aralık tahmininde çift taraflı ki-kare güven sınırlarına yer verilir (Güngör ve Bulut, 2008).

2.4. Ölçme Değişmezliği

50 yıldan uzun bir süre önce literatüre girmiş olan ölçme değişmezliği testinin önemi ancak değişmezlik testi için kullanılan istatistiksel teknikler daha çok son yıllarda ulaşılabilir hale geldiğinden araştırmacıların ölçme değişmezliğini test etmesi günümüzde daha kolay ulaşılabilir ve uygulanabilir hale gelmiştir (Putnick and Bornstein, 2016). Faktör yapılarının denklliğini araştıran ve yazan ilk araştırmacı Jöreskog (1971) olmuştur. Ölçme değişmezliği kavramı ise ilk olarak Byrne, Shavelson ve Muthe'n (1989) tarafından tanıtılmış ve daha sonra test edilmiştir.

Byrne ve Watkins'e (2003) göre ölçme değişmezliği, tüm gruplarda ölçme aracındaki maddelerin tamamıyla aynı biçimde algılanıp yorumlanması şeklinde tanımlamaktadırlar. Ölçme değişmezliğinin temelinde, karşılaştırma yapılan gruplar için yapılan ölçümlerin geçerliliğinin sağlanması yatar (Tyson, 2004). Herdman'a (1998) göre ölçme değişmezliği bir psikolojik ölçme aracının güvenilirlik ve yapı geçerliliği gibi psikometrik niteliklerinin farklı gruplardaki eşitliğinin biçimsel olarak değerlendirilmesi şeklinde ifade edilmektedir (Akt: Moraes ve Reichenheim, 2002). Ölçme değişmezliği, gizil bir yapıyı ortaya koymak için geliştirilmiş bir ölçeğin psikometrik bir özelliğidir. Aynı yapı ülkeler, kültürel birimler, zaman hususları veya ülkeler içindeki bölgeler gibi farklı gruplar arasında aynı şekilde ölçüldüğünde ölçümün değişmez olduğu kabul edilir (Horn and McArdle, 1992; Meredith, 1993; Vandenberg and Lance, 2000, Vandenberg, 2002; Millsap, 2011; Davidov ve diğerleri, 2014). Ölçme değişmezliği sağlandığında gizil değişken üzerinde yapılan gruplar arası karşılaştırmaların geçerliliği sağlanabilir; aynı zamanda gelecekte yapılacak olan çalışmalarla gizil değişken puanlarının oluşumu, belirleyicileri ve sonuçları da kıyaslanarak değerlendirilebilir (Schoot, Lugtig and Hox, 2012).

Alanyazında yapılan arařtırmalar incelendiğinde ölçme deęişmezlięi ile ilgili birçok çalıřmanın özellikle psikolojik yapıların kültürler arası genellenebilirlięini incelemek ve bu yapıların karşılaştırılabilirlięini belirlemek için yapıldıęı açıkça anlaşılmaktadır.

Kültürler arası karşılaştırma çalıřmalarının hem psikometrik hem de kavramsal bir tarafı vardır. Bu bağlamda ölçme deęişmezlięi incelenirken test yanlılıęı ve test eşitlięi gibi kavramsal ve yöntemsel durumlardan da bahsedilebilir (Ayvallı, 2016).

Test Yanlılıęı: Van de Vijver ve Tanzer'e (2004) göre test yanlılıęı, farklı kültürlerde yapılan uygulamalarda geçerlilięi tehdit eden bir unsur olmaktadır. Yanlılık bir özellik ya da yeteneğin altında yatan farklılıklardan deęil, belli bir yapının göstergeleri üzerindeki puan farklılıklarından meydana gelmektedir. Yanlılık yapı yanlılıęı, yöntem yanlılıęı ve madde yanlılıęı (deęişen madde fonksiyonu) olmak üzere üç çeşittir:

(i) Yapı yanlılıęı: Psikolojik yapıların farklı gruplarda farklı anlamlandırılmasından kaynaklanır. Ölçülen özelliğin psikolojik yapısının gruplar arasındaki farklılıęıyla ilgilidir.

(ii) Yöntem yanlılıęı: Bazı sorunların genel hali olarak ifade edilmiştir. Bu sorunlar ařaęıdaki gibi sıralanabilir:

a)örneklem arası karşılaştırma yapılamamasından kaynaklanan sorunlar,

b)ölçme aracının yapısal özellikleri, aracının karakteristik özellikleri ve aracı cevaplayanların cevaplama esnasında kültür kaynaklı nedenlerden dolayı farklı eğilimler göstermesinden kaynaklanan sorunlar,

c)uygulama sürecinin, uygulama kořullarının farklılık göstermesinden meydana gelen sorunlar şeklindedir.

(iii) Yöntemden kaynaklı yanlılıklar test sonuçlarını ve sonuçların yorumlanmasını oldukça etkiler. Madde yanlılıęı (deęişen madde fonksiyonu) testte yer alan maddelerin farklı gruplar arasında farklı anlamlara gelmesinden ve farklı biçimde yorumlanmasından kaynaklanır (Matsumoto ve Van de Vijver, 2012).

Yanlılık ve Değişmezlik: Ölçme yanlılığı ve ölçme değişmezliği terimleri birbirleriyle eş anlamlı olmamaları sebebi ile aralarındaki ayrımı yapabilmek önemlidir. Yanlılık gerçek parametrelerle tahmin edilen parametre arasındaki farktır. Yanlılık, yapı, yöntem ya da madde yanlılığı nedeniyle istenmeyen bir faktöre ait üretilen istenmeyen bir ölçme varyansı kaynağıdır (Van de Vijver ve Poortinga, 2005). Ölçme değişmezliği ise elde edilen puanların gruplar arası karşılaştırılabilirliğini ve değişkenler arasında maddelerin eşit bir biçimde katkısının olup olmadığını (özellikle faktör yük değerleri üzerinde) açıklamaya çalışır (Meredith, 1993).

Test Eşitliği: Test eşitliği farklı ölçme işlemleri sonucu elde edilen puanların farklı gruplarda farklı zamanlarda karşılaştırılabilir olmasını sağlayan bir kavramdır. Yani test sonuçlarının farklı zamanlarda ve farklı kültürlerde olsa dahi aynı anlama gelmesi demektir. Van de Vijver ve Leung (1997); yapısal eşitlik, ölçme birimi eşitliği ve tam puan eşdeğerliği olmak üzere test eşitliği kavramını üç gruba ayırmışlardır. Yapısal eşitlik, ölçülen özelliğin kültürler arasındaki teorik eşitliğinin sağlanabilmesiyle ilgilidir. Ölçme birimi eşitliği, ölçülen özelliğin eşit aralıklı ölçek ya da oran ölçeği gibi aynı ölçek düzeyinde ölçülmesi demektir. Tam puan eşdeğerliğinde ise ölçme aracı aynı ölçme birimine ve başlangıç noktasına sahiptir. Bu eşitliğe, en üst düzey eşitlikte denebilir. Bu eşitlik sağlandığında kültürler arasında karşılaştırma ve birçok istatistiksel analiz yapılabilir (Matsumoto ve Van de Vijver, 2012).

Tüm bunlar dikkate alınarak ölçmelerin kusursuz bir biçimde sağlanabilmesi yönündeki çalışmaların yapılması gerekir. Kurulan bir ölçme modeline ait yapının farklı gruplar için aynı olması, kullanılan ölçme aracındaki maddelerin faktör yüklerinin, faktör korelasyonlarının, hata varyanslarının aynı olması anlamına gelir (Bollen, 1989; Byrne, 1998; Jöreskog ve Sörbom, 1993). Ölçme değişmezliğinde deneysel olarak, gizil değişken puanları ile gözlenen puanları ilişkilendirerek bir ölçme modeli oluşturulur ve bu model üzerinden incelenebilir.

Genellikle bir tanesi üç ölçüm seviyesi değişmezliği arasında ayırım yapar; yapısal (faktör yükleri, faktör korelasyonları ve hata varyansları sabit), metrik (faktör yükleri serbest, faktör korelasyonları ve hata varyansları sabit) ve skaler (faktör yükleri ve hata varyansları serbest, faktör korelasyonları sabit) (Vandenberg ve

Lance, 2000). Metrik değişmezlik, gruplar arasında kovaryansların ve standartlaştırılmamış regresyon katsayılarının karşılaştırılması için yeterlidir. Bununla birlikte, gizli değerlerin gruplar arasında anlamlı karşılaştırması için skaler ölçme değişmezliği gereklidir. Bazı araştırmacılar ise kısmi (metrik veya skalar) ölçme değişmezliğinin anlamlı karşılaştırmalar için oldukça yeterli olduğu savını öne sürmüştür (Byrne ve diğerleri, 1989; Steenkamp ve Baumgartner, 1998).

Ölçme değişmezliği, çoklu grup doğrulayıcı faktör analizi ile dört adet hiyerarşik modelin test edilmesi ile ortaya konmaktadır. Bu modeller sırasıyla “yapısal değişmezlik (configural invariance)” (Model 1), “metrik değişmezlik (metric invariance = weak factorial invariance)” (Model 2), “güçlü faktöriyel değişmezlik (strong factorial invariance = scalar invariance)” (Model 3) ve “katı faktöriyel değişmezlik (strict factorial invariance)” (Model 4) olarak adlandırılmaktadır. Aşamalılık içeren bu dört modelin test edilmesi gereklidir (Meredith, 1993). Bu dört modele bakacak olursak Tablo 1’de isimleri sembolleri ve süreçleri şeklinde gösterilmiştir:

Tablo 1: Ölçme Değişmezliği Modelleri

Model	Hipotez	Hipotezin Adı	Sembolik İfade	Süreç
1	H_{Formu}	Yapısal Değişmezlik (Temel Model)	$\Lambda^{(1)}_{form} \therefore \Lambda^{(2)}_{form}$	Bu süreçte hata varyansları, faktör yükleri, faktör korelasyonları sabittir.
2	H_{Λ}	Zayıf Faktöriyel Değişmezlik (Metrik Değişmezlik)	$\Lambda^{(1)} \therefore \Lambda^{(2)}$	Bu süreçte faktör korelasyonları ve hata varyansları sabit ve faktör yükleri serbesttir.
3	H_{λ}	Güçlü Faktöriyel Değişmezlik (Scalar Değişmezlik)	$\Lambda^{(1)}_{ij} \therefore \Lambda^{(2)}_{ij}$	Bu süreçte faktör yükleri ve hata varyansları serbest ve faktör korelasyonları sabittir.
4	$H_{\Lambda, \Theta(\delta)}$	Katı Faktöriyel Değişmezlik (Artık Varyans Değişmezliği)	$\Theta^{(1)}_{\delta} \therefore \Theta^{(2)}_{\delta}$	Bu süreçte faktör yükleri ve faktör korelasyonları sabit ve hata varyansları serbesttir.

(Şekercioğlu, 2018).

2.4.1. Yapısal Değişmezlik (Configural Invariance)

Ölçme değişmezliğindeki en temel seviyedir. Temel model olarak adlandırılan bu modelde “Grupların faktör yapıları arasında farklılık yoktur.” hipotezi test edilir. Tüm gruplar için faktör korelasyonları, faktör yükleri ve hata varyansları eşittir. Böylece yapısal değişmezlik test edilirken faktör yükleri, faktör varyansları ve hata varyansları serbest bırakılır. Bu aşamada test edilen hipotez ölçme aracının faktör yapısının gruplar arasında değişmez olduğudur. Grupların aynı faktör yapısına sahip olup olmadıkları incelenir. Yapısal değişmezliğin sağlanması durumunda ölçülen yapının ve bu yapıya ait maddelerin gruplar arasında aynı anlama geldiği ifade edilir (Gregorich, 2006). Bu durum, aracının yapı geçerliliğinin sağlanması demektir (Salzberger, Sinkovics, ve Schlgelmich, 1999; Spini, 2003; Vandenberg ve Lance, 2000). Eğer tam tersi bir durumla karşılaşılır ve yapısal değişmezlik kanıtı elde edilemezse, bu durum gruplarda farklı yapıların ölçüldüğü anlamına geldiğinden (Wu, Li ve Zumbo, 2007), sonraki aşamalardaki hipotezlerin test edilmesinin anlamı olmayacaktır. Bu modelin iki önemli işlevi vardır:

- (i) parametreler tüm gruplar için aynı anda test edilir,
- (ii) yapısal değişmezlik modeli test uyumu için eşit başlangıç değeri üretilir (Byrne, 2008). Dolayısıyla diğer modellerden elde edilecek kriter test edilmiş olur.

2.4.2. Metrik Değişmezlik (Metric Invariance)

Meredith’e (1993) göre zayıf faktöriyel değişmezlik (weak factorial invariance) olarak da adlandırılan bu değişmezlik türünde, aşamalılık gereği yapısal değişmezliğin sağlanması durumunda bir sonraki değişmezlik türü olan metrik değişmezliğin testine geçilebilmektedir. Gözlenen puanları ilgili gizil değişkene bağlayan regresyon eğilimleri olarak adlandırılan faktör yüklerinin gruplar arasında değişmez olup olmadığı test edilir ve ayrıca metrik değişmezliğe ait hipotez kabul ediliyorsa, faktör varyans ve kovaryansları nicel karşılaştırmalarda uygulanır ve kullanılır (Gregorich, 2006). Bu durumda karşılaştırılan gruplar arasındaki ölçme birimlerinin eşit olduğu söylenmektedir (Salzberger ve diğerleri, 1999). İlk kez parametre sınırlamasına geçilen bu modelde sadece faktör yükleri serbesttir ve aynı zamanda önemli bir sabitlik olan faktör korelasyonları ve hata varyansları sabittir.

Aynı zamanda faktör yüklerinin eşit olması, maddelerin tüm gruplarda aynı biçimde yorumlanması ve herhangi bir yanlılık oluşturmaması biçiminde de yorumlanmaktadır (Knight ve Hill, 1998). Bu aşamada test edilen hipotez, gizil değişkene ait ölçme aracını oluşturan maddelere ait faktör yüklerinin gruplar arasında değişmez olduğudur. Metrik değişmezliğin sağlanması durumunda tahmini faktör varyanslarının ve kovaryanslarının grupsal karşılaştırmaları savunulabilir. Benzer bir deyişle bir madde metrik değişmezlik şartını sağlarsa, madde üzerindeki fark puanları ülkeler arasında anlamlı bir şekilde karşılaştırılabilir ve bu farklılıklar benzer yapıdaki ülkeler arası farklılıkların da göstergesidir (Steenkamp ve Baumgartner, 1998).

Ayrıca bu modelde, faktör yüklerinin eşitliği (λ), ($\Lambda^1 = \Lambda^2 = \dots = \Lambda^G$) tüm grup için test edilmiştir (Spini, 2003; Vandenberg and Lance, 2000).

Metrik değişmezliğin desteklenmemesi ortak grupların bir veya daha fazlasının veya maddelerin en azından bir alt kümesinin, gruplar arasında farklı anlamlara sahip olduğu anlamına gelmektedir (Gregorich, 2006).

Açıklanan bu değişmezlik aşağıdaki eşitlik:

$Mg \cong \hat{\tau}_g \hat{\tau}'_g + \Lambda (\hat{\alpha}_g \hat{\alpha}'_g + \Psi_g) \Lambda' + \Theta_{\varepsilon g} = M^g$ ile ifade edilmiştir (Widaman and Reise, 1997).

2.4.3. Güçlü Faktöriyel Değişmezlik (Strong Factorial Invariance)

Skaler değişmezlik ve ölçek değişmezliği olarak da adlandırılan ve ölçme değişmezliğinin güçlü bir düzeyi olan güçlü faktöriyel değişmezlik modelinde, bir ölçme aracına ait maddelerin faktör puanlarının sıfır olması durumunda elde edilen regresyon sabitinin (τ) gruplar arasında değişmez olduğu hipotezi test edilir ve aynı zamanda gizli değişkenlerle gözlenen değişkenlerin regresyon sabitinin gruplar arasında eşit olup olmadığı test edilmiştir. ($\tau^1 = \tau^2 = \dots = \tau^3$) (Schmitt and Kuljanin, 2008). Bu aşama örtük ortalamaların karşılaştırılabilmesi için önemlidir. Güçlü faktöriyel değişmezliğin sağlanması gözlenen puanların gizil puanlarla ilişkili olduğu anlamına gelir yani gizil yapıda aynı puana sahip olan bireyler hangi gruptan olursa olsun, gözlenen değişkende aynı puanı elde ederler (Milfont and Fischer, 2010).

Güçlü faktöriyel değişmezlik metrik değişmezlik modelini tanımlayanların üzerinde bir dizi kısıtlama içerir (Widaman and Reise, 1997). Yalnızca faktör yüklerinin eşitliğinin test edildiği metrik değişmezlik modeline kıyasla bu modelde regresyon sabitinin de gruplar arası eşitliği gerekmektedir. Metrik değişmezlik modeli; güçlü faktöriyel değişmezlik modeli için bir ön koşuldur ve faktör yükleri hata varyansları serbest, faktör korelasyonları sabittir. Faktör yükleri eşit olduğu gibi madde sabiti (örtük değişken yani metrik değişkenin sıfır olduğu sonuçlarda gözlenen değişkenin aldığı değer) kavramının da sabit olması gerekmektedir. Metrik değişmezlik türü sağlandığı zaman “Gözlenen değişkenlerdeki ortalama farklılıkları, örtük yapıların ortalamalarından kaynaklanmaktadır.” biçiminde yorum yapılabilir (Tucker, Özer, Lyubomirsk ve Boehm, 2006). Güçlü faktöriyel değişmezliğin sağlanmaması durumunda, örtük ortalamaların karşılaştırması belirsizdir (Cheng and Reinsvold, 2002). Gruplar arasında madde sabiti farklarının olup olmadığının tespit edilmiş olması çok güçlü bir geçerlilik kanıtı olarak kabul edilir (Vandenberg ve Lance, 2000). Gregorich’e (2006) göre güçlü faktöriyel değişmezliğin sağlanması durumunda, faktör ve gözlenen değişkenlerin ortalamaları karşılaştırılabilir.

Bu modelde, bir zayıf faktöriyel değişmezliğinde açıklanan ek kısıtlamalar dizisi vardır. Bu ek kısıtlamalar τ_g matrislerde gözlenen değişkenlerin kesişimlerine dahil edilir. Bu durumda, değişmezlik aşağıdaki eşitlik ile test edilir:

$$Mg \cong \tau \tau' + \Lambda (\alpha_g \alpha_g' + \Psi_g) \Lambda' + \Theta_{\varepsilon g} = M^{\wedge}_{g'} \quad (\text{Widaman and Reise, 1997}).$$

2.4.4. Katı Faktöriyel Değişmezlik (Strict Factorial Invariance)

Ölçme değişmezliğinin bu son modelinde test edilen hipotez, hata varyanslarının gruplar arasında $H_{\Lambda\phi}$ değişmez olduğudur ($\theta^1 = \theta^2 = \dots = \theta^G$). Katı faktöriyel değişmezlik modelinde, faktör yükleri ve regresyon sabitinin yanı sıra hata varyanslarının da eşitliği gerekmektedir ve hata varyansları serbesttir. Ayrıca bu modelde faktör korelasyonları ve faktör yükleri sabittir. Katı faktöriyel değişmezliğin sağlanması durumunda faktör ve gözlenen değişken ortalamalarıyla birlikte gözlenen varyans ve kovaryansların karşılaştırılması da mümkündür. Ayrıca ortalama ve varyans tahminlerinin gruplar arası karşılaştırmaları için katı faktöriyel değişmezlik kanıtı elde edilmelidir (Gregorich, 2006). Katı faktöriyel değişmezlik kanıtının elde edilmesi madde varyanslarının sabit olduğu anlamına gelir. Madde varyansları gözlenen değişkenler arasındaki korelasyonları etkilediğinden bu korelasyonları

karşılaştırılabilmesi için katı faktöriyel değişmezlik sağlanmalıdır (Tucker, Özer, Lyubomirsk ve Boehm, 2006). Ölçme değişmezliğine ilişkin aşamalı süreçte son aşamada yer alan bu model, en çok sınırlandırılmış modeldir. Ayrıca bu değişmezlik:

$$Mg \cong \tau \hat{\tau}' + \Lambda (\hat{\alpha}_g \hat{\alpha}_g' + \Psi_g) \Lambda' + \Theta_\varepsilon = M^g$$

Yukarıdaki eşitlik ile test edilmiştir. Ve ayrıca katı faktöriyel değişmezlikte; faktör yükleri, faktör yapısı ve madde sabitlerinin eşit olmasıyla birlikte madde artık varyanslarının da aynı zamanda eşit olması beklenir (Widaman and Reise, 1997).

2.5. Ölçme Değişmezliği Modellerinin Karşılaştırılması

Yapılan çalışmalar incelendiğinde ölçme değişmezliği çalışmalarında kurulan ölçme modellerinin karşılaştırılmasında farklı yaklaşımlar kullanılarak karşılaştırmalar yapıldığı görülmektedir. Gruplar arasında anlamlı karşılaştırmalar yapmak için ölçüm değişmezliği gereklidir. Jöreskog (1971) ve Bollen'e (1989) göre ölçme değişmezliği testi için en sık kullanılan prosedürlerden ya da yöntemlerden biri çoklu grup ya da çok gruplu doğrulayıcı faktör analizi - multigroup confirmatory factor analysis'dir (ÇGDFA - MGCFA). (Cieciuch, Davidov, Schmidt, Algesheimer ve Schwartz, 2014). Araştırmada çoklu grup doğrulayıcı faktör analizi (ÇGDFA) sonuçları modellerin aşamalı olarak karşılaştırıldığı yuvalanmış (nested) yöntem (Brown, 2006) kullanılarak karşılaştırılmıştır. Modeller sırasıyla önce Model 1 (yapısal değişmezlik) ve Model 2 (metrik değişmezlik), Model 2 ve Model 3 (güçlü faktöriyel değişmezlik) ve son olarak Model 3 ve Model 4 (katı faktöriyel değişmezlik) karşılaştırılarak sonuçlar değerlendirilmiş ve ölçme değişmezliğine ilişkin karar alınmıştır. Model uyumu karşılaştırılan modelin uyumuna eşit uyum göstermişse ya da daha kötü uyum göstermişse daha çok sınırlandırılmış ölçme modelinin doğrulandığı kabul edilmiştir (Van de Vijver ve Leung, 1997).

2.6. İlgili Araştırmalar

Bu bölümde ulusal ve uluslararası düzeyde uygulanan geniş ölçekli sınavlara (TIMSS, PISA vb.) ve ölçme değişmezliği kapsamında ele alınan ilgili çalışmalara yer verilmiştir.

Uzun ve Öğretmen (2010), TIMSS 1999 verileri üzerinde yürüttükleri çalışmalarında, öğrencilerin fen bilimleri başarısını belirleyen değişkenleri belirlemişler ve bu değişkenlerin cinsiyet bazında değişmezliğini incelemişlerdir. Özyeterlik, önem, tutum ve sınıf içi öğrenci etkinlikleri için aşamalı şekilde ölçme değişmezliği testleri yapılmış ve elde edilen uyum indeksleri sonucunda modelde yer alan değişkenlerin yalnızca metrik değişmezlik modeline kadar uyum gösterdikleri görülmüştür. Cinsiyet alanında oluşturulan grupların ise katı değişmezlik koşullarını sağlayamadıkları belirlenmiştir.

Ölçülüoğlu (2015), TIMSS 2011 Türkiye sekizinci sınıf matematik başarısını etkileyen faktörlerin bölgelere göre incelenmesini ele alan yüksek lisans çalışmasında, TIMSS 2011 verileri ile Türkiye'deki sekizinci sınıf öğrencilerinin matematik başarısını etkileyen değişkenleri modellemiş ve modelin coğrafi bölgelere göre ölçme değişmezliğini inceleyerek değişkenlerin bölgeler bakımından farklılık gösterip göstermediğini analiz etmiştir. Matematik başarısının yordandığı yapısal model, yedi coğrafi bölgede ayrı ayrı test edilmiş ve modelin bütün bölgelerde geçerli olduğu bulunmuştur. YEM ile kurulmuş model sonucunda duyuşsal özellikler değişkeninin matematik başarısını belirlemede en yüksek paya sahip olduğunu göstermiştir. Daha sonra modele alınan değişkenlerin coğrafi bölgeler arasında karşılaştırılabilmesi için matematik başarısına ilişkin kurulan modelin çoklu grup doğrulayıcı faktör analizi ile ölçme değişmezliği incelenmiştir. Modele alınan değişkenlerinin bölgelere ait regresyon katsayılarındaki farklılıklar incelendiğinde okul gizil değişkeninin İç Anadolu ve Güneydoğu Anadolu bölgesi hariç ülke genelindeki gibi matematik başarısını yordamadığı saptanmıştır.

Sarı, Arıkan ve Yıldızlı (2017), TIMSS 2015 sekizinci sınıf matematik akademik başarısını yordayan faktörler çalışmalarında, matematik başarısındaki farklılıklarda öğrenci ile ilgili değişkenlerin (öğrenci, öğretmen ve okul faktörlerinin) %34'ünün açıkladığını dile getirmişlerdir. Duyuşsal özellik olan öz-yeterlik inancı 8.sınıf öğrencilerinin matematik başarılarını yordamada en önemli değişken olduğu

ve diğ er önemli deđişkenin öğrencilerin evde sahip oldukları eğitimsel kaynakları olduğunu ifade etmiştir. Duyuşsal alan içerisinde yer alan tutum ile matematik başarısı arasında negatif yönde ilişki gözleendiğini söylemişlerdir. Araştırmada zorbalık, okula aidiyet ve öğretim etkinlikleri ile matematik başarısı arasında anlamlı bir ilişkinin olup bu ilişki matematik başarısını yordamada daha az öneme sahip olduğunu vurgulamışlardır. Tüm deđişkenler bir aradayken okullardaki 8. sınıflar arası başarı farkının %29'unu açıklanmaktadır. Öğrenmede çevre faktöründe yer alan öğrencilerden kaynaklı sorunlar ve okulda başarıya verilen önem öğrencilerin matematik başarısını yordamadaki en önemli iki deđişken olduğu sonucuna ulaşmışlardır.

Ayvallı (2016), PISA 2012 matematik okuryazarlığı testinin ölçme deđişmezliğinin incelenmesini ele aldığı yüksek lisans çalışmasında, matematik okuryazarlığı testinin Türkiye'deki cinsiyet, bölge grupları ve OECD ülkeleri arasında ölçme deđişmezliğinin sağlanıp sağlanmadığını çoklu grup doğrulayıcı faktör analizi aracılığıyla test etmiştir. Tüm gruplarda ayrı ayrı doğrulayıcı faktör analizi sağlaması yapıldıktan sonra kovaryans matris eşitliği de sağlanmış, ardından ölçme deđişmezliği testleri yapılmıştır. Yapılan analizler sonucunda Türkiye'de cinsiyet ve bölge grupları için ölçme deđişmezliğinin sağlandığı, OECD üyesi ülkeler için ise ölçme deđişmezliğinin sağlanmadığı görülmüştür.

Karakoç Alatl ı (2016), PISA 2012 okuryazarlık testlerinin ölçme deđişmezliğinin incelenmesi doktora çalışmasında, PISA 2012 matematik ve fen okuryazarlık testleri ile okuma becerileri alt testlerinin teste katılan Avustralya, Şangay-Çin, Türkiye ve Fransa grupları arasında belirli kriterlere göre ölçme deđişmezliğinin sağlanıp sağlanmadığını belirlemeyi amaçlamıştır. Dil deđişkenine göre incelenen çalışmada; Çoklu Grup Doğrulayıcı Faktör Analizi (ÇGDFA), Genelleştirilmiş Aşamalı Doğrusal Modelleme (GADM), Simultaneous Item Bias Test (SIBTEST) ve Madde Tepki Kuramı'na dayalı Olabilirlik Oran (MTK-OO) gibi farklı teknikleri kullanmıştır. Araştırmada ÇGDFA sonuçlarına göre, PISA 2012 matematik ve fen okuryazarlığı testleri ile okuma becerileri testlerinin belirlenen ülke örneklemi arasında dil deđişkenine göre ölçme deđişmezliğinin sağlanmadığını belirlemiştir. Ayrıca temel model olan yapısal deđişmezliğin sağlandığını ancak metrik deđişmezlik modelinin sağlanmadığını gözler önüne sermişlerdir.

Yandı, Köse ve Uysal (2017), PISA 2012’de bulunan veriler üzerinden farklı yöntemlerle ölçme değişmezliğinin araştırılmasını ele aldıkları çalışmalarında, PISA 2012 verileri ile istatistiksel varsayımlar açısından farklı olan yöntemler kullanılarak ölçme değişmezliği ile ilgili sonuçların incelenmesini amaçlanmışlardır. Ayrıca normallik varsayımı yapılması gereken ve gerekmeyen yöntemlerin normallik varsayımı sağlanamayan durumlarda farklı sonuçlar gösterip göstermediğine bakılmıştır. Bu amaca göre PISA 2012 alt ölçeklerinden beş maddeden oluşan Problem Çözmeye Açıklık ölçeği Türkiye ve Finlandiya örneklemi veri seti üzerinde yapısal eşitlik modellemesi içerisindeki ortalama kovaryans yapılarının değişmezliği analizi ve örtük sınıf analizi çatısı altındaki çoklu grup örtük sınıf analizi yöntemli analizler gerçekleştirilmiştir. Sonuçlara göre farklı metodlar kullanıldığında ölçme değişmezliği bulguları oldukça değişiklik göstermiştir. Bu durumda yöntemler için gerekli varsayımların mutlaka incelenmesi ve gizil ve gözlenen değişken yapıları göz önünde bulundurularak uygun yöntem seçmeye dikkat edilmesi hususuna dikkat çekmişlerdir.

Başusta ve Gelbal (2015), gruplararası kıyaslamalarda PISA öğrenci anketi örneğinde ölçme değişmezliğinin test edilmesini ele aldıkları çalışmalarında, PISA 2009 öğrenci anketinde yer alan fen bilgisi ve teknolojileri ile ilgili maddelere ait faktörleri belirleyerek bu faktörlerin cinsiyet grupları arasında ölçme değişmezliğine sahip olup olmadığını incelemişler ve Türkiye örneğini ele almışlardır (Büyüköztürk, Çakan, Tan ve Atar, 2014). Anket verilerinin analizi için yapısal eşitlik modellemesi (YEM) kullanılmış ve ölçme modeli aşama aşama sınırlandırılıp test edilmiştir. Değişmezlik testi olarak uyum indeksleri farkı alınmış ve analiz sonuçlarından elde edilen uyum indeksleri standart değerler baz alınarak karşılaştırılmış ve ölçme değişmezliğinin cinsiyet değişkeni bakımından sağlandığı sonucuna ulaşılmıştır. Buradan hareketle, ölçme modelinin kız ve erkek öğrenci grupları için aynı ve doğru şekilde ölçüm yaptığı söylenebilir.

Kıbrıslıoğlu (2015), kültürlere ve cinsiyete göre PISA 2012 matematik öğrenme modelinin ölçme değişmezliğinin incelenmesinin ele alındığı yüksek lisans çalışmasında, matematik öğrenme alt boyutu anketleri ile oluşturulan matematik öğrenme modelinin ülke ve cinsiyet grupları arasında ölçme değişmezliğinin sağlanıp sağlanmadığını belirlemeye çalışmıştır. Türkiye, Endonezya ve Çin-Şangay örneklemelerini ele almıştır. Açımlayıcı faktör analizi (AFA) ile oluşturduğu

matematik öğrenme algısı modelini doğruladıktan sonra, ölçme değişmezliği analizlerinde çoklu grup doğrulayıcı faktör analizini kullanarak modeli incelemiştir. Elde edilen araştırma sonuçlarına göre model ülkeler arasında ölçme değişmezliğinin tüm koşullarını sağlayamamış olup cinsiyet grupları arasında ise model ölçme değişmezliğini her aşamada sağlamıştır. Buradan hareketle modele ait ortalama, varyans ve kovaryansların yalnızca cinsiyet grupları arasında karşılaştırılabilir olduğu sonucuna ulaşmışlardır. Ayrıca farklılıkların cinsiyet farkından dolayı olduğunu da belirtmiştir.

Asil ve Brown (2015), PISA 2009 okuma becerileri testinin dil, kültür ve ekonomik gelişim düzeylerinde ölçme değişmezliğini test etmişlerdir. Araştırmaya, örneklem büyüklükleri göz önünde bulundurulduğundan 55 ülke ele alınmıştır. Avustralya'nın referans olarak seçildiği çalışmada, İngilizce konuşulan yalnızca üç ülkenin Avustralya ile skaler değişmezliği olduğu sonucuna ulaşılmıştır. Ayrıca elde edilen verilerden hareketle, eğitimde sosyo-ekonomik kaynakların ölçme değişmezliğinde önemli bir rol oynadığı ancak dil faktörünün ve dil eğitimi uygulamalarının etkilerinin daha az olduğu görülmüştür.

Uyar ve Doğan (2014), PISA 2009 Türkiye örnekleminde öğrenme stratejileri modelinin farklı gruplarda ölçme değişmezliğinin incelenmesinin ele alındığı yüksek lisans çalışmasında bazı bölümlerin içinde yer aldığı aynı isimli makaleden birkaç ekleme yaparak, PISA 2009 Türkiye örnekleminde öğrenci anketinde yer alan öğrenme stratejilerine ilişkin bir modeli; cinsiyet, okul türü ve istatistiksel 12 bölge gruplarına göre incelemiş ve ölçme değişmezliğinin sağlanıp sağlanmadığını belirlemeye çalışmışlardır. Sınırlı ve sınırlı olmayan modelden elde edilen değerler arasındaki farklara bakılıp, parametrelerin değişmez olup olmadığı incelenmiştir. Yaptıkları analizler sonucunda; modelin cinsiyet ve okul türleri gruplarında ölçme değişmezliğinin sağlanmadığını, istatistiksel bölge gruplarında ise tüm değişmezlik koşullarını yerine getirerek değişmezliğin sağlandığını belirlemişlerdir.

Bahadır (2012), PISA 2009'a verileri ile Türkiye'deki öğrencilerin okuma becerilerini etkileyen değişkenlerin bölgelere göre incelenmesinin ele alındığı yüksek lisans tezi çalışmasında, okuma becerisini etkilediği düşünülen aile ve öğrenci özelliklerini faktör analizi sonucunda üç boyutta ele almıştır. Anne-baba eğitim düzeyi ve öğrenciye sağlanan imkanların olduğu aile değişkeni, evde bulunan

kültürel varlıkların olduğu kulvar değişkeni ve öğrencilerin okumaya karşı tutumlarının olduğu tutum değişkeni çalışmanın boyutlarını oluşturmaktadır. Öğrencilerin okuma becerileriyle modele alınan değişkenler arasındaki uyumu test etmek için Yapısal Eşitlik Modellemesi (YEM) kullanılmıştır. Ölçme değişmezliği test edilerek modeldeki değişkenler coğrafi bölgeler arasında kıyaslanmıştır. Modelin skaler değişmezlik şartlarını sağlamaması sebebi ile kısmi değişmezlikten faydalanılarak değişmezlik aşamaları test edilerek bölgeler arasında en büyük farklılığın anne-baba eğitim düzeyi ve öğrenciye sağlanan olanaklarda olduğu görülmüştür.

Gülleroğlu (2017), PISA 2012 matematik okuryazarlığı testine katılan Türk öğrencilerin duyuşsal özelliklerinin cinsiyete göre ölçme değişmezliğinin incelenmesini ele aldığı çalışmada, matematiğe yönelik ilgi, matematik kaygısı, matematiğe yönelik benlik algısı ve matematik öz yeterliliği boyutlarının cinsiyete göre değişmezliğini tespit etmeyi amaçlamıştır ve verilere Çoklu Grup Doğrulayıcı Faktör Analizi (ÇG-DFA) uygulamıştır. Matematiğe yönelik ilgi, matematik kaygısı ve matematik benlik algısı değişkenlerinde hem kız hem de erkekler için aynı yapının var olduğunu ve bu örtük değişkenlerin gruplar için bile değişmez olduğunu ancak matematik öz yeterliliği değişkeninin bunu sağlamadığını gözler önüne sermiştir. Matematik kaygısı ve matematiğe ilgi boyutları için kız ve erkek gruplarının maddelere benzer şekilde cevap verdiği ve iki gruptan elde edilen puan karşılaştırmalarının anlamlı olduğu sonucu vurgulanmıştır.

Ertürk ve Erdinç Akan (2018), TIMSS 2015 matematik başarıları ile ilgili bazı değişkenlerin cinsiyet açısından ölçme değişmezliğinin incelenmesini ele alındığı çalışmalarında, Çok Gruplu Doğrulayıcı Faktör Analizi (ÇG-DFA) yöntemi kullanılarak öğrencilerin matematik performansı ve ilgili değişkenler yorumlanmıştır. Araştırmada yer alan matematiği sevme, matematiğe olan ilgi ve matematiğe ilişkin özgüven gizil değişkenleri için cinsiyet açısından sırasıyla biçimsel, metrik, skalar ve katı değişmezlik testleri yapılmıştır. Analizlerde biçimsel değişmezlik testine hiçbir sınırlamanın yapılmamış ve diğer değişmezlik testlerine kademeli olarak fazla sınırlama getirilerek arasındaki CFI karşılaştırmalı uyum iyiliği indeks farkları kıyaslanmıştır. Araştırmada cinsiyet gruplarında katı değişmezlik koşulunu sadece “matematiği sevme” gizil değişkeninin karşıladığı ancak test edilen

üç gizil değişkenin biçimsel değişmezlik koşulunu karşıladığı bulgusu gözler önüne serilmiştir.

Koğar ve Yılmaz Koğar (2017), TIMSS 2015 matematik başarıları ile öğretmenlerin matematik konularına yönelik hazırlıklı gelme düzeylerinin ilişkisini ele aldıkları çalışmalarında, Türkiye ve Singapur örneğinde bulunan başarı testi ve öğretmen anketinden yola çıkarak matematik öğretmenlerinin matematik öğrenme alanlarını, öğretmedeki yeterlik düzeylerini ortaya koyarak 8. sınıf öğrencilerinin bu alandaki başarılarının öğretmenlerin konuya hazırlıklı gelmesinden mi yoksa öğrencilerin bu konu hakkındaki başarılarından kaynaklı mı farklı sonuçlar verdiğini ayrıca ülkeden ülkeye farklılaşıp farklılaşmadığını analiz etmişlerdir. Singapur'daki öğretmenlerin Türkiye'deki öğretmenlere kıyasla matematik konularına daha iyi hazırlandıkları sonucuna varılmıştır ve derse hazırlık ne oranda artarsa öğrenci başarıları da aynı oranda artmaktadır. TIMSS'te bulunan matematik öğrenme alanları ile TEOG'ta bulunan matematik öğrenme alanları örtüşmektedir ve konulara hazırlık düzeyi ile matematik başarıları arasında pozitif uyum bulunan bu çalışmada öğretmenin konulara hazırlıklı gelmesindeki artışın öğrencilerin TEOG matematik başarılarında da yükseliş sağlayabileceği düşünülmektedir.

BÖLÜM III

YÖNTEM

Bu bölümde araştırmanın modeli, evren ve örneklem, verilerin elde edilmesi ve verilerin analizine ilişkin bilgiler yer almaktadır.

3.1. Araştırma Modeli

Bu araştırmanın modelini, TIMSS 2015 uygulamasında yer alan sekizinci sınıf matematik başarı testi yapılarının OECD üyesi ülkeler arasında ölçme değişmezliğini sağlama durumunun belirlenmesi oluşturmaktadır. Bu bağlamda çalışma, TIMSS 2015 değerlendirmesine ait bilişsel testin geçerlilik düzeyini belirlemeye yönelik olduğu için ilişkisel tarama modelinde yürütülmüştür. Tarama modelinde yürütülen çalışmalarda, evreni temsil eden belirli bir örneklem grubu üzerinde yapılan çalışmalarla evren hakkında genel bir yargıya varabilme amaçlanır (Bailey, 1982).

3.2. Evren ve Örneklem

TIMSS 2015 uygulamasında 8. sınıflardan 39 ülke ve 301.472 öğrenci, 4. sınıflardan 45 ülke ve 253.546 öğrenci olmak üzere toplam 61 ülke 555.018 öğrenci yer almıştır. Bu ülkeler, stratejik yapısı ile güçlü olan Amerika, Almanya, İngiltere, Rusya gibi ülkeler, eğitim seviyesi ile çok çok ilerde olan Kore, Kanada, Japonya, Singapur, HongKong gibi ülkeler; ülkede eğitime verilen önemin az olması sebebi ile eğitimde geride kalan Suudi Arabistan, Filistin, Ürdün gibi ülkeler; normal standartta eğitimle üst seviyelere çıkmaya çalışan Türkiye, Macaristan, Slovak Cumhuriyeti gibi ülkeler ve diğer ülkeler olmak üzere 61 ülkeden oluşmaktadır.

4.sınıf ve 8.sınıf ya da her iki sınıf düzeylerinde araştırmaya katılabilecek olan TIMSS 2015 uygulamasına katılan ülkeler eğitim sistemleri farklılıkları, gereksinimleri ve siyaset ilişkileri değerlendirilerek bu projeye dâhil edilmektedir. Öğrenciler bu programa seçim aşamasında 4. sınıfta ve 8. sınıfta olmasına göre elenip seçilmekte, yaşları da bu anlamda sırasıyla 9.5 ve 13.5 olmalıdır. Bu bağlamda

Türkiye’de ilkokul ve ortaokul bünyesinde örgün öğretimde okuyan 4. ve 8. sınıf öğrencilerinden TIMSS’in örneklem dağılımı ve başarıyı ölçme şeması görülmektedir. Bu projede diğer uluslararası sınavlarda ele alınan okuryazarlığın aksine başarı şeması ele alınmıştır.

ÖDSGM her TIMSS senesinde, Türkiye’deki bütün ilk ve orta dereceli okulların yer aldığı 4. ve 8. sınıf düzeylerinde bir liste hazırlayarak, TIMSS uygulayıcılarıyla paylaşmaktadır. Bu bilgiler doğrultusunda oluşturulan okullar, sınıflar ve yaş aralıklarından oluşan liste doğrultusunda IEA bu projeyi yapacağı okulları, sonrasında ise ÖDSGM hangi okuldan hangi şubenin projeye katılacağını tesadüfi yolla ve farklı yordama şekilleriyle seçmektedir.

TIMSS 2015 araştırmasına katılan ülkelerin tümünde, test maddeleri 14 kitapçığa ayrılarak uygulanmıştır. Bu çalışmada, bütün kitapçıklar tek tek analiz edilmiş ve en yüksek kişi sayısına sahip olması sebebi ile tüm ülkelerde uygulanmış olan 11 no’lu kitapçık seçilerek analizler bu kitapçıktaki matematik başarısını ölçmeyi amaçlayan maddeler üzerinden gerçekleştirilmiştir.

Araştırmanın evrenini OECD ülkelerinden TIMSS 2015 uygulamasına katılan 8. sınıf öğrencileri oluşturmaktadır. TIMSS 2015 uygulamasında 11 nolu kitapçığı alan bireylerin OECD ülkeleri için dağılım tablosu Tablo 2’deki şekildedir.

Tablo 2: OECD Ülkelerine Göre Örneklem Dağılımı

Ülke	N	%
Amerika	741	10,26
Avusturalya	732	10,13
Hong Kong	288	3,99
İngiltere	342	4,73
İrlanda	339	4,69
İsrail	393	5,44
İsveç	303	4,19
İtalya	321	4,44
Japonya	345	4,78
Kanada	631	8,74
Kore	384	5,32
Macaristan	354	4,90
Norveç	342	4,73
Şili	348	4,82
Slovenya	301	4,17
Türkiye	441	6,11
Yeni Zelanda	618	8,56
TOPLAM	7.223	100

Tablo 2 incelendiğinde 8. sınıf öğrencisi bulunan 17 OECD üyesi ülkede 7.223 kişiden çalışmaya dâhil edilen bireylerin dağılımı incelendiğinde en fazla katılımcıya sahip ülke %10,26 (741 kişi) ile Amerika, en az katılımcıya sahip olan ülke 3,99 (288 kişi) ile Hong Kong'dur. Diğer ülkelerin katılımcı oranları ise %10,13 ve %4,17 (732 ve 301 kişi) arasında değişmektedir.

3.3. Araştırma Verileri

TIMSS uygulaması, öğrencilerin akademik performanslarının ölçüldüğü bilişsel testler ve açık uçlu sorular ile öğrencilerin bütün olarak değerlendirilmesinin amaçlandığı okul ve öğrenci anketlerinden meydana gelir. Bu bağlamda genel olarak matematik ve fen bilimleri eğitim-öğretim düzeyinin kültürel, siyasal ve sosyal değişkenlerden etkilenmeden bu alandaki başarının uluslararası düzeye çıkarılması hedeflenmektedir. Bu hedeflerden yola çıkarak genel olarak diğer sınavlar gibi okuryazarlığı değerlendirmeyip akademik başarıyı değerlendiren TIMSS öğrencilerin uluslararası ve ulusal eğitim sistemlerindeki eksik, fazla ya da farklı, aynı yönleri belirleyerek öğrencilerin sınıf düzeylerindeki başarılarını analiz ederek yorumlamaktadır. Bu projeye her öğrencinin matematik ve fen bilimlerine ilgileri, başarıları, eğitim sistemindeki doğru ve yanlışlar, öğretim programları, öğretmenler, idareciler ve okullar ile ilgili veriler elde edilmektedir.

Bu uygulamada bilişsel testler, öğrencilerin matematik başarısı ağırlıklı olmak üzere, fen başarısı ve okuduğunu anlama, bilişsel düzeylerini kullanabilmeye ait performanslarını ölçmeyi amaçlar. Kâğıt-kalem testi olarak yapılan sınav, 14-14 toplam 28 bloktan oluşur. PISA'nın matematik ve fen okuryazarlığını ölçtüğü bir sınav olmasının aksine TIMSS'in genel olarak matematik ve fen başarısını ölçtüğü bir sınav olması sebebi ile her bir bloktaki sorular 4. ve 8. sınıf seviyesinde ve bilgiyi anlamadan çok bilme ve uygulama basamakları yoğun bir şekilde işlenen bir sınav olması ile tanınır.

TIMSS'in öğrenme alanları, alt öğrenme alanları ve kazanımları genel olarak 4. ve 8. sınıf öğrencilerinin matematik ve fen bilimleri dersinde öğrendikleri kısımlardan seçilmektedir ve böylece öğrencilerin bu öğrenme alanı, alt öğrenme alanı ve kazanıma dair tutumları ve kavramları öğrenme süreçleri uluslararası düzeyde belgelenmiş olur. Bununla birlikte 4. sınıftan 8. sınıfa kadar matematik ve fen bilimleri derslerinde öğrenme düzeyleri ve öğrenme becerileri aşamalandırılır ve gruplara ayrılır. Bu süreç her dört senede bir tekrar eder, böylece eğitim ve öğretim sürecinin değerlendirilmesi aşaması aksamadan devam eder. Bu bağlamda uluslararası düzeyde bilişsel alanda tüm verilere ulaşabildiğimiz ve sürekliliği olan nadir sınavlardan biri olması açısından önemlidir. Akabinde TIMSS'in geniş veri ağı vasıtasıyla her ülke kendi okul, öğretmen, öğrenci performansını değerlendirip

geleceğe yönelik başarıyı koruma ya da başarıyı arttırma yönünde kestirimsel planlar yapabilecek, düşük olan başarıların sebepleri öğrenilecek, bu uygulamaya katılan diğer ülkelerle aramızdaki benzer, farklı yanlar ve farklı yanların sebepleri görülebilecek ve bu ülkelerle iletişim halinde olunabilecektir.

TIMSS başarı testlerinin yanı sıra öğrenci ve öğretmen başarısını etkileyen bilişsel düzey, eğitim düzeyi ve sosyal çevre hakkında öneriler toplayan, iki tür olan öğrenci ve öğretmen anketlerinden oluşmaktadır. Böylece okul idarecileri, ilgili okul öğretmen ve öğrencileri hakkında detaylı bilgi sahibi olunacak, özellikle de öğrencilerin okulda yaşadıkları zorluklar, ders dinleme esnasında ve ders başarısında karşılaştıkları zorluklar, eğitim düzeyleri ve bu eğitimden ne derece keyif aldıkları, eğitimden, öğretmenden ve idareden beklentiler, daha iyi bir eğitim ortamı ve akabinde başarılı bir eğitim için öğrenci ve öğretmenlerin ne tür istekleri olduğu, eğitim ve öğretime bakış açıları aydınlığa kavuşacaktır. Test başarısının analizi kıyaslanması ve sebeplerinin yanında dile getirilemeyen sorunlar da bu anketlerle gözler önüne serilecektir. Böylece komisyonda bulunan ve sınav detayları konusunda karar veren kişiler bütün bu şikâyet, dilek ve öneri gibi aşamaları gözlemleyerek raporlaştıracaklardır. Genel bir ifadeyle anketlerle öğrenci ve öğretmenlerin sosyal görüşlerine ağırlık verilmektedir. Karar vericiler genel olarak her ülkeden temsilci seçilme usulü ile belirlenir. Bu kişiler matematik ve fen bilimleri başarı testlerinde hangi test maddesinin kullanılacağını ve hangi sırayla hangi kitapçıklarda uygulanacağını belirlerler. Test maddelerinin seçilme kriterlerine bakacak olursak test maddelerinin araştırılması, bulunması ve geliştirilmesi aşamalarıyla öğrencilerin fen bilimleri ve özellikle matematik alanlarındaki bilişsel düzeyleri, okuduğunu anlama ve yorumlama becerileri ölçülecek maddeler tercih edilmektedir. Bu test maddeleri 6 farklı kitapçıkta (her biri matematik ve fen bilimleri kısımlarına sahip) her sınıf eşdeğerinde ve pilot uygulamada tercih edilirken, gerçek uygulamasında 14 (her biri matematik ve fen bilimleri kısımlarına sahip) farklı kitapçık tercih edilmektedir. Hazırlanan bu kitapçıkları Ulusal Merkez öğrencilerin sınava gireceği birimlere iletmektedir ve her bir kitapçık tek tek öğrenci adına basılması adına güvenilirliği oldukça yüksek uluslararası bir sınav olduğu söylenebilir.

Verilere TIMSS'e ait resmi web sitesinin International Database - SPSS Data ilgili bölümünden ulaşılmış ve <http://timssandpirls.bc.edu/timss2015/international-database/> adresinden indirilmiştir. Araştırmada kullanılan maddeler Tablo 3'te gösterilmiştir.

Tablo 3: Analize Dâhil Edilen Matematik Başarı Testi Maddeleri

Maddeler	
	T15_G8_SPSS_Data_pt1.zip (AAD-COT 148MB) Updated*
Matematik Başarısı	T15_G8_SPSS_Data_pt2.zip (CQU-KAZ 129MB)*
	T15_G8_SPSS_Data_pt3.zip (KOR-OMN 134MB)*
	T15_G8_SPSS_Data_pt4.zip (QAT-ZAF 130MB)*
	TN15_SPSS_Data.zip (36MB)*

TIMSS değerlendirmesinde çoktan seçmeli test ve uçlu soru türleri kullanılmaktadır. 14 kitapçıktan seçilen 11 nolu kitapçıkta açık uçlu ve çoktan seçmeli maddeler bulunmaktadır. Bu araştırmada maddelere ait puanlamalar madde türüne göre farklılık gösterdiği için sadece matematik başarısına ait, 16 çoktan seçmeli ve 13 açık uçlu olmak üzere toplam 29 madde analize dâhil edilmiştir.

TIMSS uygulamasında çoktan seçmeli maddeler dört seçenekten oluşmaktadır. Bu tür maddelerde yanlış verilen yanıtlar doğru yanıtları etkilememektedir ve sunulan dört seçenekten biri doğru olarak kabul edilmektedir. Çoktan seçmeli maddelerin aksine açık uçlu maddelerde öğrenci kendi yanıtını yanıt kısmına yorumu ve sonucu ile yazmalıdır. Genellikle hızlı soru çözen öğrencilerin çoktan seçmeli maddelerde, yorum gücü kuvvetli olan öğrencilerin ise açık uçlu maddelerde başarılı oldukları görülmüştür. Açık uçlu madde türlerinde öğrenciler soruyu okuyup, anlamlandırır ve bilişsel düzeylerde onlardan istenen açıklamaları kendi yorum güçlerini de katarak

yani sözel, sayısal ve şekilsel şemalarla destekleyerek, verileri kullanarak açıklar ve yaparlar. Daha sonra bu maddeler dereceli puanlama anahtarı ve cevap anahtarlarıyla değerlendirilir.

Çoktan seçmeli maddelerin puan değerleri her doğru cevaba karşılık gelen puan “1”, her yanlış cevaba karşılık gelen puan “0” olarak belirlenmiş; açık uçlu maddelerin puan değerleri puanlama anahtarına bakılarak kısmi, boş ve yanlış olanlar yanlış, doğru olanlar doğru olacak şekilde atanmıştır.

Bahsedilen TIMSS 2015 matematik 8. sınıf soru başlıklarına ve cevap anahtarına Tablo 4’te yer verilmiştir.

Tablo 4: TIMSS 2015 Matematik 8. Sınıf İngilizce – Türkçe Madde Başlıkları ve Cevap Anahtarı

TIMSS 2015		Cevap
Matematik 8. Sınıf Soru Başlıkları		Anahtarı
1. Soru	length of shorter string piece	(B)
Türkçesi	(kısa dizi parçası uzunluğu)	
2. Soru	which is a prime number	(C)
Türkçesi	(hangisi başlangıç numarasıdır)	
3. Soru	answer to Robert's problem	(1)
Türkçesi	(Robert’in sorununa cevap ver)	
4. Soru	fraction of diagram shaded	(1)
Türkçesi	(gölgeli kesir şeması)	
5. Soru	which statement is true	(B)
Türkçesi	(hangi ifade doğrudur)	
6. Soru	value of given expression	(B)
Türkçesi	(verilen ifadenin değeri)	

Tablo 5'ün devamı: TIMSS 2015 Matematik 8. Sınıf İngilizce – Türkçe Madde Başlıkları ve Cevap Anahtarı

TIMSS 2015		Cevap
	Matematik 8. Sınıf Soru Başlıkları	Anahtarı
7. Soru	identify an equivalent expression	(A)
Türkçesi	(eşdeğer bir ifade tanımlayınız)	
8. Soru	solve for simultaneous equations	(1)
Türkçesi	(simultane denklemler için çözüm)	
9. Soru	lines of symmetry for pentagon	(1)
Türkçesi	(pentagon için simetri çizgileri)	
10. Soru	length of line segment ab	(1)
Türkçesi	(ab sınırı parçası uzunluğu)	
11. Soru	number of cubes that fit in a box	(D)
Türkçesi	(kutunun içine uygun olan küp sayısı)	
12. Soru	girls playing video games	(C)
Türkçesi	(video oyunu oynayan kızlar)	
13. Soru	difference in percentages	(C)
Türkçesi	(fark yüzdesi)	
14. Soru	raspberry as favorite flavor	(D)
Türkçesi	(ahşap favori bayrak)	
15. Soru	explain why Pat is correct	(1)
Türkçesi	(Pat'ın neden doğru olduğunu açıklayın)	
16. Soru	Randy's blue balloons	(D)
Türkçesi	(Randy'nin mavi balonları)	

Tablo 6'ün devamı: TIMSS 2015 Matematik 8. Sınıf İngilizce – Türkçe Madde Başlıkları ve Cevap Anahtarı

TIMSS 2015		Cevap
Matematik 8. Sınıf Soru Başlıkları		Anahtarı
17. Soru	least number of crates for Harry's apples	(1)
Türkçesi	(Harry'nin elmaları için en az kesim sayısı)	
18. Soru	fill in new price/discount in bicycle sale table	(2)
Türkçesi	(bisiklet satış tablosundaki yeni fiyatı / indirimi doldurun)	
19. Soru	number of girls in a class given boys: girls	(1)
Türkçesi	(erkek : kız oranı verilen bir sınıftaki kız sayısı)	
20. Soru	identify expression equivalent to $(3x+6)/9x$	(B)
Türkçesi	$((3x + 6)/9x$ 'e eşdeğer olan ifade)	
21. Soru	equation for Bruno's monthly cell phone cost	(A)
Türkçesi	(Bruno'nun aylık cep telefonu maliyeti denklemi)	
22. Soru	room divided into classroom and laboratory	(A)
Türkçesi	(sınıf ve laboratuvara dönüşen oda)	
23. Soru	find cost of taxi ride with table	(1)
Türkçesi	(taksimetre maliyetini tablo ile bul)	
24. Soru	shaded square reflected over the line xy	(D)
Türkçesi	xy çizgisinde bulunan kare içerisindeki çember	
25. Soru	sculpture from solid cube - number of faces	(1)
Türkçesi	(katı cisimlerde küpün yüzey sayısı)	

Tablo 7'ün devamı: TIMSS 2015 Matematik 8. Sınıf İngilizce – Türkçe Madde Başlıkları ve Cevap Anahtarı

TIMSS 2015		Cevap
Matematik 8. Sınıf Soru Başlıkları		Anahtarı
26. Soru	number of cubes in solid figure	(C)
Türkçesi	(katı cisimlerde küplerin sayısı)	
27. Soru	total bowls to raise average to 15	(1)
Türkçesi	(toplam değerleri 15'e yükseltmek)	
28. Soru	fishing histogram - proportion of fish released	(D)
Türkçesi	(histogram – çıktı oranları)	
29. Soru	draw a spinner given chance of outcomes	(1)
Türkçesi	(topaç çiziminde verilen değişim sonuçları)	

Tablo 4'te sorulara başlıklar bağlamında bakıldığında Türkiye'de eskiden kullanılan ezber dayalı sorular değil, artık yenilikçi eğitim anlayışıyla günümüzde uygulanan anlamaya, keşfetmeye, yorumlama gücüne dayalı sorular kullanılmakta olduğu gözler önündedir. Ayrıca sorular bilgi, kavrama, uygulama basamaklarını yoklamaktadır, soruların isimleri ve örnek sorulardan da (ekler kısmında TIMSS 2015, 2011 ve 2007 senelerine ait örnek sorulara yer verilmiştir.) anlaşılacağı üzere günlük hayattan soruların ya da geometrik cisimler ve katı cisimlerle alakalı soruların sıkça kullanıldığı gözükmektedir.

Bu sorular ve örnekler bağlamında bazı ifadelerin sınav anahtarının ülkesinin kültürünü daha yakın düzeyde yansıttığı ifade edilebilir. Bunlara örnek olarak sorularda Randy, Harry gibi yabancı isimlerin sıkça kullanılması, bazı geometrik ifadelerin bizdeki anlamından daha farklı kullanılması türünden ifadelere yer verilebilir.

TIMSS uygulamasında, sekizinci sınıf seviyesinde matematik ve fen bilimleri cevap anahtarında bilişsel alan ve öğrenme alanı olmak üzere iki kısım bulunmaktadır. Bu boyutlara Tablo 5'te yer verilmiştir. TIMSS 2015 verilerine

dayanarak 8. sınıf matematik sorularının uygulama ağırlıklı olduğu görülmektedir. Sonrasında bilgi ve en son sırada da akıl yürütme becerisi gelmektedir.

Tablo 8: TIMSS Matematik Maddelerinin 8. Sınıf Bilişsel Alanlara Göre Yüzdese Dağılımı

Bilişsel Alan	8. Sınıf Matematik
Bilgi	%35
Uygulama	%40
Akıl Yürütme	%25

Genellikle sorular bilgi, kavrama, uygulama basamaklarını yoklamaktadır, ekler kısmında TIMSS 2015, 2011 ve 2007 senelerine ait örnek sorulara yer verilmiş olup bu soruların cevap anahtarında her bir soru için bilişsel alanı ve alt öğreme alanı verilmiştir. Böylece her sorunun neyi yordadığını, hangi amaca hizmet ettiğini açıkça görebilmekteyiz.

Tablo 6'da bulunan TIMSS matematik maddelerinin bilişsel alanlara göre yüzdesi incelendiğinde çoktan seçmeli maddeler içerisinde en yüksek oranın bilgi basamağından (%36) sonrasında uygulama (%39) ve akıl yürütme (%25) ağırlıklı olduğu söylenebilir. Çoktan seçmeli maddelerin genel olarak bilgiye dayalı olması sebebi ile bu şekilde sonuçlandığı söylenebilir. Ayrıca kısa cevaplı maddelerden en yüksek oranın uygulama basamağında olması kısa cevaplı sorularda öğrencilerin uygulama ve fikir geliştirebilme gücünün gelişmesi açısından önemlidir.

Tablo 9: TIMSS Matematik Maddelerinin Bilişsel Alanlara Göre Yüzdesi

Bilişsel Düzey	Çoktan Seçmeli Maddeler A(B)*	Kısa Cevaplı Maddeler A(B)*	Toplam Madde A(B)*	Test Puanındaki Ağırlığı %
Bilgi	53(53)	27(30)	80(83)	%36
Uygulama	47(47)	38(44)	85(91)	%39
Akıl Yürütme	18(18)	34(40)	52(58)	%25
Toplam	118(118)	99(114)	217(232)	%100
Test Puanındaki Ağırlığı	%51	%49		

A(B)*: Toplam madde sayısı (Toplam alınabilecek puan)

Türkiye açısından Tablo 6'ya bakacak olursak ülkemizin test analizlerinin de genellikle bilgi düzeyinde olduğu anlaşılmaktadır. Son zamanlarda yapılan ölçme değerlendirme çalışmalarıyla soru tarzları değiştirilmiş, daha uzun anlama ve yorumlamaya dayalı sorular yazılmaya ve sınavlarda sorulmaya başlanmıştır. 8. sınıf örnek sınav sorularına ÖDSGM sayfasından bakılabilir.

Bu verilere dayanarak Tablo 7'ye bakıldığında 8. sınıf matematik sorularının öğrenme alanı ve alt öğrenme alanları gözükmekte, sayılar ve cebir öğrenme alanlarının daha yoğun olduğu, geometri ve veri olasılık öğrenme alanlarının kısmen daha az olduğu belirtilmektedir. Alt öğrenme alanı bakımından çoğu kazanıma yer vermeye çalışıldığı gözükmektedir.

Tablo 10: TIMSS Matematik Maddelerinin Sekizinci Sınıf Matematik Öğrenme Alanlarına Göre Yüzdesel Dağılımı

Öğrenme Alanı	Alt Öğrenme Alanı	Yüzde
Sayılar	Doğal Sayılar	%30
	Kesirler, Ondalık Sayılar ve Tam Sayılar	
	Oran, Orantı ve Yüzde	
Cebir	Örüntüler	%30
	Cebirsel İfadeler ve İşlemler	
	Eşitlik ve Eşitsizlikler	
	Formüller ve Fonksiyonlar	
Geometri	Geometrik Şekiller	%20
	Geometrik Ölçme	
	Dik Koordinat Sistemi (Bir Noktanın Yeri, Yansıma, Öteleme ve Dönme)	
Veri ve Olasılık	Veri Kümesinin Özellikleri	%20
	Verileri Yorumlama	
	Olasılık	

Bu bilgiler ışığında Tablo 8'e bakacak olursak matematik başarı testinin kapsam olarak sayılar, cebir, geometri, veri ve olasılık öğrenme alanlarından oluştuğunu görmüştük. En çok cebir ve sayılar öğrenme alanında, geometri, veri ve olasılık alanında kısmen daha az çoktan seçmeli madde bulunmaktadır. Kısa cevaplı maddelerin kıyaslaması da çoktan seçmelilerle benzer şekildedir. Toplama bakacak olursak en çok cebir daha sonra sayılar, geometri, veri ve olasılık olarak sıralayabiliriz.

Tablo 11: TIMSS 8. Sınıf Matematik Maddelerinin Matematik Öğrenme Alanlarına Göre Yüzdesi

8. Sınıf Matematik Başarı Testinin Kapsamı	Çoktan Seçmeli Maddeler A(B)*	Kısa Cevaplı Maddeler A(B)*	Toplam Madde A(B)*	Test Puanındaki Ağırlığı %
Sayılar	31(31)	30(36)	61(67)	%29
Cebir	37(37)	33(39)	70(76)	%33
Geometri	25(25)	18(19)	43(44)	%19
Veri ve Olasılık	25(25)	18(20)	43(45)	%19
Toplam	118(118)	99(114)	217(232)	%100
Test Puanındaki Ağırlığı	%51	%49		

A(B)*: Toplam madde sayısı (Toplam alınabilecek puan)

Sekizinci sınıflarda matematik başarı testi kırk beş dakika, fen bilimleri başarı testi de kırk beş dakika olacak şekilde toplam doksan dakika olarak belirlenmiştir. Matematik ve fen bilimleri testlerinin arasında verilen mola süresi beş dakika olarak belirlenmiştir. Matematik ve fen bilimleri başarı testi bitiminde de öğrencilere on beş dakika mola süresi vererek, daha sonrasında öğrencilere anketi cevaplandırabilmeleri için ekstra otuz dakika süre eklenmiştir.

Genel olarak TIMSS uygulamasında öğrenci anketlerini dolduran öğrenciler daha sonra başarı testlerini yaparlar ve 2 kısım cevaplandırılır. Bu anket türleri öğrenci anketi, öğretmen anketi ve okul anketi olmak üzere üçe ayrılır ve sekizinci sınıflara, matematik-fen bilimleri öğretmenlerine ve idacilerle okul aile birliği bünyesindeki kişilere uygulanır.

Bu anket çeşitlerini ele alacak olursak, okul anketi soruları genel olarak; öğretmenlerin işsel bilgileri, öğrencilerin kişisel bilgileri, matematik ve fen bilimleri dersleri ile alakalı bütün kaynaklar, okul-aile birliğinin okulda yürüttüğü faaliyetler, eğitim-öğretim süresi ve kurs detayları, okulun başarısı, konumu ve sosyal özellikleri, veli profili vb. sorulardan oluşmaktadır. Öğretmen anketine sekizinci sınıf matematik ve fen bilimleri öğretmenleri katılır. Bu ankette sekizinci sınıf öğretmenlerinin kişisel bilgileri, meslek tecrübeleri, akademik kariyer deneyimleri, ders anlatma stilleri, öğrenciye karşı tutumları, pedagoji alanında ne kadar iyi olduğu ve kendini genel olarak ne seviyesinde gördüğü, kurs ve ders saatlerinin süresi, matematik ve fen bilimleri kaynaklarına yönelik görüşleri, kendini öğrencilerin gözünde nasıl gördüğü, öğretmenliği nasıl tanımladığı, derslere yönelik görüşleri gibi sorulardan oluşmaktadır. Son kısım olan öğrenci anketinde öğrencilerin ev ödevlerine ve derslere karşı tutumlarına, okul çıkışlarında nasıl zaman geçirdiklerine, matematik ve fen bilimleri derslerini anlama düzeyleri ve kalıcılık oranlarına, bu derslere yönelik olumlu tutum geliştirebilme becerilerine, bilgisayar kullanımı ve kısıtına karşı bakış açılarına, ev içi ders çalışma ortamlarına ve ders gereçlerine yönelik sorular bulunmaktadır. Yani genel bir ifadeyle bu anketlerle okul kültürü ve okul yapısı detaylı şekilde ele alınmış, okulun adeta profili çıkarılmıştır. Bu açıdan sürekli uluslararası ve detaylı bir sınav olması sebebi ile analizlerde datalarına çokça başvurulmaktadır.

TIMSS döngüleri ve Türkiye'nin katılımına Tablo 9'da bakacak olursak; TIMSS'e 41 ülkenin katılım gösterdiği 1995 ve 46 ülkenin katılım gösterdiği 2003 yıllarında ülkemiz katılmamıştır. 38 ülkenin katılım gösterdiği 1999 senesinde sadece 8. sınıf düzeyinde katılmış olup, 1999 katılımı ilk katılımımızı oluşturmuştur. Daha sonra 59 ülkenin katılım gösterdiği 2007 senesi 8. sınıf düzeyinde, 66 ülkenin katılım gösterdiği 2011 senesi ilk kez 4. ve 8. sınıf düzeyinde ve en yakın tarih olan 2015 senesinde 61 ülkenin katılımı ile 4. ve 8. sınıf düzeyinde sınava katılmıştır. 2019 senesinde yapılan TIMSS'in verileri henüz yayınlanmamıştır.

Tablo 12: TIMSS'e Ait Yıllara Göre Temel Bilgiler ve Katılım Durumu

Yıl	Katılımcı Ülke Sayısı	Türkiye'nin Katılım Durumu
1995	41	Katılmamıştır.
1999	38	Sekizinci sınıf düzeyinde katılmıştır.
2003	46	Katılmamıştır.
2007	59	Sadece sekizinci sınıf düzeyinde katılmıştır.
2011	66	Dördüncü ve sekizinci sınıf düzeyinde katılmıştır.
2015	61	Dördüncü ve sekizinci sınıf düzeyinde katılmıştır.

Tablo 10'a bakacak olursak, TIMSS 2011 Türkiye verilerine dayanarak ülkemizden toplam 67 şehir, bu şehirlerde bulunan 239 okul, bu okullarda görev yapan 239 matematik öğretmeni, 23 fen ve teknoloji öğretmeni ve bu okullarda öğrenim gören 7286 öğrenci katılmıştır. Bu uygulamaya sınıf öğretmenleri katılmamıştır.

Tablo 13: Türkiye'de TIMSS 2011 Uygulamasına Katılan Kişilere Dair Sayısal Veriler

8.Sınıf	
İl Sayısı	67
Okul Sayısı	239
Öğrenci Sayısı	7286
Sınıf Öğretmeni	-
Matematik Öğretmeni	239
Fen ve Teknoloji Öğretmeni	238

TIMSS ulusal sınavında Tablo 11'e bakacak olursak; ülkemiz 1999 yılında 38 ülkenin katıldığı uygulamada 429 puanla 31. sırada, 2007 yılında 50 ülkenin katıldığı uygulamada 432 puanla 30. sırada, 2011 yılında 45 ülkenin katıldığı uygulamada 452 puanla 24. sırada yer almıştır. Bu bilgilerden hareketle başarımızın puan ve sıralama bazında giderek arttığı söylenebilir.

Tablo 14: 1999, 2007 ve 2011'de TIMSS 8. Sınıf Matematik Başarısının Ülke Sayısı Bağlamında Sıralaması

Yıl	Puan	Sıralama	Ülke Sayısı
2011	452	24	45
2007	432	30	50
1999	429	31	38

Tablo 11'deki veriler ışığında TIMSS 2011 uygulamasında on dört ülke belirlenen kıstas ortalama puanın üzerinde yer almış, buna karşın otuz bir ülke ise belirlenen kıstas ortalama puandan daha az puan almıştır.

Ayrıca TIMSS 1999, 2007 ve 2011'de matematik alanındaki öğrenci başarısını sekizinci sınıf cinsiyete göre Tablo 12'de ki bilgiler ışığında bakacak olursak; 1999 yılında Türkiye ortalaması baz alındığında erkek ortalamasının (429) kız ortalamasını (428) geçtiği, 2007 yılında kız ortalamasının (432) ve erkek ortalamasının (432) eşitlendiği ve 2011 yılında kız ortalamasının (457) erkek ortalamasını (448) geçtiği söylenebilir. Ayrıca 1999 yılında uluslararası ortalama baz alındığında erkek ortalamasının (489) kız ortalamasını (485) geçtiği, 2007 yılında kız ortalamasının (453) erkek ortalamasını (448) geçtiği ve 2011 yılında da kız ortalamasının (469) erkek ortalamasını (465) geçtiği söylenebilir. Bu bilgilerden hareketle TIMSS'in ilk yıllarında erkek başarısının kız başarısını geçtiği, son yıllarda ise kız başarısının erkek başarısının önüne geçtiği gözler önündedir.

Tablo 15: 1999, 2007 ve 2011'de TIMSS 8. Sınıf Cinsiyete Göre Matematik Alanındaki Öğrenci Başarısı

	Türkiye Ortalaması		Uluslararası Ortalama	
	Kız	Erkek	Kız	Erkek
2011	457	448	469	465
2007	432	432	453	448
1999	428	429	485	489

Örneğin bu alanda yapılmış bir çalışmaya bakacak olursak; dil öğrenme stratejileri ile ilgili çalışmaların bir kısmında cinsiyetten kaynaklanan farklar bulunmuş ve bu çalışmalarda genellikle kızlar lehine bulgulara rastlanmıştır. (Green, 1992; Noguchi, 1991; Green ve Oxford,1993; Oxford, 1993; Akt. Shmais, 2003; Oxford ve Nyikos, 1989; Sheorey,1999). Ayrıca Hong-Nam ve Leavell'in (2006) çalışmasına kız öğrencilerin erkek öğrencilere kıyasla çoğu stratejiyi daha sıklıkla kullandığı ve uyguladığı gözlemlenmiştir.

Tüm bu durumlarının yola çıkış yorumu olan ve cinsiyet faktörlerine bağlı olarak Oxford, Nyikos ve Ehrman (1988) cinsiyetler arası farka göre dil öğrenme stratejileri seçimini etkileyen faktörleri, kızların toplum tarafından onaylanma isteğinin ağır basması ve kızların var olan normları kabul etmede daha istekli oluşu olarak, yani genel bir ifadeyle kızların toplumun genel normlarından kaynaklı olarak kendilerini topluma kanıtlama çabası içinde olması sebebi ile daha başarılı olmaya yatkın oldukları şeklinde tanımlamıştır.

Tablo 13'de yer alan bilgiler alındığında matematik başarı sıralamalarına ve oranlarına bakacak olursak; 1999 yılında genel olarak yığılmanın alt düzey altı (400 puan altı) ve alt düzey (400-475 puan) arası olduğu, orta düzeyin (475-550 puan) oranının orta seviyelerde, özellikle üst düzey (550-625 puan) ve ileri düzeyde (625 üstü) çok az kişinin bulunması oldukça dikkat çekicidir. 2007 yılında genel olarak yığılmanın gene alt düzey altı (400 puan altı) ve alt düzey (400-475 puan) arası olduğu ve gene orta düzeyin (475-550 puan) oranının orta seviyelerde, üst düzey (550-625 puan) ve ileri düzeyde (625 üstü) diğer düzeylere nazaran daha az kişi oluşu gözlemlenmiştir. Ancak 1999 yılına oranla 2007 yılında yüzdenin üst düzeye doğru kısmen de olsa kayması gözle görülür bir farktır. 2011 yılında genel olarak yığılmanın gene alt düzey altı (400 puan altı) ve alt düzey (400-475 puan) arası

olduğu ve gene orta düzeyin (475-550 puan) oranının orta seviyelerde, üst düzey (550-625 puan) ve ileri düzeyde (625 üstü) diğer düzeylere nazaran daha az kişi oluşu gözlemlenmiştir. Ancak 1999 ve 2007 yıllarına oranla 2011 yılında yüzdenin üst düzeye doğru oldukça artış göstermesi de önemli bir farka işaret etmektedir.

Tablo 16: 1999, 2007 ve 2011’de TIMSS Uygulamasının Yeterlik Düzeyine Göre Matematik Başarı Yüzdesi

Yıl	İleri Düzey (625 üstü)	Üst Düzey (550 ve 625 arası)	Orta Düzey (475 ve 550 arası)	Alt Düzey (400 ve 475 arası)	Alt Düzey Altı (400 altı)
2011	%7	%13	%20	%27	%33
2007	%5	%10	%18	%26	%41
1999	%1	%6	%20	%38	%35

2007 yılından 2011 yılına kadar olan süreçte matematik başarısında en fazla artış ve başarı gösteren ülkeler sıralamasında Tablo 14’e bakacak olursak Katar 103 puanla 1. sırada, Suudi Arabistan 65 puanla 2. sırada, İsrail 53 puanla 3. sırada, Filistin 37 puanla 4. sırada, Rusya 27 puanla 5. sırada, Gana 22 puanla 6. sırada, Gürcistan 21 puanla 7. sırada, Türkiye 20 puanla 8. sırada, Singapur 18 puanla 9. sırada, İtalya 18 puanla 10. sırada yer almaktadır.

Tablo 17: 2007 Yılından 2011 Yılına 8. Sınıf Matematik Başarısında En Fazla Artış Gösteren 10 Ülke

Sıra	Ülke	Artış (Puan)
1	Katar	103
2	Suudi Arabistan	65
3	İsrail	53
4	Filistin	37
5	Rusya	27
6	Gana	22
7	Gürcistan	21
8	Türkiye	20
9	Singapur	18
10	İtalya	18

Malone ve Haimes (2003)'e göre TIMSS çalışmaları öğrenci başarılarının değerlendirilmesinden de daha geniş bir kapsama sahiptir. TIMSS uygulamaları öğrencilerin başarılarına ilişkin verilerin yanı sıra ülkelerin öğretim programları ve öğretim uygulamalarıyla bunların sınıf içinde-dışında öğretmenler ve öğrenciler üzerindeki etkileri hakkında geniş çapta bilgi edinilmesini sağlamaktadır. Ayrıca alanyazında yapılan araştırmalar sonucunda araştırmacılar çalışmalarında Amerika'daki öğrencilerin Almanya'dakilerden farklı olarak yazı yazmak için bilgisayar kullanma ve elektronik iletişim gerçekleştirme özelliklerinin fen ve matematik başarıları üzerinde anlamlı farklılığa neden olduğunu tespit etmişlerdir (Balım, Evrekli, İnel, Deniz, 2009). Tüm bu bilgilerden ve daha birçok araştırmadan yola çıkarak, ülkelerin başarılarının artış ve düşüş sebepleri siyaset, teknoloji, gelişmişlik seviyesi, sosyo-ekonomik durum, aile-kültür yapısı ve gelenek gibi faktörlerden etkilenecek değişebilmektedir.

3.4. Verilerin Analizi

Veriler analiz edilmeden önce tüm sayıtlar incelenmiş, merkezi eğilim ölçüleri, basıklık ve çarpıklık katsayıları, KR-20 güvenilirlik katsayıları hesaplanmıştır. Ölçeğin geçerliliği, Kolmogorov-Smirnov testi ile sınanmış, elde edilen sonuçların normal dağılıma uygun olduğu belirlenmiştir. Verilere AFA (Açımlayıcı Faktör Analizi) yapılmış ve tek faktörlü yapıda olduğu bulunmuştur. Yapı geçerliliğine ilişkin kanıt elde edebilmek amacıyla DFA (Doğrulayıcı Faktör Analizi) yapılmıştır. Daha sonra kovaryans matrislerinin eşitliği test edilerek, maddelere ait ölçme değişmezliği OECD üyesi ülkelere göre ÇGDFA (Çok Gruplu Doğrulayıcı Faktör Analizi) yöntemi ile test edilmiştir. Analizler çeşitli veri analizi paket programları kullanılarak yapılmış, asimtotik kovaryanslar matrisi üzerinden model parametrelerinin kestirilmesinde en çok ağırlıklandırılmamış küçük kareler yöntemi (unweighted least square-ULS) kullanılmıştır.

İki kategorili veriler için tetrakorik korelasyon kullanılırken, polikorik korelasyon ikiden fazla kategoriye sahip veriler için kullanılır. Ordinal veriler için geliştirilen tahmin yöntemlerinden bazıları en küçük kareler ağırlıklı (WLS), ağırlıklandırılmamış en küçük kareler (ULS) ve çapraz ağırlıklandırılmış en küçük kareler (DWLS) yöntemleridir (Babakus, 1985).

Alanyazında yapılan arařtırmalar sonucunda normalde istatistiksel yöntemler ierisinde yer alan regresyon özümlemesi için en ok kullanılan yöntem en küçük kareler yöntemidir. Olası birçok regresyon yöntemlerinin dıřında, genellikle matematiksel hesaplamalardaki kolaylıđından dolayı veri analizi uygulamalarında en küçük kareler yöntemi (WLS) kestiricileri yaygın olarak tercih edilmektedir. (Alma, 2008)

Fakat alanyazında yapılan bazı arařtırmalar sonucunda ađırlıklandırılmamıř küçük kareler (ULS) yöntemlerinin, ordinalde daha az standart hata olan parametreleri daha iyi tahmin ettiđi bulunmuřtur. Parametrelerin tahmininde ađırlıklandırılmamıř en küçük kareler (ULS) yöntemi daha düşük örneklem boyutlarında daha iyi sonuçlar verdiđi ve apraz ađırlıklandırılmıř en küçük kareler (DWLS) yönteminin ise yüksek örneklem boyutlarında daha iyi sonuçlar verdiđi gözlenmiřtir. Ayrıca ađırlıklandırılmamıř küçük kareler (ULS) yöntemi, bir kerede tüm parametreleri tahmin yeteneđi gibi bazı özelliklere sahiptir ve en az yineleme sayısına sahip parametreleri tahmin etmiřtir bu durumda tüm veri kümeleri ve ilgili örnek verinin parametrelerini tahmin etmek için daha dođru bir yöntemdir (Kođar ve Kođar, 2015). Bu yüzden analizde ilgili alıřmada OECD ülkeleri ile analiz yapılması sebebi ile örneklem azalmıř ve bu sebeple ađırlıklandırılmamıř en küçük kareler (ULS) yöntemi kullanılmıřtır.

Analizlerin yapılabilmesi için veri setinin temel varsayımları karşılayıp karşılamadıđı test edilmiřtir. Bu varsayımlar ait sonuçlar ařađıda verilmiřtir.

Tablo 18: OECD Üyesi Ülkelerin 8. Sınıf Matematik Başarı Testi Puanları için Test İstatistikleri, Normallik Testleri ve Güvenilirlik Katsayıları

Ülke	N	X	Mod	Medyan	S	Ranj	K _y	B _s	KR-20
Amerika	741	14.21	19	14	6.14	29	.109	-.733	.87
Avusturalya	732	13.87	11	13	6.42	28	.073	-.890	.89
Hong Kong	288	19.99	25	22	5.93	28	-1.155	.744	.89
İngiltere	342	15.47	12	15	6.25	28	.132	.749	.88
İrlanda	339	15.64	21	16	6.34	27	-.220	-.795	.88
İsrail	393	14.10	10	13	7.54	28	.110	-1.197	.92
İsveç	303	13	6	13	6.45	28	.191	-.697	.89
İtalya	321	12.57	9	12	5.70	28	.136	-.360	.86
Japonya	345	12.24	24	19	6.58	26	-.427	-.830	.90
Kanada	631	15.70	17	17	5.65	28	-.238	-.611	.85
Kore	384	22.70	28	24	5.34	27	-1.008	.532	.83
Macaristan	354	15.40	16	15	6.85	28	-.031	-.929	.90
Norveç	342	12.62	12	12	5.43	27	.351	-.329	.84
Şili	348	10.25	11	10	5.28	26	.598	.055	.84
Slovenya	301	14.41	13	14	5.61	25	.204	.589	.85
Türkiye	441	10.95	5	10	6.42	29	.669	-.404	.89
Yeni Zelanda	618	13.84	9	14	6.28	28	.148	-.849	.88

K_y: Çarpıklık katsayısı

B_s: Basıklık katsayısı

KR-20: Güvenilirlik katsayısı

Tablo 15’de yer alan sonuçlara göre, matematik okur-yazarlığı test puanlarının OECD ülkelerine ait merkezi eğilim ölçülerinin birbirine yakın olduğu söylenebilir. Normallik için basıklık ve çarpıklık katsayıları incelendiğinde 17 ülkeden 14 ülkeye ait değerlerin ± 1 arasında yer alıyor olması veri setinin normale yakın bir dağılım gösterdiği anlamına gelir. Ülkelere ait güvenilirlik katsayıları (KR-20) incelendiğinde ise güvenilirliğin .83 ile .92 arasında değiştiği görülmüştür. KR-20 değeri .80’de küçük ise madde çıkarılır, .80 - .90 arasında ise küçük düzeltmeler yapılır, .90 ve üzerinde ise kullanılır. Buradan hareketle güvenilirlik katsayılarının hepsinin .80 üstü olması sebebi ile madde çıkarılmaması kararı alınmıştır. Testin maddelerinin çoğunlukla iç tutarlığının iyi olduğu ifade edilebilir.

Bazı ülkeler için kabul edilebilir düzeyin üstünde, bazı ülkeler içinse altında olması, yapılan araştırma bir ölçme aracı geliştirme çalışması olmadığı için sadece raporlaştırılmış ve herhangi bir işlem yapılmamıştır. ÇGDFA yapılmadan önce ülkelere ait kovaryans matrislerinin eşitliği test edilmiştir. Test sonuçları Tablo 16’da verilmiştir.

Tablo 19: OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Kovaryans Matrislerinin Eşitliği Testi Sonuçları

Grup	S-B_{χ²}(sd)	P	χ²/sd	RMSEA	GFI	CFI	SRMR
OECD Üyesi Ülkeler	18179.21(6960)	.00	2.61	.063	.94	.92	.063

Tablo 16’da χ^2/sd oranı 3’ün altında çıktığı, RMSEA değerinin .08’in altında, SRMR değerinin .08’in altında, CFI değerinin .90’ın üstünde ve GFI değerinin .90’ın üzerinde çıktığı görülmektedir. Uyum indeksleri değerlendirildiğinde χ^2/sd manidarlığı oldukça iyi, RMSEA değerinin iyi uyum, SRMR değerinin iyi uyum, CFI ve GFI değerlerinin oldukça iyi uyum gösterdiği görülmüştür. Ükelere ait kovaryans matrislerinin oldukça iyi bir uyum gösterdiği ifade edilebilir. Ükelere ait DFA sonuçlarının çok sayıda ülke olması sebebiyle ÇGDFA sonuçlarından ayrı tablo halinde sunulması uygun görülmüş ve sonuçlar Tablo 17’de verilmiştir.

Tablo 20: OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Doğrulayıcı Faktör Analizi Sonuçları

Ülke	S-B χ^2 (sd)	p	χ^2 /sd	RMSEA	GFI	CFI	SRMR
Amerika	804.71(377)	.00	2.13	.039	.93	.97	.040
Avustralya	818.38(377)	.00	2.17	.040	.93	.97	.039
Hong Kong	552.84(377)	.00	1.47	.040	.87	.98	.053
İngiltere	542.46(377)	.00	1.44	.036	.90	.97	.048
İrlanda	485.21(377)	.00	1.29	.029	.91	.98	.045
İsrail	595.53(377)	.00	1.58	.038	.90	.98	.042
İsveç	734.07(377)	.00	1.95	.056	.86	.94	.058
İtalya	707.66(377)	.00	1.88	.052	.87	.93	.057
Japonya	554.64(377)	.00	1.47	.037	.89	.98	.047
Kanada	689.57(377)	.00	1.83	.036	.93	.96	.040
Kore	583.17(377)	.00	1.55	.038	.89	.97	.050
Macaristan	556.80(377)	.00	1.48	.037	.90	.98	.046
Norveç	496.84(377)	.00	1.32	.031	.91	.97	.047
Şili	553.27(377)	.00	1.47	.037	.90	.96	.050
Slovenya	460.58(377)	.00	1.22	.027	.91	.98	.049
Türkiye	827.03(377)	.00	2.19	.052	.88	.96	.049
Yeni Zelanda	796.14(377)	.00	2.11	.042	.92	.96	.042

Tablo 17'deki sonuçlara göre 17 ülkenin χ^2 /sd oranının 3'ün altında olduğu görülmektedir. Bu durum da bize χ^2 /sd sonucunun iyi olduğunu göstermektedir. Uyum indekslerine bakıldığında tüm ülkelere ait CFI değerleri 0.90'nın üzerinde yani iyi uyum göstermektedir. GFI değerleri Hong Kong, İsveç, İtalya, Japonya, Kore ve Türkiye için iyi uyuma yakın, diğer ülkeler için iyi uyum göstermiştir. Hong Kong, İsveç, İtalya hariç diğer ülkelere ait SRMR değerlerinin .05'in altında olduğu, Hong Kong'a ait SRMR değerinin .053, İsveç'e ait SRMR değerinin .058, İtalya'ya ait SRMR değerinin .057 olduğu görülmektedir. Uyum indeksleri genel olarak değerlendirildiğinde tüm ülkelere ait modellerin 17 OECD üyesi ülke için ayrı ayrı doğrulandığını kararı alınmıştır

Normallik: Tek deęişkenli ve çok deęişkenli normallik olmak üzere iki çeşit olan normallik veri setindeki deęişkenlere ait gözlemlerin normal dağılım göstermesi anlamına gelir. Yapısal Eşitlik Modellemesi çerçevesinde çok deęişkenli normalliğin hesaplanması güç olduğundan, tek deęişkenli normallik hesaplama yöntemlerinden olan basıklık ve çarpıklık katsayılarına bakılarak verilerin normalliğine karar verilmiştir (Weston ve Gore, 2006). Elde edilen sonuçlara bulgular bölümünde yer verilmiştir.

Çoklu Bağlantılılık Problemi: Bağımsız deęişkenler arasında güçlü ilişkilerin olmasına bağlantı (collinearity) veya çoklu bağlantı (multicollinearity) adı verilmektedir. Bağımsız deęişkenlerin birbirleriyle güçlü ilişkilerinin olduğu durumlarda çoklu bağlantı problemi vardır ve deęişkenlerin arasındaki korelasyonun genellikle .90 ve üzerinde olduğu durumlarda gözlenir. Yaygın bir şekilde .00 ile .30 arasındaki korelasyonlar 'düşük', .31 ile .70 arasındaki korelasyonlar 'orta' .71 ve üzeri korelasyon katsayıları 'yüksek' ilişkinin göstergesidir. Bu nedenle yüksek ilişkinin olduğu durumlarda dikkatli olunmalıdır. (Çokluk, Şekercioğlu ve Büyüköztürk, 2010).

Çoklu regresyon modeline yeni bağımsız deęişkenler ilave edildiğinde, R^2 'de önemli bir gelişme sağlanamazsa, çoklu doğrusal bağlantı problemi bulunabilir. (Albayrak, 2005). İki deęişken arasındaki basit korelasyon katsayısı manidar, buna rağmen kısmi korelasyon katsayısı manidar değilse, bu durum çoklu doğrusal bağlantı problemi için bir işaret olabilir. Bağımlı deęişkenler ile bağımsız deęişkenler arasında ilişki yoksa $R^2=0$ olacağından, $VIF=1$ olur [$VIF = 1/(1-R^2) = 1/(1-0) = 1$]. Bağımlı ve bağımsız deęişkenler arasında mükemmel bir ilişki varsa, $R^2=1$ olacağından $VIF=\infty$ olur [$VIF = 1/(1-R^2) = 1/(1-1) = \infty$]. $R^2=.90$ ise, $VIF=10$ olur [$VIF = 1/(1-0.9) = 10$]. Genel bir kural olarak VIF 'in 10'a eşit veya daha büyük ($VIF \geq 10$) olması durumunda, çoklu doğrusal bağlantı probleminin söz konusu olacağı vurgulanır (Depster, 1992: 683-684, Akt: Albayrak, 2005).

Bağımsız deęişkenler için tolerans deęerleri TV (Tolerance Value) = $1-R^2$ 'dir. Böylece daha küçük tolerans deęeri için daha büyük VIF deęerine karşılık gelir. Alanyazında tolerans deęerlerinin .10'dan büyük olması durumunda, deęişkenler arası çoklu bağlantı probleminin olmadığını göstermektedir. (Field, 2005; Mertler ve Vannata, 2005). $CN=[\text{Maksimum Özdeęer} / \text{Minimum Özdeęer}]$,

CI = [Maksimum Özdeğer / Minimum Özdeğer] / 2 formülleri kullanılarak hesaplanır. Gujarati (1995), CI değerinin 10 ile 30 arasında ve CN değerinin 100-1000 arasında olmasının ‘orta düzeyde’, CI değerinin 30’dan büyük ve CN değerinin 1000’den büyük olması ise ‘yüksek düzeyde’ çoklu doğrusal bağlantı probleminin göstergesidir (Akt: Albayrak, 2005). Değişkenlerin seçilmesi ve buna bağlı olarak modelin kurulması regresyon çözümlemesinde önemli yer tutar (Alma ve Vupa, 2008).

Tablo 21: OECD Ülkelerinin Pearson Korelasyon Değer Aralığı

Ülke	N	Pearson Korelasyon Değer Aralığı	Çoklu Bağlantı İstatistiği	
			VIF Değeri	Tolerans Değeri
Amerika	741	.021 - 0.195	1.075 - 1.546	0.717 - 0.923
Avusturalya	732	.042 - 0.258	1.043 - 1.761	0.707 - 0.884
Hong Kong	288	.100 - 0.339	1.078 - 2.049	0.709 - 0.927
İngiltere	342	.068 - 0.323	1.085 - 1.776	0.725 - 0.922
İrlanda	339	.015 - 0.300	1.116 - 1.573	0.719 - 0.896
İsrail	393	.110 - 0.312	1.197 - 1.936	0.711 - 0.835
İsveç	303	.039 - 0.268	1.153 - 1.818	0.709 - 0.867
İtalya	321	.029 - 0.234	1.151 - 1.877	0.719 - 0.869
Japonya	345	.020 - 0.270	1.163 - 1.827	0.736 - 0.912
Kanada	631	.016 - 0.138	1.097 - 1.560	0.722 - 0.965
Kore	384	.086 - 0.397	1.136 - 1.758	0.771 - 0.872
Macaristan	354	.101 - 0.298	1.078 - 1.656	0.726 - 0.947
Norveç	342	.014 - 0.125	1.157 - 1.656	0.726 - 0.947
Şili	348	.052 - 0.232	1.111 - 1.580	0.736 - 0.955
Slovenya	301	.024 - 0.159	1.183 - 1.553	0.807 - 0.915
Türkiye	441	.017 - 0.138	1.182 - 1.838	0.732 - 0.914
Yeni Zelanda	618	.019 - .0205	1.100 - 1.786	0.774 - 0.909
Toplam	7.223	-		
Değer Aralığı	-	.014 - 0.397	1.043 - 2.049	0.707 - 0.965

Bağımsız değişkenler arasındaki korelasyonlara Tablo 18'deki değerler için bakılmış ve değişkenler arasındaki korelasyonların genel olarak .014 ile .397 arasındaki değiştiği görülmüştür. Tolerans değerleri .01'den büyük olduğu durumlarda çoklu bağlantı probleminin olmadığı söylenebilir (Tabachnick ve Fidell, 2013). Tolerans değerleri .71 ile .97 arasında değişmektedir. Elde edilen değerler dikkate alındığında çoklu bağlantı problemi olmadığını söylemek mümkündür. Bu çerçevede, veri setine ilişkin tüm sayıtların karşılandığı söylenebilir.

Yapılan araştırmada veri setine ilişkin yapı geçerliliği kanıtlarını ortaya koymak amacıyla her bir grup için ayrı ayrı DFA yapılmıştır. DFA sonuçları bulgular bölümünde verilmiştir. Araştırmada, S-B χ^2 düzeltmesi farklı örneklem büyüklüğüne ve puan dağılımına sahip veri setlerinde, örneklem büyüklüğünün χ^2 üzerindeki etkisini azalttığı için, değeri olarak S-B χ^2 kullanılmış ve model uyumuna karar verilirken χ^2 , χ^2 /sd, CFI ve SRMR uyum indeksleri dikkate alınmıştır.

Ölçme modelleri karşılaştırılırken ise uyum indekslerindeki değişimler incelenerek, bu indeksler arasındaki farka bakılmıştır. İlk olarak karşılaştırılan modellerin, χ^2 istatistiği ve serbestlik dereceleri arasındaki (sd) fark ($\Delta\chi^2$ ve Δ sd) incelenir. Karşılaştırma sonucunda elde edilen farkın anlamlı olup olmadığı χ^2 tablosundaki değerler $p \leq .01$ ya da $p \leq .05$ düzeyinde karşılaştırılarak tespit edilir. Eğer χ^2 tablosundaki değer elde edilen sonuçtan büyükse modeller arasında anlamlı fark olduğu kabul edilir (Kline, 2005). Bu araştırmada χ^2 olarak S-B χ^2 değeri temel alındığı için fark derecesi belirlenirken T_s değerleri hesaplanmış ve modeller arasındaki farkın manidarlığı elde edilen T_s değeri ile χ^2 tablosundaki kritik değerler $p \leq .05$ düzeyinde karşılaştırılarak belirlenmiştir. Bununla birlikte model karşılaştırmalarında Δ CFI, Δ SRMR, Δ NFI, Δ GFİ, Δ RMSEA, Δ TLI gibi uyum indekslerinin değerlendirilmesi önerilmiştir (Cheng ve Reinsvold, 2002). Araştırmada Δ CFI ve Δ SRMR değerleri temel alınarak karşılaştırmalar yapılmıştır. (Çelebi, Korumaz, Güner ve Kaya, 2016).

Bu karşılaştırmalar için kabul düzeyleri,

$n > 300$ olduğunda metrik değişmezlik için $\Delta CFI \geq -.010$ ve $\Delta SRMR \geq .015$, güçlü ve katı faktöriyel değişmezliğin karşılaştırılmasında $\Delta CFI \geq -.010$ ve $\Delta SRMR \geq .010$.

$n < 300$ olduğunda metrik değişmezlik için $\Delta CFI \leq -.005$ ve $\Delta SRMR \geq .025$, güçlü ve katı faktöriyel değişmezliğe ait karşılaştırmalarda $\Delta CFI \geq -.005$, $\Delta SRMR \geq 0.005$ olarak belirlenmiştir (Chen, 2007).

Tüm durumlara bakıldıktan sonra ölçme değişmezliğine karar verme aşamasında dikkat edilmesi gereken bazı hususlar bulunmaktadır.

Şekercioğlu'na (2018) göre bunlardan ilki χ^2 ' de örnek sayısı arttıkça önem de artmaktadır, bu sebeple alternatif modeller, model çerçevesindeki faktör yapılarının veya kovaryans matrisleri arasındaki uyumun uygun değerlerde olup olmadığını araştırır. Bu değerler arasında ilk önce χ^2 değeri ve serbestlik derecesi karşılaştırılır. Bu bağlamda, daha sınırlandırılmış modelden elde edilen χ^2 değeri, daha az sınırlandırılmış modelden elde edilen χ^2 değeri ve serbestlik derecesi arasındaki "delta" değeri (delta fark demektir ve sembolü Δ 'dır) hesaplanır. Bu hesaplama ile $\Delta\chi^2$ ve Δsd değerleri belirlenmiş olur. Bu belirleme sonucunda elde edilen χ^2 değerinin anlamlılık düzeyi, dağılım tablosundaki kritik değerler karşılaştırılarak $p \leq .01$ veya $p \leq .05$ düzeyinde kontrol edilir (Byrne, 2010; Jöreskog, 1971; Kline, 2005; Lee ve Leung, 1982; Steiger, 2007; Van den Bergh ve Van Ranst, 1998).

Bu durumda, H_0 ve H_1 hipotezleri aşağıdaki formatta geliştirilebilir:

H_0 : Daha sınırlandırılmış model ile daha az sınırlandırılmış model arasında uyum açısından anlamlı bir fark yoktur.

H_1 : Daha sınırlandırılmış model ile daha az sınırlandırılmış model arasında uyum açısından anlamlı bir fark vardır.

Bundan dolayı, belirli bir Δ_{sd} seviyesindeki serbestlik derecesi farkları temelinde hesaplanan $\Delta\chi^2$, kritik tablo değerlerinden düşükse, H_0 kabul edilir. Başka bir söyleyiş ile iki model arasında uyum açısından anlamlı bir fark yoktur, bu nedenle araştırmacı, $\Delta\chi^2$ 'ye dayanarak ölçme değişmezliği hakkında karar verebilir. Öte yandan, belirli bir Δ_{df} seviyesindeki serbestlik derecesi farkları temelinde hesaplanan $\Delta\chi^2$ kritik tablo değerlerinden daha fazla ise, H_0 reddedilir. Bu nedenle, iki model arasında uyum açısından anlamlı bir fark vardır ve eğer bu fark alternatif hipotezin lehine ise, analizi yapan kişi ölçme değişmezliğinin $\Delta\chi^2$ temelinde verilmediğini kabul edebilir.

Ölçme değişmezliği analizinin uygulandığı birçok çalışmada, dağılımlar belirli toleranslar içinde normalden uzak olabilir. Büyük örneklerde normallik olmadığında, Satorra-Bentler düzeltmesinden elde edilen χ^2 değeri (S-B χ^2), örnekteki insan sayısı ve çıktı dağılımı normal olduğunda çıktı χ^2 'ye yakın değerler üretir. Çeşitli dağılımlar ve örneklem büyüklüklerindeki kovaryans yapı modellerini değerlendirmek için S-B χ^2 kullanılır ve epey güvenilir istatistiksel testtir (Byrne, 2006; Everitt ve Howell, 2005). Ölçme değişmezliği kanıtı elde etmek için gerçekleştirilen çoklu grup doğrulayıcı faktör analizi gibi S-B χ^2 yalnızca her grubun dağılımı normal dağılımdan uzaksa hesaplanabilir. Maksimum olabilirlik yöntemiyle gerçekleştirilen, çoklu grup doğrulayıcı faktör analizi iç içe modeller arasında ölçme değişmezliği sonucu açısından S-B χ^2 ölçekli fark için T_s değeri hesaplanmalıdır.

Ayrıca veri analizinin yorumlanmasında kullanılan C_0 , C_d ve T_s değerlerine ilişkin hesaplamalar aşağıda verilmiştir. Aşağıdaki eşitliklerde T_0 , iç içe model için normal maksimum olasılık χ^2 değeri, T_1 , karşılaştırma modeli (daha az kısıtlanmış model) için normal maksimum olasılık χ^2 değeridir ve c_d test düzeltmesi farkının derecesidir. d_0 iç içe geçmiş modelin serbestlik derecesi, d_1 karşılaştırma modelinin serbestlik derecesi, c_0 iç içe modelin düzeltme derecesi ve c_1 karşılaştırma modelinin düzeltme derecesidir.

T_0^* iç içe geçmiş modelin S-B χ^2 değeri, diğer yandan T_1^* karşılaştırma modelinin S-B χ^2 değeridir. S-B χ^2 fark dereceleri için hesaplanan T_s , dağılım tablosundaki kritik değerlerle karşılaştırılarak ölçme değişmezliğinin desteklenip desteklenmediği belirlenebilir (Brown, 2006; Satorra ve Bentler, 2011).

$$T_s = (T_0 - T_1) / C_d$$

$$C_d = [(d_0 * c_0) - (d_1 * c_1)] / (d_0 - d_1)$$

$$C_0 = T_0 / T_0^*$$

$$C_1 = T_1 / T_1^*$$

Bu formül son zamanlarda birçok araştırmada modeller arasındaki farkları değerlendirmenin yanı sıra, uygun indekslerden faydalanmak için yaygın bir alternatif olarak kullanılmaktadır. Cheung ve Rensvold'un görüşüne bakarsak (2002), önemsiz bir değer elde edilmesi durumunda boş hipotezi reddetmemeye özen gösterilmelidir. χ^2 , büyük örneklem için istatistiksel olarak hassas bir testtir fakat modelin bütünsel uyumu için hızlı ve uygun bir test değildir. Bu bağlamda, sonucu görebilmek için alternatif uygunluk indeksleri önerilmelidir. Karşılaştırmalı uygunluk indeksleri CFI, NNFI / TLI, RMSEA vb. en sık tavsiye edilenler indeksler arasındadır. Bu bağlamda bakıldığında, modelin genel uygunluğunu değerlendirmek ve raporlamak için pek çok uygunluk indeksinin bir arada kullanıldığı görülmektedir. Ölçme değişmezliğinin kanıtı için, modeller arasındaki farklılıklar, Δ RMSEA, Δ CFI, Δ SRMR, Δ NNFI, Δ IFI, Δ AIC, Δ EVCI, , Δ TLI gibi değerler aracılığı ile bulunur. (Şekercioğlu, 2018).

BÖLÜM IV

BULGULAR

Bu bölümde araştırmanın amaçları doğrultusunda kullanılan veri seti için TIMSS 2015 matematik başarısının, OECD üyesi ülkelere göre değişmezliğine ilişkin bulgulara ve söz konusu gruplara ait test istatistiklerine, normallik testlerine ve güvenilirlik katsayılarına yer verilmiştir.

4.1. OECD Üyesi Ülkeler İçin Ölçme Değişmezliğine İlişkin Bulgular

Bu başlık altında araştırmanın problemi olan TIMSS 2015 Türkiye uygulaması 8. sınıf matematik başarı testi OECD ülkeleri için ölçme değişmezliğinin sağlanıp sağlanamadığı sorusuna cevap aranmıştır.

TIMSS 2015 matematik başarısının, ülkeler arasında ölçme değişmezliğinin sağlanıp sağlanamadığını tespit etmek amacıyla yapılan çoklu grup doğrulayıcı faktör analizi (ÇGDFA) sonuçları Tablo 19’da verilmiştir.

Tablo 22: OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları

	S-B χ^2 (sd)	M.K.	$\Delta\chi^2$ (Δ sd)	χ^2 /sd	$\Delta\chi^2$ / (Δ sd)	CFI	Δ CFI	SRMR	Δ SRMR	Karar
Model 1^A	23382.36 (7337)	-	-	3.1869	-	.86	-	.069	-	
Model 2^B	17923.45 (6873)	M1-M2	5458.91 (464)	2.6078	.5791	.91	-.05	.048	.021	H₀: Ret
Model 3^C	10735.69 (6409)	M2-M3	7187.76 (464)	1.6750	.9328	.96	-.05	.042	.006	H₀: Ret
Model 4^D	23030.71 (6873)	M3-M4	-12295.02 (-464)	3.3508	- 1.6758	.86	.10	.061	-.019	H ₀ : Kabul

* p<.05

** Model karşılaştırmaları (M=Model)

A Yapısal (Biçimsel) Değişmezlik (Temel Model) (Faktör yükleri, faktör korelasyonları ve hata varyansları sabit)

B Zayıf Faktöriyel (Metrik) Değişmezlik (Faktör yükleri serbest, faktör korelasyonları ve hata varyansları sabit)

C Güçlü Faktöriyel (Ölçek / Scalar) Değişmezlik (Faktör yükleri ve hata varyansları serbest, faktör korelasyonları sabit)

D Katı Faktöriyel (Tam) Değişmezlik (Artık Varyans Değişmezliği) (Hata varyansları serbest, faktör yükleri ve faktör korelasyonları sabit)

İlk olarak, matematik testi puanı için gerçekleştirilen analiz sonucunda elde edilen uygunluk indeksleri incelendiğinde, uygunluk indekslerinin genel olarak kabul seviyesini karşıladığı söylenebilir.

Tablo 23: OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları

	S-B χ^2 (sd)	M.K.	$\Delta\chi^2(\Delta sd)$	χ^2/sd	$\Delta\chi^2/(\Delta sd)$	CFI	ΔCFI	SRMR	$\Delta SRMR$	Karar
Model	23382.36	-	-	3.1869	-	.86	-	.069	-	-
1^A	(7337)									

* p<.05

** Model karşılaştırmaları (M=Model)

A_{Yapısal (Biçimsel) Değişmezlik (Temel Model) (Faktör yükleri, faktör korelasyonları ve hata varyansları sabit)}

Tablo 20’de görüldüğü üzere ülkelerin sahip olduğu faktör yükleri, faktör korelasyonları ve hata varyanslarının sabit çıktığı Model 1 (yapısal değişmezlik)’de kovaryans matrislerinde faktör yüklerinin buna bağlı olarak faktör korelasyonlarının ve daha sonrasında hata varyanslarının tüm gruplar için eşit olup olmadığının test edilmesi sonucunda S-B χ^2 (sd) sonucunun 23382.36(7337) olduğu, S-B χ^2 ve serbestlik derecesi oranının $\chi^2/sd=3.18695$ çıktığı, 5’in altında ve 3’e yakın olması sebebi ile iyi, CFI değerinin .86 olduğu yani .90’a yakın olması sebebi ile kabul edilebilir düzeye yakın ve SRMR değerinin .069 olduğu yani .1’in altında .05 ile .08 arasında olması sebebi ile iyi uyum gösterdiği görülmektedir. Uyum indekslerinin değerlerinin kabul edilebilir düzeylerde olduğu ve modelin doğrulandığı söylenebilir. Bu durumda yapısal değişmezliğin sağlandığını söylemek mümkündür.

Tablo 24: OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları

	S-B χ^2 (sd)	M.K.	$\Delta\chi^2(\Delta sd)$	χ^2/sd	$\Delta\chi^2/(\Delta sd)$	CFI	ΔCFI	SRMR	$\Delta SRMR$	Karar
Model	17923.45	M1-M2	5458.91	2.6078	.5791	.91	-.05	.048	.021	H₀:
2^B	(6873)		(464)							Ret

* p<.05

** Model karşılaştırmaları (M=Model)

A_{Yapısal} (Biçimsel) Değişmezlik (Temel Model) (Faktör yükleri, faktör korelasyonları ve hata varyansları sabit)

B_{Zayıf Faktöriyel} (Metrik) Değişmezlik (Faktör yükleri serbest, faktör korelasyonları ve hata varyansları sabit)

Faktör yükleri, faktör korelasyonları ve hata varyanslarının sabit olduğu Model 1 (yapısal değişmezlik) ile faktör korelasyonları ve hata varyanslarının sabit, faktör yüklerinin serbest bırakıldığı Model 2 (zayıf faktöriyel değişmezlik) karşılaştırıldığında S-B χ^2 (sd) sonucunun 17923.45(6873) olduğu, χ^2/sd oranının 2.6078'e düştüğü yani oranın 3'ün altına düşerek daha da iyileştiği görülmektedir. $\Delta\chi^2(\Delta sd)$ 'nin 5458.91(464) olduğu, böylece $\Delta\chi^2/\Delta sd$ oranının da .5791 olduğu görülmektedir. S-B χ^2 fark derecesinin manidar olup olmadığını tespit etmek için T_s değeri hesaplanmış ve bu değer 7161.3943 olarak bulunmuştur. T_s değerinin χ^2 dağılımı tablosundaki kritik değerden büyük olduğu tespit edilmiştir, $\chi^2_{fark}(464)=515.2183$, p<.05. Bu durumda modeller arasındaki farkın manidar olduğunu ifade etmek mümkündür. CFI değerinin .86'dan .91'e yani .90 ile 1 arasına yükselmesi değer giderek daha da iyileştiği ve iyi uyum gösterdiği söylenebilir. SRMR değerinin .069'dan .048'e düşmesi değer giderek daha da iyileştiği göstermekte iyi uyum gösterdiğini ifade etmektedir. Ayrıca ΔCFI değerinin <-.01 ve $\Delta SRMR \geq .025$ kabul düzeylerini sağlayıp sağlamadığına bakılmıştır. ΔCFI değerinin -.05 (<-.01) olması dolayısı ile manidar bir fark vardır. $\Delta SRMR$ değerinin .021 (<.025) olması dolayısı ile farkın manidar olmadığı ifade edilebilir.

Model 1 ve Model 2 karşılaştırmasında Model 2’de genel olarak değerlerin iyileştiği, diğer bir ifadeyle Model 1’de sağlanan ölçme değişmezliğinin daha da iyileştiği ve Model 1 ile Model 2 yani yapısal değişmezlik ile zayıf faktöriyel (metrik) değişmezlik arasındaki değerler genel olarak incelendiğinde, manidar bir fark olduğu söylenebilir. Bu durumda Model 1 ve Model 2 arasında manidar bir fark olması sebebi ile H_0 ret edilmiştir.

Tablo 25: OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları

	S- B_{χ^2} (sd)	M.K.	$\Delta\chi^2$ (Δ sd)	χ^2 /sd	$\Delta\chi^2$ / Δ (sd)	CFI	Δ CFI	SRMR	Δ SRMR	Karar
Model 3^C	10735.69 (6409)	M2-M3	7187.76 (464)	1.6750	.9328	.96	-.05	.042	.006	H₀:Ret

* $p < .05$

** Model karşılaştırmaları (M=Model)

B Zayıf Faktöriyel (Metrik) Değişmezlik (Faktör yükleri serbest, faktör korelasyonları ve hata varyansları sabit)

C Güçlü Faktöriyel (Ölçek / Scalar) Değişmezlik (Faktör yükleri ve hata varyansları serbest, faktör korelasyonları sabit)

Faktör yükleri serbest, faktör korelasyonları ve hata varyanslarının sabit olduğu Model 2 (zayıf faktöriyel değişmezlik) ile faktör yükleri ve hata varyanslarının serbest, faktör korelasyonlarının sabit olduğu Model 3 (güçlü faktöriyel değişmezlik) karşılaştırıldığında, S- B_{χ^2} (sd) sonucunun 10735.69(6409) olduğu, $\Delta\chi^2$ (Δ sd)’nin 7187.76(464) olduğu, böylece χ^2 /sd oranının 1.6750’e düştüğü görülmektedir. $\Delta\chi^2$ / Δ sd oranının 1.6750 olduğu ve oldukça iyileştiği tespit edilmiştir. S- B_{χ^2} fark derecesinin manidar olup olmadığını tespit etmek için T_s değeri hesaplanmış ve bu değer 13212.7076 olarak bulunmuştur. T_s değerinin χ^2 dağılımı tablosundaki kritik değerden büyük olduğu tespit edilmiştir, $\chi^2_{\text{fark}}(464) = 515.2183$, $p < .05$. Bu durumda modeller arasındaki farkın manidar olduğu ifade etmek mümkündür. CFI değerinin .91’den .96’ya yani .90 ile 1 arasında 1’e daha yakın bir değere yükselmesi değer giderek daha da iyileştiğini göstermekte iyi uyum gösterdiğini ifade etmektedir. SRMR değerinin .048’den .042’ye düşmesi değer .00 ile .05 arasında bir değere gelmesi ve .00’a giderek yaklaşması sebebi ile değer giderek daha da iyileştiği göstermekte iyi uyum gösterdiğini ifade etmektedir. Ayrıca

Δ CFI değerinin $<-.01$ ve Δ SRMR $>.01$ kabul düzeylerini sağlayıp sağlamadığına bakılmıştır. Δ CFI değeri $-.05$ ($<-.01$) olması dolayısı ile manidar bir fark vardır. Δ SRMR değerinin $.006$ ($<.01$) olması dolayısı ile manidar bir fark yoktur denilebilir. Model 2 ve Model 3 karşılaştırmasında Model 3'te sağlanan ölçme değişmezliğinin daha da iyileştiği ve Model 2 ve Model 3 yani zayıf faktöriyel (metrik) değişmezlik ile güçlü faktöriyel değişmezlik arasındaki değerler ($S-B\chi^2(sd)$, χ^2/sd , $\Delta\chi^2/(\Delta sd)$, CFI, SRMR, Δ CFI, Δ SRMR, T_s değeri ve χ^2_{fark} kıyaslaması) genel olarak incelendiğinde, manidar bir fark olduğu ifade edilebilir. Bu durumda Model 2 ve Model 3 arasında manidar bir fark olması sebebi ile H_0 ret edilmiştir.

Tablo 26: OECD Üyesi Ülkelerin Matematik Başarı Testi Puanları için Çoklu Grup Doğrulayıcı Faktör Analizi Sonuçları

	S- $B\chi^2(sd)$	M.K.	$\Delta\chi^2(\Delta sd)$	χ^2/sd	$\Delta\chi^2/(\Delta sd)$	CFI	Δ CFI	SRMR	Δ SRMR	Karar
Model	23030.71	M3-M4	-12295.02	3.3508	-1.6758	.86	.10	.061	-.019	H_0 :
4^D	(6873)		(-464)							Kabul

* $p < .05$

** Model karşılaştırmaları (M=Model)

^CGüçlü Faktöriyel (Ölçek / Scalar) Değişmezlik (Faktör yükleri ve hata varyansları serbest, faktör korelasyonları sabit)

^DKatı Faktöriyel (Tam) Değişmezlik (Artık Varyans Değişmezliği) (Hata varyansları serbest, faktör yükleri ve faktör korelasyonları sabit)

Faktör yükleri ve hata varyanslarının serbest, faktör korelasyonlarının sabit olduğu Model 3 (güçlü faktöriyel değişmezlik) ve hata varyanslarının serbest, faktör yükleri ve faktör korelasyonlarının sabit olduğu Model 4 (katı faktöriyel değişmezlik) karşılaştırıldığında, $S-B\chi^2(sd)$ sonucunun 23030.71(6873) ve $\Delta\chi^2(\Delta sd)$ 'nin -12295.02(-464) olduğu, χ^2/sd oranının 3.3508'e yükseldiği yani oranın 3 ile 5 arasında bir değer olarak yeterli uyum sağladığı ancak Model 2 ve Model 3 kıyaslamasına göre oldukça kötüleştiği görülmektedir. $\Delta\chi^2/\Delta sd$ oranının -1.6758 olduğu ve oldukça kötüleştiği tespit edilmiştir. $S-B\chi^2$ fark derecesinin manidar olup olmadığını tespit etmek için T_s değerine bakılmış ve değerlerinin oldukça kötü olması dolayısı ile hesaplanamamış, ayrıca χ^2_{fark} değeri değer aralığı dışında çıkmış ve hesaplanamamıştır. Bu sebepten T_s değerinin χ^2 dağılımı tablosundaki değerle kıyaslaması yapılmamış ve sonuç vermemiştir. Bu durumda modeller arasındaki

farkın manidar olmadığını ifade etmek mümkündür. CFI değerinin .96'dan .86'ya bir başka deyişle .90 değerinin altına düşmesi uyumun kötüleştiğini göstermektedir. SRMR değerinin .042'den .0461'e yükselmesi, değer .00 ile .05 arasında olmasına rağmen Model 3'e göre değer kötüleşmesi sebebi ile iyi uyum göstermediğini ifade etmektedir. Ayrıca ΔCFI değerinin $<-.01$ ve $\Delta SRMR >.01$ kabul düzeylerini sağlayıp sağlamadığına bakılmıştır. ΔCFI değerinin .10 ($>-.01$) manidar bir fark olmadığı ve değer biraz daha kötüleştiği tespit edilmiştir. $\Delta SRMR -.019$ ($<.01$) manidar bir fark olmadığı yani SRMR değerlerinin de kötüleştiği, sonuçların genel olarak kötüleştiği ve manidar bir fark gözlenmediği tespit edilmiştir. Model 3 ve Model 4 karşılaştırmasında, Model 4'te genel olarak değerlerin kötüleştiği bu durumda Model 3 ve Model 4 yani güçlü faktöriyel değişmezlik ile katı faktöriyel değişmezlik arasında manidar bir farkın olmadığını söylemek mümkündür. Bu durumda Model 3 ve Model 4 arasında manidar bir fark olmaması sebebi ile H_0 kabul edilmiştir.

Tüm model karşılaştırmalarına ait bulgular değerlendirildiğinde karşılaştırılan dört model içerisinde en iyi çalışan model güçlü faktöriyel değişmezlik (Model 3) modelidir. Bu doğrultuda yapılan analizler sonucunda TIMSS 2015 matematik başarısının faktör yapısının OECD üyesi ülkeler için eşit olmadığı ve ölçme değişmezliğini sağlamadığı kararı alınmıştır.

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

Bu bölümde araştırmanın bulgularına ilişkin sonuçlara ve bu sonuçlara dayalı olarak önerilere yer verilmiştir.

5.1. Sonuç ve Tartışma

TIMSS 2015 matematik başarısının OECD üyesi ülkeler için ölçme değişmezliğine ilişkin sonuçlar aşağıda verilmiştir.

TIMSS 2015 matematik başarısı OECD üyesi ülkeler arasında ölçme değişmezliğinin sağlanıp sağlanmadığını belirlemek için test istatistikleri, normallik ve güvenilirlik hesaplamaları yapılmıştır. Bu hesaplamalar sonucunda veri setinin normale yakın bir dağılım gösterdiği kabul edilmiştir. Güvenilirlik hesaplamaları sonucunda ise ülkeler ilişkin KR-20 iç tutarlılık güvenilirlik katsayısının bazı ülkelerin kabul düzeyinin altında sonuç gösterdiği, bazı ülkelerin kabul düzeyinde değer gösterdiği tespit edilmiştir.

17 ülke arasında kovaryans matrislerinin uyum düzeyini belirleyebilmek için yapılan kovaryans matrislerinin eşitliği testi sonucunda, ülkeler arasında kovaryans matrislerine ait modelin iyi uyum gösterdiği sonucuna varılmıştır. TIMSS 2015 matematik başarısının OECD üyesi ülkelere göre ölçme değişmezliğinin çoklu-grup doğrulayıcı faktör analizi ile dört model temel alınarak test edilmesi sonucunda temel model olarak kurulan Model 1'e ait uyum indekslerinin kabul edilebilir düzeyde olduğu ve modelin doğrulandığı tespit edilmiştir. Bu durumda OECD ülkelerine göre yapısal değişmezliğin sağlandığı ifade edilebilir. Model 1'e alternatif olarak kurulan Model 2, Model 3 ve Model 4'ün yuvalanmış model (nested) temel alınarak karşılaştırılması sonucunda, Model 2 ve Model 3'e ait uyumun oldukça iyileştiği ancak manidar farklar olması sebebi ile ölçme değişmezliği olmadığı belirlenmiştir. Model 3 güçlü faktöriyel değişmezlik ve Model 4 katı faktöriyel değişmezlik arasında manidar bir farkın olmadığını söylemek mümkündür.

Tüm model karşılaştırmalarına ait bulgular değerlendirildiğinde karşılaştırılan dört model içerisinde en iyi çalışan model güçlü faktöriyel değişmezlik (Model 3) modelidir. Bu doğrultuda TIMSS 2015 matematik başarısının faktör yapısının OECD üyesi ülkeler için eşit olmadığı ve ölçme değişmezliğini sağlamadığı kararı alınmıştır. Ölçme değişmezliğinin en iyi güçlü faktöriyel değişmezlik düzeyinde sağlandığı sonucuna ulaşılmıştır. TIMSS 2015 Matematik başarı değişmezliğinin 11 no'lu kitapçık üzerinden değerlendirildiği bu araştırmanın bulguları, OECD üyesi ülkeler arasında ölçme değişmezliğinin sağlanamadığını göstermektedir.

Ülke değişkenine ilişkin sonuçlar değerlendirildiğinde, geniş ölçekli sınavlarda kullanılan ölçme araçlarının farklı dillere ve kültürlere uyarlamasının sınırlılık oluşturduğu söylenebilir.

1. Bazı araştırma sonuçları farklı ülkelerde uygulanan geniş ölçekli çalışmalarda farklı ülkeler arasında ölçme değişmezliğinin sağlandığını raporlaştırmışlardır (Marsh ve diğerleri, 2006; Akyıldız, 2009).
2. Bazı çalışmalar farklı dil ve kültürlere sahip bireyler arasında kullanılan ölçme araçlarında ölçme değişmezliğinin sağlanamadığını raporlamıştır (Ercikan ve Koh 2005; Öğretmen, 2006; Wu ve diğerleri 2007; Marsh ve diğerleri, 2013; Kıbrıslıoğlu, 2015; Karakoç Alatlı, 2016).

Bununla birlikte, geniş ölçekli sınav uygulamalarının sonuçları değerlendirilirken ölçme değişmezliğinin sağlanamıyor olması, özellikle ülkeler arası karşılaştırmalar ve yorumlamalar yapılırken göz önünde bulundurulmalıdır. Tüm sonuçlar değerlendirildiğinde, TIMSS 2015 Türkiye uygulamasında matematik başarısına ait elde edilen sonuçların farklı ülkeler yani, farklı kültür ve dillere sahip katılımcılardan elde edilen sonuçların karşılaştırılabilirliği ve buna bağlı olarak yapılan yorumlar ölçme değişmezliğinin sağlanamaması yani uygulanan ölçme aracının tüm katılımcılar için aynı anlamı ifade etmemesi nedeniyle tartışmaya açık hale gelmektedir.

5.2. Öneriler

Araştırma süreci ve elde edilen sonuçlar değerlendirildiğinde, uygulamaya yönelik ve araştırmacıya yönelik olmak üzere iki grupta öneriler verilebilir. Bu öneriler şu şekildedir:

a) Uygulamaya yönelik öneriler:

1. Araştırma sadece TIMSS 2015'e katılan OECD üyesi ülkelere ait veri seti üzerinden yürütülmüştür. Gelecek çalışmalarda bu araştırmada yer almayan diğer ülkelerin de bulunduğu ve demografik değişkenlerin dâhil edildiği araştırmalar yapılabilir.
2. Bu araştırma yalnızca TIMSS 2015 uygulaması 11 no'lu kitapçıkta bulunan 8. sınıf matematik başarısına ilişkin maddeler üzerinden yürütülmüştür. Sonuçların genellenebilmesi için gelecek araştırmalarda farklı kitapçıklarda ya da farklı başarı alanlarına (okuma becerileri, anketler ve fen başarısı) ilişkin ölçme değişmezliği incelenebilir.
3. Bu araştırmada ölçme değişmezliği çoklu-grup doğrulayıcı faktör analizi kullanılarak test edilmiştir. Gelecek araştırmalarda farklı yöntemler kullanılarak ölçme değişmezliği test edilirse, yöntemler arası karşılaştırmalar yapılması ve hangi yöntemin daha etkili olduğuna karar verilmesi mümkün olabilir.
4. Araştırmada ülkeler arasında ölçme değişmezliğinin sağlanamadığının belirlenmiş olması, farklı gruplar (cinsiyet, dil vb.) arasında karşılaştırmalar yapılmadan önce yapılan yorumların anlamlı ve daha sağlıklı olabilmesi için uygulanan ölçme aracına ilişkin ölçme değişmezliği çalışmalarının yapılmasının oldukça önem taşıdığını göstermiştir.

b) Uygulayıcıya yönelik öneriler:

1. Her ülkenin kendine has yaşanmışlığı, tarihi ve kültürel değer ve gelenekleri olduğu bir gerçektir. Bu sebeple madde yazımı ve çeviri komisyonlarında her bir ilgili kültürdeki temsilcilerin de yer alması gerekmektedir. Sadece birebir çevirinin yeterli olmadığı bilinmelidir.
2. Bu araştırmaya göre özellikle ortalamalar kullanılarak yapılan hipotez testlerinin bulgularının tartışmalı olduğu bilinmelidir. Bu bulgulara dayalı olarak verilecek eğitimsel ve politik kararların da tartışmalı olacağı gözden kaçırılmamalıdır.
3. Geniş ölçekli olarak yürütülen ölçme ve değerlendirme uygulamalarında, ölçülen özelliğin ya da yapılan uygulamanın türüne göre ölçme değişmezliğinin farklılık gösterdiğini, elde edilen sonuçlara göre sonucu bulma, karşılaştırma ve yorumlama yapılmadan önce ölçme değişmezliği çalışmalarının titizlikle yapılması gerektiği göz ardı edilmemelidir.

KAYNAKÇA

- Albayrak, A. S. (2005). Çoklu doğrusal bağlantı halinde en küçük kareler tekniğinin alternatifini yanlı tahmin teknikleri ve bir uygulama. *ZKÜ Sosyal Bilimler Dergisi*, 1(1), 2005, 105-126. [Çevrim-içi: <https://www.academia.edu/7252199/D4>. Albayrak Ali Sait 2005] Erişim Tarihi: 4 Temmuz 2018.
- Alma, Ö. ve Vupa, Ö. (2008). Doğrusal regresyon çözümlemesinde çoklu bağlantı probleminin sapan değer içeren küçük örnekleme incelenmesi. *Sdü Fen Edebiyat Fakültesi Fen Dergisi (E-Dergi)*, 31, 97-107.
- Alma, Ö. ve Vupa, Ö. (2008). Regresyon analizinde kullanılan en küçük kareler ve en küçük medyan kareler yöntemlerinin karşılaştırılması. *Sdü Fen Edebiyat Fakültesi Fen Dergisi (E-Dergi)*, 3(2), 2008, 219-229
- Anastasi, A. (1968). *Psychological testing* (Third Edition). Oxford: Macmillan
- Asil, M., & Brown, G. T. L. (2015). Comparing OECD PISA reading in English to other languages: Identifying potential sources of non-invariance. *International Journal of Testing*. Advance online publication. [Çevrim-içi: <https://www.tandfonline.com/doi/abs/10.1080/15305058.2015.1064431>] Erişim Tarihi: 14 Temmuz 2018.
- Ayvallı, M. (2016). *PISA 2012 matematik okuryazarlığı testinin ölçme değişmezliğinin incelenmesi. An investigation into the measurement invariance of PISA 2012 mathematical literacy test*. Yayınlanmamış Yüksek Lisans Tezi, Akdeniz Üniversitesi, Antalya, 2016.
- Babakus, E. (1985). *The sensitivity of maximum likelihood factor analysis given violations of interval scale and multivariate normality*. Doctoral dissertation, The University of Alabama. [Çevrim-içi: <https://elibrary.ru/item.asp?id=7467190>] Erişim Tarihi: 14 Eylül 2018.

- Bahadır, E. (2012). *Uluslararası öğrenci değerlendirme programına (PISA 2009) göre Türkiye'deki öğrencilerin okuma becerilerini etkileyen değişkenlerin bölgelere göre incelenmesi*. Yayınlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara, 2012.
- Bailey, K.D. (1982). *Methods of social research*. (Second Edition). New York: The Free Press.
- Balcı, A. (2013). *Sosyal bilimlerde araştırma yöntem, teknik ve ilkeler* (10. baskı). Ankara: Pegem Akademi
- Balım, A.G, Evrekli, E., İnel, D. ve Deniz, H. (2009). Türkiye'de PISA 2006'da ki durumu üzerine bir inceleme: Fen bilimleri yeterlilik düzeyinin bilgi ve iletişim teknolojilerinin kullanımına göre değerlendirilmesi. *e-Journal of New World Sciences Academy*, 4(3), 2009.
- Başusta, N.B. ve Gelbal, S. (2015). Gruplararası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: PISA öğrenci anketi örneği. Examination of measurement invariance at groups' comparisons: a study on PISA student questionnaire. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30(4), 80-90.
- Bayar, V. ve Bayar, S. A. (2013). TIMSS 2011 matematik başarıları ulusal değerlendirme raporu. *Türk Eğitim Sendikası TIMSS 2011 Matematik Başarıları Ulusal Değerlendirme Raporu*, Ankara. [Çevrim-içi: http://www.turkegitimsen.org.tr/upload_doc/00_2012_y/00_yok/TIMSS.docx] Erişim tarihi: 4 Nisan 2018.
- Bayır, E., Çakıcı, Y. ve Atalay, Ö. (2016). Fen bilimleri öğretmenlerinin bilimin doğasına ilişkin görüşleri: Bilişsel harita örneği. *Kastamonu Eğitim Dergisi*, 24(3), 1419-1436.
- Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, 88(3), 588.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Newyork: Guilford Publications, Inc.

- Büyüköztürk, Ş. (2002). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem Akademi.
- Büyüköztürk, Ş., Çakmak, E., Akgün, Ö.E., Karadeniz, Ş. ve Demirel, F. (2008). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi.
- Büyüköztürk, Ş., Çakan, M., Tan, Ş. ve Atar, H. Y. (2014). *TIMSS 2011 Ulusal Matematik ve Fen Raporu. 8. Sınıflar*. Ankara: MEB YEGİTEK Yayınları. [Çevrim-içi: <http://timss.meb.gov.tr/wp-content/uploads/TIMSS-2011-8-Sinif.pdf>] Erişim Tarihi: 4 Nisan 2018.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley
- Bollen, K. A. (1990). Overall fit in covariance structure models: two types of sample size effects. *Psychological bulletin*, 107(2), 256.
- Byrne, B.M. (1998). *Structural equation modeling with LISREL, PRELIS and SIMPLIS: basic concepts, applications and programming*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Byrne, B. M. (2006). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. (Second Edition). California: Sage Publications, Inc.
- Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, 20(4), 872-882.
- Byrne, B. M. (2013). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press.
- Byrne, B. M. & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, 34(2), 155-175.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56(2), 81.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling*, 14(3), 464-504.

- Cheung, G. W. & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal Of Cross-Cultural Psychology*, 31(2), 187-212.
- Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, 9(2), 233-255.
- Cheng, Y. & Reinsvold, R. B. (2002). *Factors Associated with foreign language writing anxiety*. Retrieved 25, 2002. doi: <https://doi.org/10.1111/j.1944-9720.2002.tb01903.x>
- Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., & Schwartz, S.H. (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: a cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology*, 5, 982.
- Crocker, L. and Algina, J. (1986). *Introduction to classical and modern test theory*. Toronto: Holt, RineHart, and Winston, Inc.
- Cronbach, L. J. and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281.
- Çelebi, N., Korumaz, M., Güner ve Kaya (2016). Teachers loyalty to their supervisors and organizational commitment. *Educational Research and Reviews*, 11(12), 1161-1167.
- Çelebi, N., Güner, H., Kaya G. T., ve Korumaz, M. (2014). Neoliberal eğitim politikaları ve eğitimde fırsat eşitliği bağlamında uluslararası sınavların (PISA, TIMSS ve PIRLS) analizi. *Tarih Kültür ve Sanat Araştırmaları Dergisi*, 3(3), 33-75. doi: 10.5897/ERR2016.2808 (Yayın No: 2864277)
- Çepni, S., Ayas, A., Johnson, D. ve Turgut, F. (1997). *Fizik Öğretimi*. Ankara: YÖK/Dünya Bankası, Milli Eğitimi Geliştirme Projesi
- Çokluk, Ö., Şekercioğlu, G. ve Büyüköztürk, Ş. (2010-2012). *Sosyal Bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları*. Ankara: Pegem Akademi.

- Dempster, A. P., M. Schatzoff & N. Wermuth (1977). A simulation study of alternatives to ordinary least square. *Journal of American Statistical Association*, 72, 77-91.
- EARGED (2014). *TIMSS 2011 ulusal matematik ve fen raporu: 8. sınıflar*. [Çevrim-içi: <http://timss.meb.gov.tr/wp-content/uploads/TIMSS-2011-8-Sinif.pdf>]
Erişim tarihi: 01 Eylül 2018.
- Ergül, H. (1999). Uzaktan öğretimde kalite verimlilik ve üretkenlik. *Anadolu Üniversitesi Açık Öğretim Fakültesi Kurgu Dergisi*, 16, 283-296.
- Ertürk, Z., Erdiñç Akan, O. (2018). TIMSS 2015 matematik başarısı ile ilgili bazı deęişkenlerin cinsiyete göre ölçme deęişmezlięinin incelenmesi. *Kuramsal Eęitimbilim Dergisi, UBEK-2018, 204-226, Aralık 2018 Journal of Theoretical Educational Science, ICSE-2018, 204-226, December 2018* [Çevrim-içi: <http://www.keg.aku.edu.tr>] Erişim Tarihi: 08 Temmuz 2018
- Everitt, B. S. & Howell, D. C. (2005). *Encyclopedia of statistics in behavioral sciences*. Chichester: John Wiley & Sons, Ltd.
- Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural equation modeling: a multidisciplinary journal*, 6(1), 56-83.
- Floyd, F. J. & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological assessment*, 7(3), 286.
- Gierl, M.J. (2000). Construct equivalence on translated achievement tests. *Canadian Journal of Education*, 25(4), 280-296.
- Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical care*, 44(11), 78-94.

- Gülleroğlu, H. D. (2017). PISA 2012 matematik uygulamasına katılan Türk öğrencilerin duyuşsal özelliklerinin cinsiyete göre ölçme deęişmezlięinin incelenmesi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 37, 1, GEFAD / GUJGEF 37, 1: 151-175 (2017) Yayına Kabul Tarihi: 27.12.2016
- Güngör, M., Bulut, Y. (2008). *Ki-kare testi üzerine*. Fırat Üniversitesi, Fen – Edebiyat Fakültesi, Matematik Bölümü, Doęu Anadolu Bölgesi Araştırmaları, 23119 – ELAZIĞ.
- Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! Testing! One, two, three–testing the theory in structural equation models!. *Personality and Individual Differences*, 42(5), 841-850.
- Hong-Nam, K & Leavell, A. (2007). A comparative study of language learning strategy use in an EFL context: Monolingual Korean and bilingual Korean-Chinese university students. *Asia Pasific Education Review*, 8(1), 71-78.
- Hu, L.T. & Bentler, P.M. (1999). Cut off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- Huck, S.W.(2007). *Reading statistics and research*. Boston: Pearson Education.
- Joppe, M. (2000). *The Research Process*. Retrieved February 25, 1998. [Çevrim-içi: <http://www.rverson.ca/~mjoppe/rp.htm>] Erişim Tarihi: 8 Temmuz 2018
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409-426.
- Jöreskog, K. G. and Sörbom, D. (1993). *LISREL 8: structural equation modeling with the simplis command language*. USA: Scientific Software International, Inc.
- Karakoç Atatlı, B. (2016). *Uluslararası öğrenci değerlendirme programı (PISA - 2012) okuryazarlık testlerinin ölçme deęişmezlięinin incelenmesi*. (Yayımlanmamış Yüksek Lisans Tezi). Ankara Üniversitesi, Ankara.
- Kelley, T. L. (1927). *Interpretation of educational measurements*. New York: Macmillan.

- Kenny, D. A. & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural equation modeling*, 10(3), 333-351.
- Kıbrıslıođlu, N. (2015). *PISA 2012 matematik öğrenme modelinin kültürlere ve cinsiyete göre ölçme değışmezliđinin incelenmesi: Türkiye-Çin, (Şangay)-Endonezya örneđi*. (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. (Second Edition) New York: Guilford Press.
- Kođar, H., Yılmaz Kođar, E. (2015). Comparison of different estimation methods for categorical and ordinal data in confirmatory factor analysis. *Eđitimde ve Psikolojide Ölçme ve Deđerlendirme Dergisi Journal of Measurement and Evaluation in Education and Psychology*, 6(2), 2015, 351-364
- Kođar, H.ve Yılmaz Kođar, E. (2017), Öğretmenlerin matematik konularına yönelik hazırlık düzeylerinin matematik başarısı ile iliřkisi: TIMSS 2015 Türkiye ve Singapur örneđi. *Başkent Üniversitesi Eğitim Dergisi*, 4(2), 2017, 108-121,
- Korkmaz, H. (2004). *Fen ve teknoloji eğitiminde alternatif deđerlendirme yaklaşımları*, Yeryüzü Yayınevi, Ankara.
- Lee, S. Y. & Leung, T. K. (1982). Covariance structure analysis in several populations. *Psychometrika*, 47, 297-308.
- LoBiondo-Wood, G. & Haber, J. (2014). *Nursing research: Methods and critical appraisal for evidence-based practice*. Missouri: Elsevier Mosby.
- Malone, J. & Haines, D. (2003). Australian students' performance in the 1999 TIMSS repeat. *Australian Senior Mathematics Journal*, 17(1), 34-38.
- Maruyama, G. M. (1998). *Basics of structural equation modeling* (First Edition). CA: Sage Publications. Inc.

- Mark, B. A. & Wan, T.T.H. (2005). Testing measurement equivalence in a patient satisfaction instrument. *Western Journal of Nursing Research*, 27(6), 772-787. [Çevrim-içi: <http://wjn.sagepub.com/cgi/reprint/27/6/772.pdf>]. Erişim Tarihi: 14 Ekim 2005
- Martin, M.O. & Mullis, I.V.S. (Eds.) (2012). *Methods and procedures in TIMSS and PIRLS 2011: Characteristics of national samples*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. [Çevrim-içi: http://timss.bc.edu/methods/pdf/T11_Characteristics.pdf] Erişim tarihi: 16 Şubat 2015.
- Martin, M.O. & Mullis, I.V.S. (Eds.) (2012). *Methods and procedures in TIMSS and PIRLS 2011: Stratified Two-Stage Cluster Sample Design*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. [Çevrim-içi: http://timss.bc.edu/methods/pdf/T11_Characteristics.pdfhttp://timss.bc.edu/methods/pdf/2Stage_Sample_Design.pdf] Erişim tarihi: 16 Şubat 2015.
- Matsumoto, D. & Van de Vijver, F. J. (Eds.). (2011). *Cross-cultural research methods in psychology*. NY:Cambridge University Press.
- McDonald, R.P. & Ho, M.-H.R. (2002). Principles and practice in reporting statistical equation analyses. *Psychological Methods*, 7(1), 64-82.
- McIntosh, C. (2006), Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett (2007). *Personality and Individual Differences*, 42(5), 859-67.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543.
- Mertler, C.A. & Vannatta, R. A. (2005). *Advanced and multivariate statistical methods. practical application and interpretation*. (Third Edition). CA: Pyrczak Publishing.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). Washington, DC: American Council on Education and National Council on Measurement in Education.

- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Miles, J. & Shevlin, M. (2007). A time and a place for incremental fit indices. *Personality and individual differences*, 42(5), 869-874.
- Milfont, T. L. & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111-121.
- Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance*. New York, NY: Routledge.
- Moraes, C.L. & Reichenheim, M.E. (2002). Cross-Cultural measurement equivalence of the revised conflict tactics scales (CTS2) portuguese version used to identify violence within couples. *Saúde Pública*, 18, 3. [Çevrim-içi: http://www.scielo.br/scielo.php?script=sci_arttext&pid] Erişim Tarihi: 26 Ekim 2018
- Mulaik, S.A., James, L.R., Van Alstine, J., Bennet, N., Lind, S. & Stilwell, C.D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105(3), 430-45.
- Mullis, I., Martin, M.O., Foy, P. & Arora, A. *TIMSS 2011 international results in mathematics*. 2012, 1.
- Murphy, K. R. & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications* (6th Edition). Upper Saddle River, NJ: Prentice Hall
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory*. (Third Edition). New York: McGraw-Hill, Inc.
- Oxford, R. (1993). Instructural implications of gender diferences in second/foreign langage (L2) learning styles and strategies. *Applied Language Learning*, 4(1-2), 65-94

- Oxford, R. (1993) & Nyikos, M. (1989). Variables affecting choice of language learning strategies by university students. *The Modern Language Journal*, 73, 291-300.
- Oxford, R., Nyikos, M. & Ehrman, M. (1988). Vive la difference? Reflections on sex differences in use of language learning strategies. *Foreign Language Annals*, 21(4).
- Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü. [Çevrim-içi: <https://odsgm.meb.gov.tr/kurslar/>] Erişim Tarihi: 22 Şubat 2018
- Ölçülüoğlu, R. (2015). *TIMSS 2011 Türkiye sekizinci sınıf matematik başarısını etkileyen değişkenlerin bölgelere göre incelenmesi*. Yayınlanmış yüksek lisans tezi. [Çevrim-içi: www.openaccess.hacettepe.edu.tr] Erişim Tarihi: 18 Ocak 2019
- Pektaş, M., İncikabı, L. ve Yaz, Ö. (2015). Orta Öğretim Fen Ders Kitaplarının TIMSS Çerçevesine Göre Analizi. *Adıyaman University Journal of Educational Sciences*, 5(1), 29-48
- Putnick, D.L. & Bornstein, M.H. (2016). *Measurement invariance conventions and reporting: The state of the art and future directions for psychological research*. *Developmental Review* 41, 2015, 71-90.
- Raykov, T. & Marcoulides, G. A. (2008). *An Introduction to Applied Multivariate Analysis* (First Edition). NY: Taylor & Francis Group.
- Sarı, M. H., Arıkan, S. ve Yıldızlı, H. (2017). 8. Sınıf Matematik Akademik Başarısını Yordayan Faktörler-TIMSS 2015. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi Journal of Measurement and Evaluation in Education and Psychology*, 8(3), 2017, 246-265.
- Satorra, A. & Bentler, P. M. (2011). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika: A Journal of Quantitative Psychology*, 66(4), 507-514.
- Schmitt, N. & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18, 210-222.

- Schoot, R., Lugtig, P. & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486–492.
- Sheorey, R. (1999). An examination of language learning strategy use in the setting of an indigenized variety of English. *System*, 28(2), 173-190.
- Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research*, 45, 83-117.
- Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher*, 36(8), 477-481.
- Steenkamp, B., E., M. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78-107
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42, 893-898.
- Suhr, D.D. (2006). *Exploratory or confirmatory analysis*. NC: SAS Institute.
- Şekercioğlu, G. (2018). *Measurement invariance: Concept and implementation*. *International Online Journal of Education and Teaching (IOJET)*, 5(3), 609-634. [Çevrim-içi: <http://iojet.org/index.php/IOJET/article/view/439/257>] Erişim Tarihi: 02 Temmuz 2018
- Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlilik*. Ankara: Seçkin Yayıncılık.
- Tabachnick, B.G. & Fidell, L.S. (2013). *Using multivariate statistics* (6th ed.). New Jersey: Pearson.
- TIMSS 2015 Ulusal Raporu [Çevrim-içi: http://timss.meb.gov.tr/?page_id=147] Erişim Tarihi: 17.03.2018
- TIMSS and PIRLS web adresi [Çevrim-içi: <http://timssandpirls.bc.edu/>] Erişim Tarihi: 22 Temmuz 2018

- Tucker, K. L., Özer, D. J., Lyubomirsk, S. & Boehm, J. K. (2006). Testing for measurement invariance in the satisfaction with life scale: A comparison of Russians and North Americans. *Social Indicators Research*, 78, 341–360.
- Uyar, Ş. ve Doğan, N. (2014). PISA 2009 Türkiye örnekleminde öğrenme stratejileri modelinin farklı gruplarda ölçme değişmezliğinin incelenmesi. *Uluslararası Türk Eğitim Bilimleri Dergisi*, 2, 30-43.
- Uzun, B. ve Öğretmen, T. (2010). Fen başarısı ile ilgili bazı değişkenlerin TIMSS-R Türkiye örnekleminde cinsiyete göre ölçme değişmezliğinin değerlendirilmesi. *Eğitim ve Bilim*, 35, 155, 26-35.
- Weston, R. & Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist*, 34(5), 719-751.
- Wheaton, B., Muthen, B., Alwin, D. F. & Summers, G. F. (1977). Assessing reliability and stability in panel models. *Sociological methodology*, 8(1), 84-136.
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research*, 281-324. Washington, DC: American Psychological Association.
- Wu. A. D., Li, Z. & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multigroup confirmatory factor analysis: A demonstration with TIMSS data practical assessment, *Research and Evaluation*, 12(3), 1-26.
- Van de Vijver, F. J. R. & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage
- Van de Vijver, F. J. R. & Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. In R. K. Hambleton, P. F. Merenda, ve C. D. Spielberger (Eds.), *Adapting Educational and Psychological Tests For Cross-Cultural Assessment*, 39-63. Mahwah, NJ: Lawrence Erlbaum.

Van de Vijver, F. & Tanzer, N. K. (2004). *Bias and equivalence in cross-cultural assessment: An overview.* [Çevrim-içi: http://resilienceresearch.org/files/articlevandevijver_tanzer.pdf] Erişim Tarihi: 21 Ekim 2018

Vandenberg, R. J. & Lance, C. E. (2000). A review and synthesis of the MI literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-69, 4-70. [Çevrim-içi: <http://orm.sagepub.com/cgi/reprint/3/1/4.pdf>] Erişim Tarihi: 24 Ekim 2005

Yandı, A., Köse, İ. ve Uysal Ö. (2017). Farklı yöntemlerle ölçme değişmezliğinin incelenmesi: PISA 2012 örneği. *Mersin Üniversitesi Eğitim Fakültesi Dergisi - Mersin University Journal of the Faculty of Education "Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 13(1), 2017, 243-253. doi: <http://dx.doi.org/10.17860/mersinefd.305952>.

Yılmaz, V. (2004). Lisrel ile yapısal eşitlik modelleri: Tüketici şikâyetlerine uygulanması. *Sosyal Bilimler Dergisi*, 1.

EKLER

EK 1: TIMSS 2015 SEKİZİNCİ SINIF MATEMATİK ÖRNEK SORULARI

Sekizinci sınıf TIMSS 2015 matematik örnek sorularına aşağıda yer verilmiştir.

SORU 1

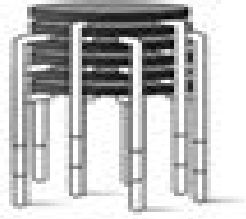
Aşağıdakilerden hangisi $\frac{3}{4}$ 'e en yakın değerdir?

- A) 0,34
- B) 0,43
- C) 0,74
- D) 0,79

SORU 2

Can'ın evinde aşağıdaki gibi üst üste konulmuş tabureler bulunmaktadır.

Şekil 27. TIMSS 2015 Sekizinci Sınıf Matematik 2. Madde Şekli



Her bir taburenin yüksekliği 49 cm'dir.

2 tabure üst üste konulduğunda yükseklikleri 55 cm olmaktadır.

Buna göre 6 tane tabure üst üste konulduğunda en üstte bulunan taburenin yerden yüksekliği ne kadar olur?

- A) 79 cm
- B) 85 cm
- C) 110cm
- D) 165 cm

SORU 3

$\frac{2}{3}x+1$ bir tam sayıdır.

Buna göre x ile ilgili aşağıdaki ifadelerden hangisi kesinlikle doğrudur?

- A) x bir sayıdır
- B) x bir çift sayıdır
- C) x, 3'ten büyük bir sayıdır
- D) x, 3'e bölünebilen bir sayıdır

SORU 4

Aşağıdaki her bir ifadenin doğru olması için kutulara < , > ya da = sembollerinden uygun olanı yerleştiriniz.

0,35 0,350

0,35 0,4

0,35 0,305

0,35 0,035

SORU 5

Elinizde 2 yumurta ve 0,3 litre süt ile yapılan bir kek tarifi bulunmaktadır.

Sizin 5 yumurtanız var ve yapabileceğiniz en büyük keki yapmak istiyorsunuz.

Yapılabilecek en büyük keki hazırlamak için kaç litre süte ihtiyacınız vardır?

Yanıt: _____ litre

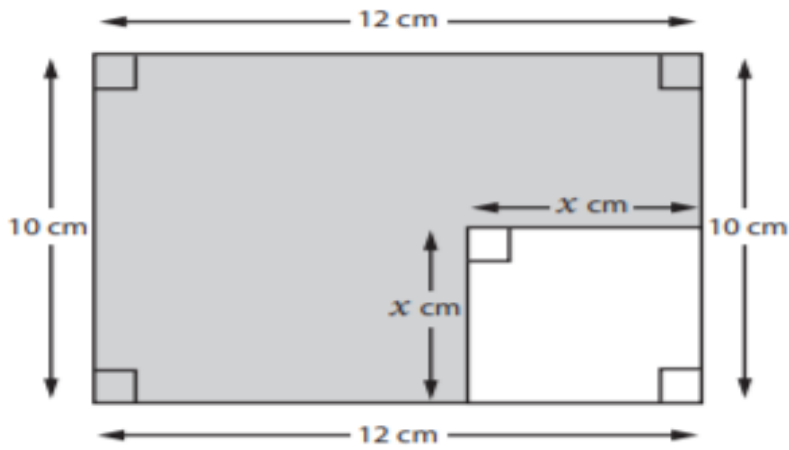
SORU 6

$\frac{a^2}{2} - 6a + 36$ ifadesinin $a = 3$ için değeri kaçtır?

- A) 58,5
- B) 27
- C) 22,5
- D) 21

SORU 7

Şekil 28. TIMSS 2015 Sekizinci Sınıf Matematik 7. Madde Şekli



Yukarıdaki şeklin taralı bölgesinin alanını x cinsinden yazınız.

Yanıt: _____ cm^2

SORU 8

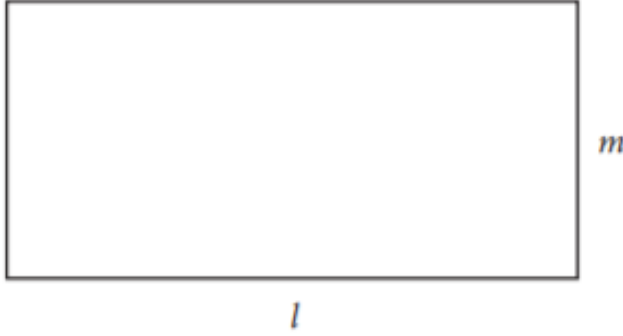
$$y = \sqrt{x - 9}$$

Yukarıdaki ifadede $x = 25$ iken y 'nin degeri kaçtır?

- A) 3
- B) 4
- C) 8
- D) 16

SORU 9

Şekil 29. TIMSS 2015 Sekizinci Sınıf Matematik 9. Madde Şekli



Yukarıdaki şekil uzun kenarı l , kısa kenarı m olan bir dikdörtgendir.

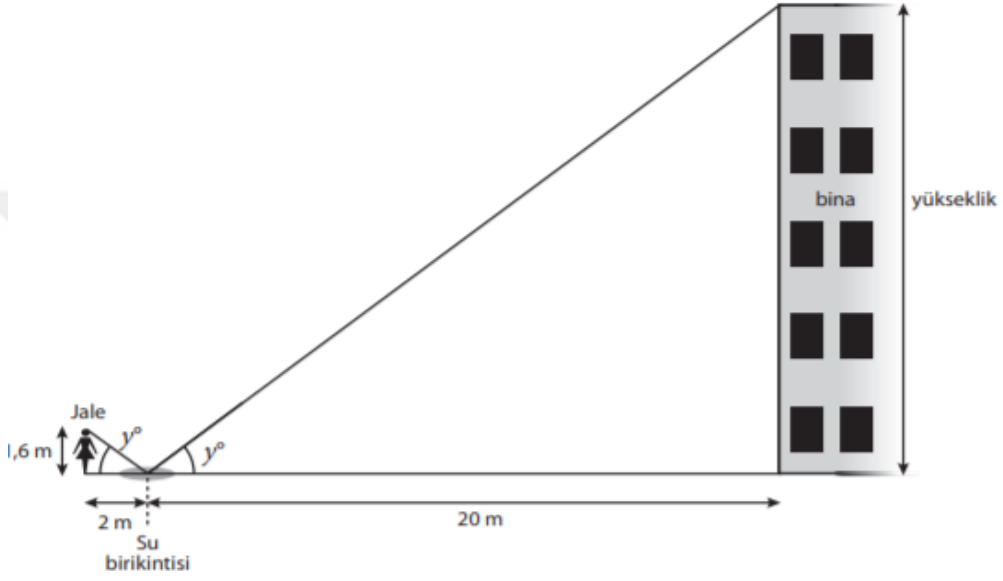
Eger bu dikdörtgenin uzun kenarı iki katına çıkarılır ve kısa kenarı aynı kalırsa, yeni dikdörtgenin alanını (A) aşağıdaki formüllerden hangisi verir?

- A) $A = 2l + 2m$
- B) $A = 2l + 4m$
- C) $A = 2lm$
- D) $A = 4lm$

SORU 10

Jale bir su birikintisinin yanında duruyor. Bu su birikintisinde kar_ısındaki binanın tepe noktasının yansımasını g_orebiliyor. Jale'nin g_orus_uz çizgisi su birikintisi ile y° 'lik a_aci yapıyor ve su birikintisinden aynı a_aci ile yansıyor.

Şekil 30. TIMSS 2015 Sekizinci Sınıf Matematik 10. Madde Şekli

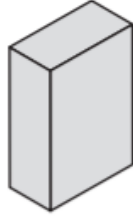


Uzaklıklar ve yükseklik yukarıdaki resimde gösteriliyor. Buna göre binanın yüksekliği ne kadardır?

Yanıt: _____ m

SORU 11

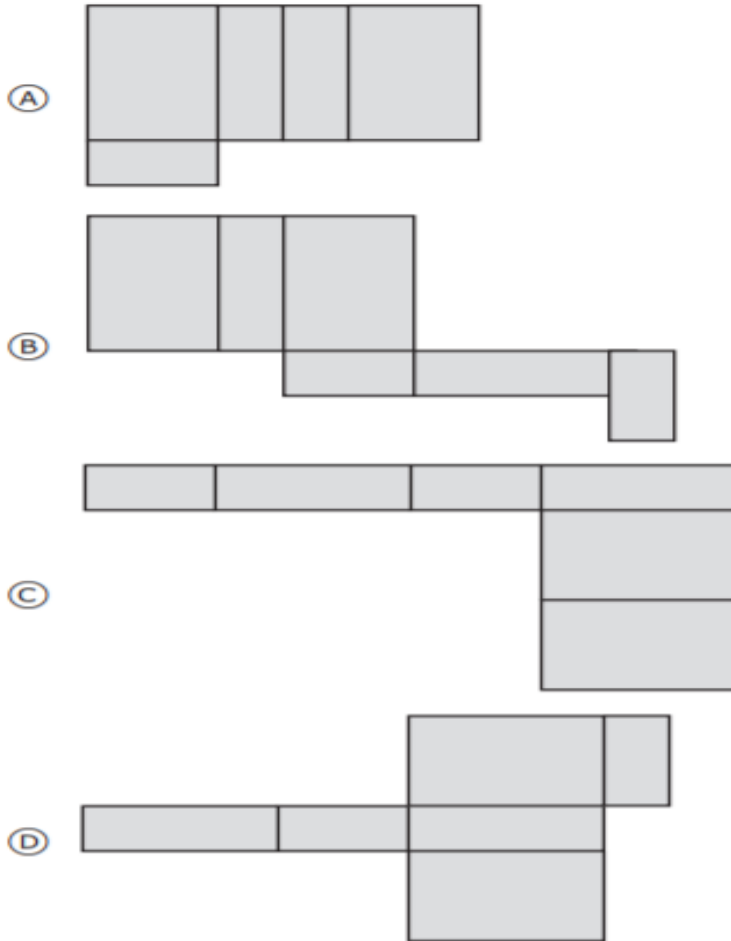
Şekil 31. TIMSS 2015 Sekizinci Sınıf Matematik 11. Madde Şekli



Yukarıda dikdörtgenler prizması şeklinde bir kutu verilmiştir.

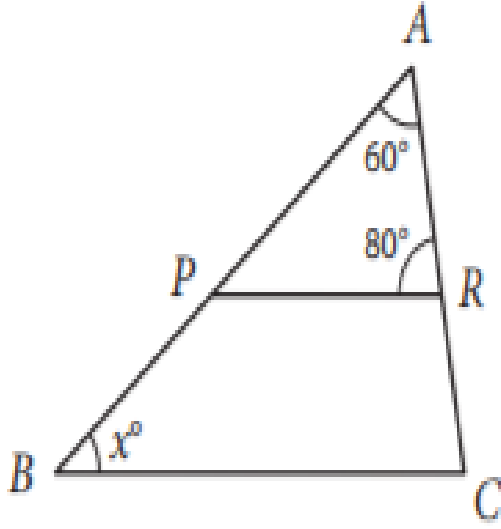
Aşağıdaki şekillerden hangisi katlandığında yukarıdaki kutu elde edilir?

Şekil 32. TIMSS 2015 Sekizinci Sınıf Matematik 11. Madde Şekli



SORU 12

Şekil 33. TIMSS 2015 Sekizinci Sınıf Matematik 12. Madde Şekli



PR ve BC kenarları birbirine paraleldir.

Şekilde verilenlere göre x 'in degeri nedir?

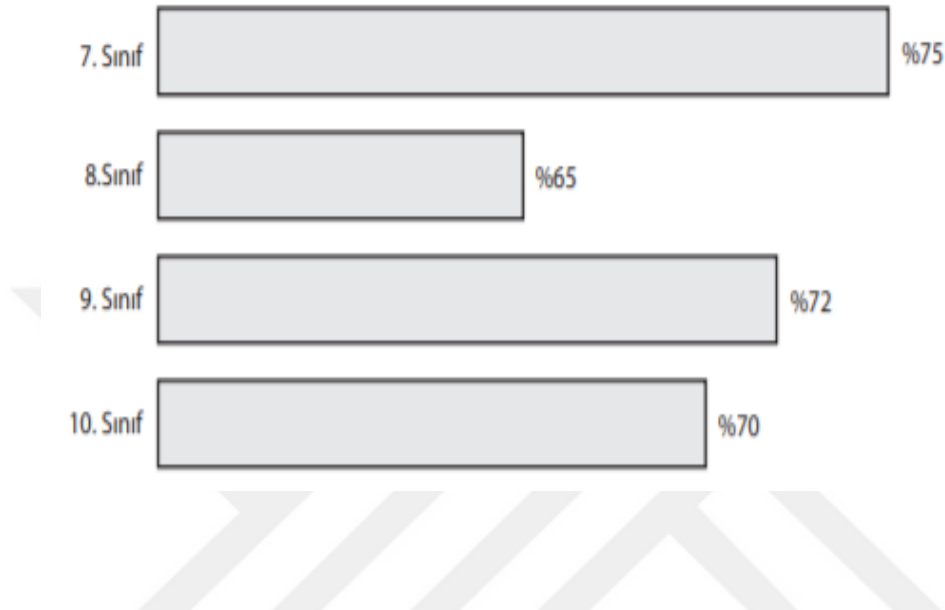
Yanıt: _____

SORU 13

Okul Spor Araştırması — 7-10. Sınıflar

Futbolu en sevdiği spor olarak seçen öğrencilerin yüzdesi:

Şekil 34. TIMSS 2015 Sekizinci Sınıf Matematik 13. Madde Şekli



Batu'nun okulunda 7. sınıftan 10. sınıfa kadarki öğrencilere en sevdikleri spor sorulmuştur. Her bir sınıf seviyesinde 100 öğrenci bulunmaktadır. Yukarıdaki grafik, futbolu seçen öğrencilere ait sonuçları göstermektedir.

Batu, 7. ve 8. sınıflara ait sonuçları karşılaştırmıştır. Bu karşılaştırma sonucunda 7. sınıftaki futbolu seçen öğrenci sayısının 8. sınıftaki futbolu seçen öğrenci sayısının iki katı kadar olduğu sonucuna ulaşmıştır.

Grafik, Batu'nun bu hatayı yapmasına nasıl yol açmıştır, açıklayınız.

SORU 14

Öğrencilerin Sevdiği Televizyon Programları

Şekil 35. TIMSS 2015 Sekizinci Sınıf Matematik 14. Madde Şekli



Grafik, 240 öğrencinin en çok sevdiklerini söyledikleri televizyon program türlerini göstermektedir.

Aşağıdakilerden hangisi Tarih programlarını sevdiğini söyleyen öğrencilerin sayısıdır?

- A) 20
- B) 30
- C) 40
- D) 60

SORU 15

Veri açıklığı en küçük ve ortalaması en büyük olan sayı listesi hangi seçenekte verilmiştir?

- A) 6 8 12 23 46
- B) 6 8 12 28 46
- C) 6 8 12 23 51
- D) 6 8 12 18 51

EK 2: TIMSS 2011 SEKİZİNCİ SINIF MATEMATİK ÖRNEK SORULARI

Sekizinci sınıf TIMSS 2011 matematik örnek sorularına aşağıda yer verilmiştir.

SORU 1

Burcu her biri 6 yumurta alan kutulara yumurtaları yerleştiriyor.

Burcu'nun 94 tane yumurtası vardır.

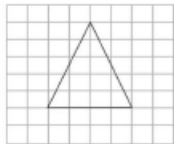
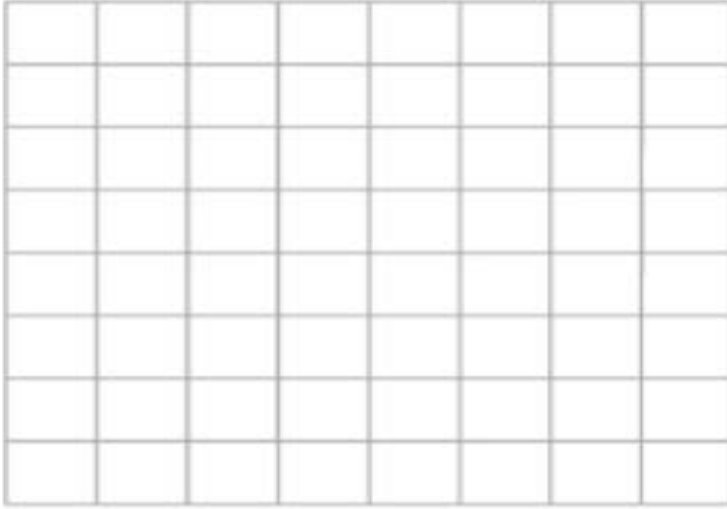
Buna göre tüm yumurtaları yerleştirmek için Burcu'nun en az kaç kutuya ihtiyacı vardır?

Yanıt: _____ kutu

SORU 2

Aşağıdaki kareli bölümde, her bir karenin kenar uzunluğu 1 cm'dir. Bu bölüme tabanı 4 cm, yüksekliği 5 cm olan bir ikizkenar üçgen çiziniz.

Şekil 36. TIMSS 2011 Sekizinci Sınıf Matematik 2. Madde Şekli



Yanıt:

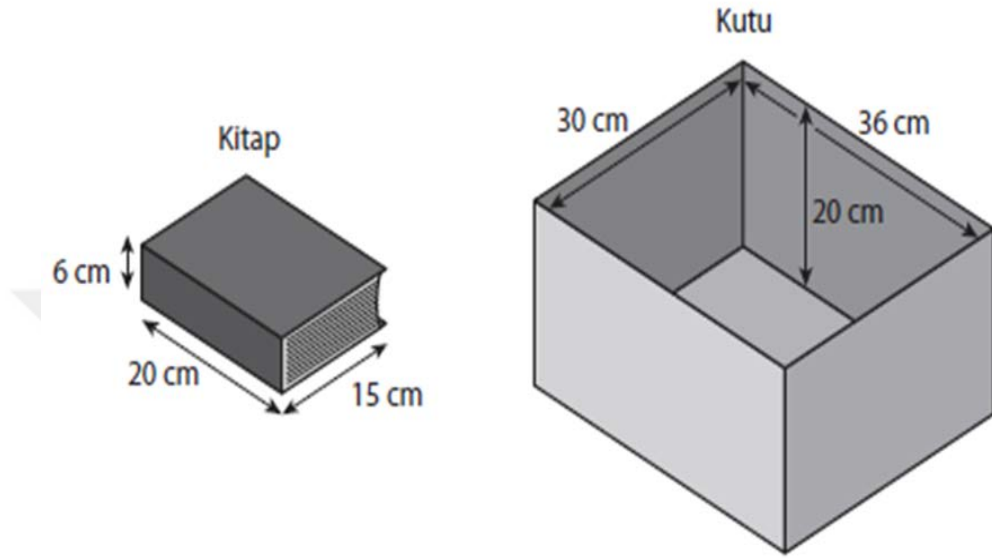
Not: Kareli bölümün her hangi bir yerine bu ölçülerdeki bir üçgen çizilebilir.

SORU 3

Banu ařağıdaki dikdörtgenler prizması řeklindeki kutuya itapları yerleřtiriyor.

Bütün kitaplar aynı büyüklüktedir.

řekil 37. TIMSS 2011 Sekizinci Sınıf Matematik 3. Madde řekli



Buna göre kutunun içine en fazla kaç kitap sığar?

Yanıt: _____

SORU 4

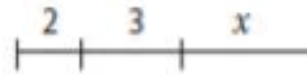
Aşağıdakilerden hangisi, $2x+3x$ ifadesinin anlamını göstermek için kullanılabilir?

Şekil 38. TIMSS 2011 Sekizinci Sınıf Matematik 4. Madde Şekli

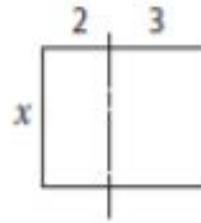
(A) Bu parçanın uzunluğu:



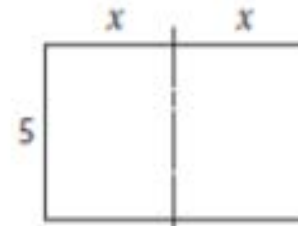
(B) Bu parçanın uzunluğu:



(C) Bu şeklin alanı:



(D) Bu şeklin alanı:



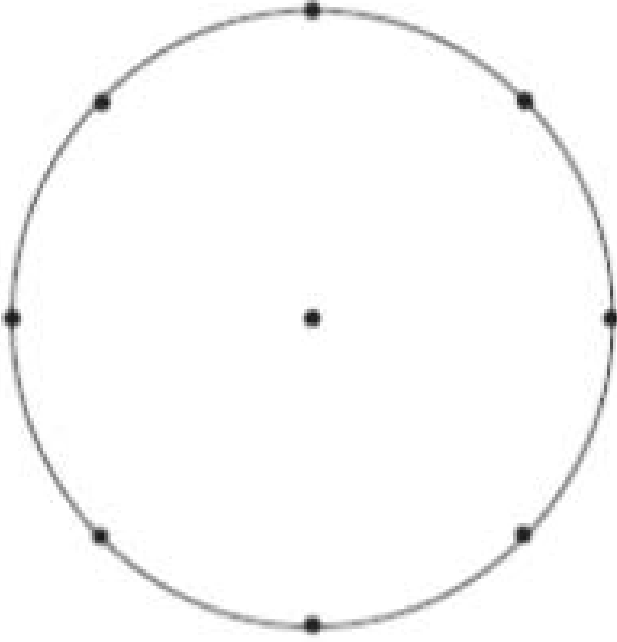
Yanıt: C

SORU 5

Bir okuldaki 400 öğrencinin 50'si mühendislik fakültelerine, 100'ü tıp fakültelerine, 150'si eğitim fakültelerine gitmeyi, kalanı ise çalışma hayatına girmeyi planlıyor.

Yukarıdakilerin her birini yapmayı planlayan öğrencilerin oranlarını gösteren bir daire grafiği yapmak için aşağıdaki çemberi kullanınız. Yaptığınız grafikteki alanları isimlendiriniz.

Şekil 39. TIMSS 2011 Sekizinci Sınıf Matematik 5. Madde Şekli



SORU 6

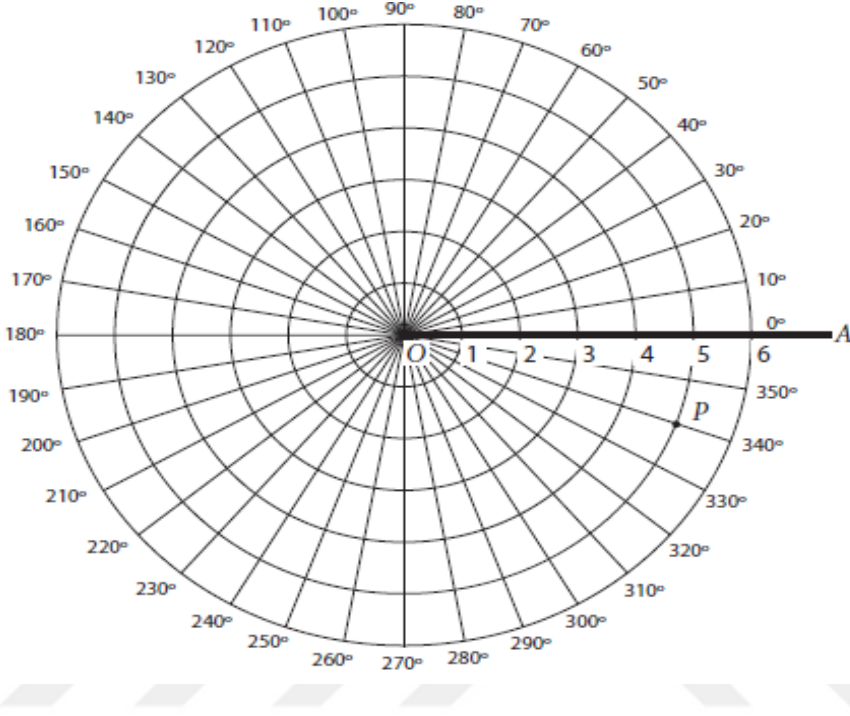
$\frac{3x}{8} + \frac{x}{4} + \frac{x}{2}$ ifadesini sadeleştiriniz. İşlemlerinizi gösteriniz.

Yanıt: _____

SORU 7

Aşağıdaki şekil, noktaların konumlarını belirlemek için kullanılan bir sistemi göstermektedir.

Şekil 40. TIMSS 2011 Sekizinci Sınıf Matematik 7. Madde Şekli



Bu sistemde P noktasının yeri, P noktasunun O'dan uzaklığıyla ve QA hattından saat ibresinin tersi yönde hareket ederek OP hattına kadar yapılan dönüş açısıyla gösterilmektedir. Buna göre, P noktasının koordinatları $(5, 120^\circ)$ ' dir.

A) Yukarıdaki daire grafiği üzerinde $B(3, 30^\circ)$ ve $C(4, 340^\circ)$ noktalarını işaretleyiniz.

B) BOC açısını çiziniz. BOC açısının ölçüsü nedir?

BOC açısı= _____

Yanıt: B ve C noktaları şekilde gösterilmiştir. BOC açısı 90° dir.

SORU 8

Cem ve Ceyda okul başkanlığına aday oldular. Seçim sonuçları şu şekildedir:

Cem %80

Ceyda %20

Rastgele seçilmiş bir öğrencinin Cem'e oy veriş olma olasılığı hakkında ne söylenebilir?

- A) Öğrenci kesinlikle Cem'e oy vermiştir.
- B) Öğrencinin Cem'e oy vermiş olma olasılığı vardır.
- C) Öğrencinin Cem'e oy vermiş olma olasılığı yoktur
- D) Öğrenci kesinlikle Ceyda'ya oy vermiştir.

Yanıt: B

SORU 9

$$x+y=12 \text{ ve } 2x+5y=36$$

Buna göre x ve y'nin değeri nedir?

- A) $x=2, y=10$
- B) $x=4, y=8$
- C) $x=6, y=6$
- D) $x=8, y=4$

Yanıt: D

SORU 10

Aşağıdakilerden hangisi $4(3+x)$ 'e eşittir?

- A) $12+x$
- B) $7+x$
- C) $12+4x$
- D) $12x$

Yanıt: C

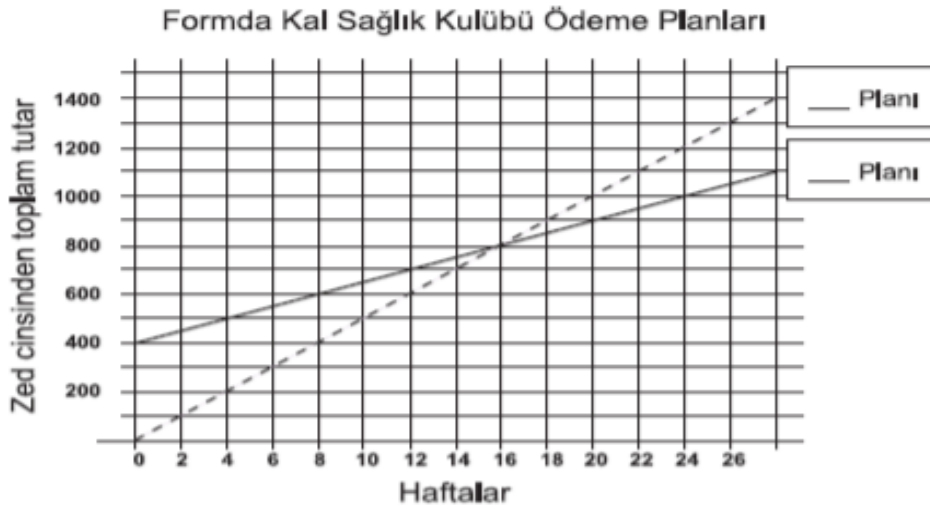
EK 3: TIMSS 2007 SEKİZİNCİ SINIF MATEMATİK ÖRNEK SORULARI

Sekizinci sınıf TIMSS 2007 matematik örnek sorularına aşağıda yer verilmiştir.

SORU 1

Formda Kal Sağlık Klübü müşterilerine 2 farklı ödeme planı sunmaktadır. A planının başlangıç ödemesi 400 zed ve bir haftalık ödeme 25 zed'dir. B planının başlangıç yoktur fakat bir haftalık ödeme 50 zed'dir. Aşağıdaki grafik A planı ve B planının fiyatlarını karşılaştırmaktadır.

Şekil 41. TIMSS 2007 Sekizinci Sınıf Matematik 1. Madde Şekli



- A. Grafik üzerinde verilen doğrulardan hangisinin A, hangisinin B planına ait olduğunu verilen kutucuklara yazınız.
- B. Kaçınıcı haftada A ve B planları için aynı miktarda ödeme yapmış oluruz?
- C. 24. haftada, iki plan arasındaki toplam fiyat farkı nedir?

ÖĞRENME ALANI	Veri ve Olasılık
BİLİŞSEL ALAN	Uygulama
DOĞRU YANIT	Puanlama Rehberi

SORU 2

Bir bahçıvan çimlendireceği alana ekmek için 4,45 kg çayır tohumu ile 2,735 kilogram yonca tohumunu karıştırıyor. Bahçıvanın elinde kaç kilogram tohum karışımı var?

Yanıt: _____

ÖĞRENME ALANI	Sayılar
BİLİŞSEL ALAN	Uygulama
DOĞRU YANIT	Puanlama Rehberi

SORU 3

Aşağıdaki karelerden her birinin içine, bu ifadenin mümkün olan en yüksek değerde olmasını sağlayacak şekilde + ya da – işareti koyunuz.

Şekil 42. TIMSS 2007 Sekizinci Sınıf Matematik 3. Madde Şekli

$$-5 \square -6 \square 3 \square -9$$

ÖĞRENME ALANI	Sayılar
BİLİŞSEL ALAN	Uygulama
DOĞRU YANIT	Puanlama Rehberi

SORU 4

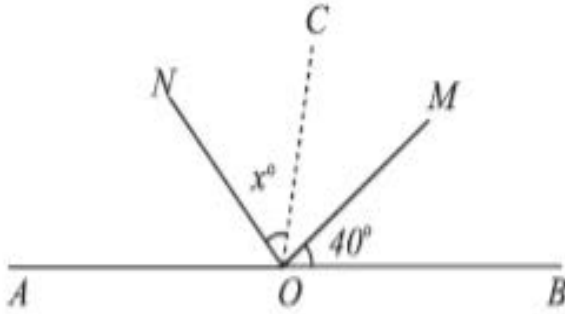
$\frac{x}{3} > 8$ ifadesi aşağıdakilerden hangisine denktir?

- A) $x < 5$
- B) $x < 24$
- C) $x > \frac{8}{3}$
- D) $x > 5$
- E) $x > 24$

ÖĞRENME ALANI	Cebir
BİLİŞSEL ALAN	Bilgi
DOĞRU YANIT	E

SORU 5

Şekil 43. TIMSS 2007 Sekizinci Sınıf Matematik 5. Madde Şekli



Yukarıdaki şekilde A,O ve B noktaları bir doğru üzerindedir. OM, COB açısını; ON, AOC açısını iki eşit parçaya böldüğüne göre x kaç derecedir?

Yanıt: _____

ÖĞRENME ALANI	Geometri
BİLİŞSEL ALAN	Uygulama
DOĞRU YANIT	Puanlama Rehberi

SORU 6

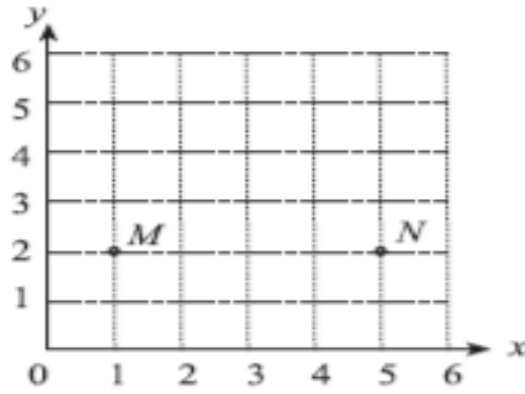
Alanı 100 metrekare olan bir karenin çevresinin uzunluğu ne kadardır?

Yanıt: _____

ÖĞRENME ALANI	Geometri
BİLİŞSEL ALAN	Uygulama
DOĞRU YANIT	Puanlama Rehberi

SORU 7

Şekil 44. TIMSS 2007 Sekizinci Sınıf Matematik 7. Madde Şekli



Yukarıdaki şekilde M ve N noktaları gösterilmiştir. Mustafa MNP'nin ikizkenar üçgen olmasını sağlayacak bir P noktası aramaktadır. P noktasının koordinatları aşağıdakilerden hangisi olabilir?

- A) (3,5)
- B) (3,2)
- C) (1,5)
- D) (5,1)

ÖĞRENME ALANI	Geometri
BİLİŞSEL ALAN	Uygulama
DOĞRU YANIT	A

SORU 8

Bir firma, bir yıl içinde 1426 ton suni gübre satmıştır. Bu firmanın suni gübre satışı sonraki yılda %15 azalmıştır. Bu firmanın ikinci yıldaki suni gübre satışı yaklaşık olarak kaç tondur?

- A) 200
- B) 300
- C) 1200
- D) 1600
- E) 1700

ÖĞRENME ALANI	Sayılar
BİLİŞSEL ALAN	Uygulama
DOĞRU YANIT	C

SORU 9

$$3(2x-1)+2x=21$$

İfadesinde x'in değeri nedir?

- A) -3
- B) $-\frac{11}{2}$
- C) $-\frac{11}{4}$
- D) 3

ÖĞRENME ALANI	Sayılar
BİLİŞSEL ALAN	Bilgi
DOĞRU YANIT	D

SORU 10

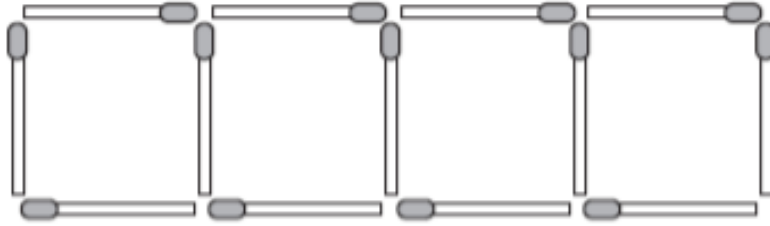
Ebru'nun ceketlerinin sayısı Serpil'in eketlerinin sayısından 3 fazladır. Ebru'nun ceketlerinin sayısı n ise Serpil'in n cinsinden kaçane ceketi vardır?

- A) $n-3$
- B) $n+3$
- C) $3-n$
- D) $3n$

ÖĞRENME ALANI	Cebir
BİLİŞSEL ALAN	Bilgi
DOĞRU YANIT	A

SORU 11

Şekil 45. TIMSS 2007 Sekizinci Sınıf Matematik 11. Madde Şekli



Şekildeki 4 kareyi yapmak için 13 kibrit çöpü kullanılmıştır. Aynı yöntemle 73 kibrit kullanılarak kaç tane kare yapılabilir?

Yanıtınızı bulmanızı sağlayan hesaplanınızı gösteriniz.

Yanıt: _____

ÖĞRENME ALANI	Cebir
BİLİŞSEL ALAN	Sonuç Çıkarma
DOĞRU YANIT	Puanlama Rehberi

(TIMSS Ulusal Raporu)

Sorulardan da anlaşıldığı gibi birçok kazanımın yer aldığı bilme, kavrama ve uygulama düzeyinde sorularla öğrencilerin konuyu bilip bilmedikleri, analitik düşünme güçleri yoklanmaktadır. Öğrencilerin konuyu ne derece anladığına bakmak için sık sık açık uçlu sorular kullanılmış olup, öğrencilerin o soru hakkında yorum yapması, fikir beyan etmesi istenmiştir. Çoktan seçmeli ve açık uçlu sorular tercih edilmiştir. Açık uçlu soruların cevapları TIMSS Ulusal Raporu'nda verilmemiş olup Puanlama Rehberi'ne yönlendirilmiştir. Genel bir ifadeyle, TIMSS 2015 yayınlanan örnek sorulara ve TIMSS 2011 ve TIMSS 2009 sorularının tamamına bakıldığında birçok kazanımın, alt öğrenme alanı, öğrenme alanının ve birçok bilişsel düzeyin yoklandığı, öğrencileri konuyu ezberlemekten çok düşünmeye ve sorgulamaya yönlendirdiği söylenebilir.

EK 4: ÖZGEÇMİŞ SAYFASI

Kişisel Bilgiler	
Ad-Soyad	Özge ÖNCÜ (GÖÇER)
Doğum Yeri ve Tarihi	Altınyayla/BURDUR 04.01.1993
İletişim	
E-mail	ozgegocer07@hotmail.com ozge.gocer.07@gmail.com
Eğitim Bilgileri	
Lisans Öğrenimi	Akdeniz Üniversitesi Eğitim Fakültesi İlköğretim Matematik Öğretmenliği Lisans (2011-2015)
	Anadolu Üniversitesi İktisat Lisans (2012-2016)
Yüksek Lisans Öğrenimi	Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü Eğitimde Ölçme ve Değerlendirme Tezli Yüksek Lisans (2016-2019)
Bildiği Yabancı Diller	İngilizce (C1)
Yabancı Dil Eğitimi	St. Giles College London U.K.
İş Tecrübesi	
Stajlar	Leyla Kahraman Sevim Ertünü Ortaokulu (2014)
	Meryem Mustafa Ege Ortaokulu (2014)

	Dr. Cahit Ünver Ortaokulu (2015)
Çalıştığı Kurumlar	Antalya Mecdude Başakıncı Ortaokulu (2015-2016) (MEB/Matematik Öğretmeni)
	Antalya Kozağacı Ortaokulu (2016-2018) (MEB/Matematik Öğretmeni)
	Antalya İsmail Hakkı Kaya Ortaokulu (2018-....) (MEB/Matematik Öğretmeni)
Projeler	Tubitak 4006
Akademik Çalışmalar	
Makaleler	JASSS The Journal of Academic Social Science Studies – International Journal of Social Science, Aralık 2018 Doi number: http://doi.org/10.9761/JASSS7789 Number: 73, p. 71-79, Winter 2018 ‘Yayımlanmış Makalelerin Bilimsel Araştırma Süreçlerine Uygunluk Düzeylerinin Değerlendirilmesi’
	JRET Eğitim ve Öğretim Araştırmaları Dergisi - Journal of Research in Education and Teaching, Mayıs 2017, Cilt: 6, Sayı: 2, Makale No: 35, ISSN: 2146-9199 ‘Teog Sınav Sürecinde Okul İdarecilerinin Öğretmenler Üzerinde Yarattığı Etki Üzerine Bir Araştırma’
Bildiriler	EYFOR Kongresi – 8. Uluslararası Eğitim Yönetimi Forumu Ankara TOBB ETÜ 19-21 Ekim 2017

	‘Öğretmen Adaylarının Matematik Alan Derslerine Yönelik Tutumlarının İncelenmesi’
	8th International Congress on New Trends in Education – 8. Uluslararası Eğitimde Yeni Yönelimler Kongresi ICONTE Antalya 18-20 Mayıs 2017 ‘Teog Sınav Sürecinde Okul İdarecilerinin Öğretmenler Üzerinde Yarattığı Etki Üzerine Bir Araştırma’
	IV th International Eurasian Educational Research Congress - EJER Kongresi Denizli Pamukkale Üniversitesi 11-14 Mayıs 2017 ‘Yayımlanmış Makalelerin Bilimsel Araştırma Süreçlerine Uygunluk Düzeylerinin Değerlendirilmesi’
	26 th International Conference on Educational Sciences - 26. Uluslararası Eğitim Bilimleri Kongresi ICES – UEBK Side/Manavgat 20-23 Nisan 2017 ‘İlbap Sürecinde Okul İdarecilerinin Öğretmenler Üzerinde Yarattığı Etki Üzerine Bir Araştırma’
	5. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi Antalya Akdeniz Üniversitesi (Uluslararası katılımlı) 1-3 Eylül 2016 ‘Performans Takip Sınavında (PTS) Kullanılan Alt Testlerin Psikometrik Özelliklerinin İncelenmesi’

EK 5: BİLDİRİM SAYFASI

BİLDİRİM

Hazırladığım tezin tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin kâğıt ve elektronik kopyalarının Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

Tezimin tamamı her yerden erişime açılabilir.

Tezim sadece Akdeniz Üniversitesi yerleşkelerinden erişime açılabilir.

Tezimin 1 yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin tamamı her yerden erişime açılabilir.

02 / 05 /2019

Özge ÖNCÜ

EK 6: İNTİHAL RAPORU

Home Page | Use this | Messages | Notices | Login | Community | Help | Logout

turnitin

Assigned | Status | Date Due | Library | Calendar | Division | Preferences

001481140 NONE - 028 FEB - 028 FEB

About this page
This is your assignment entry. To view a paper's status, select the paper's Summary Report card in the summary column. A green icon indicates that the Summary Report has not yet been generated.

028 Feb
NONE - 028 FEB - 028 FEB

Submit File

FILE	STATUS	DATE	SCORE	RESPONSE	FILE	WFILE ID	DATE
028 Feb	028 FEB	028 FEB	100%			153759132	15/11/2016

[Done] [Go to] [Report] [Edit assignment settings] [View my submissions]

Dr. Öğr. Üy. Halim Kaya
Halim